

# DIAGNOSING ACUTE HEART FAILURE IN THE EMERGENCY DEPARTMENT

CRAIG JAMES FERGUSON

A thesis submitted in partial fulfilment of the  
requirements of the Manchester Metropolitan  
University for the degree of Doctor of Philosophy

Faculty of Health, Psychology and Social Care  
The Manchester Metropolitan University  
In collaboration with Manchester Royal Infirmary  
June 2014

PhD            2014

Table of Contents.....	2
List of Figures.....	9
List of Tables.....	14
Abbreviations.....	16
Abstract.....	19
Declaration.....	21
Acknowledgements.....	21
Dedication.....	22

## **CHAPTER 1 BACKGROUND .....23**

<b>1.1 Heart Failure .....</b>	<b>23</b>
1.1.1 The Burden of Heart Failure .....	23
1.1.2 The Definition of Heart Failure.....	23
1.1.3 Demographics.....	25
1.1.4 The Consequences of Heart Failure.....	27
1.1.5 The Aetiology of Heart Failure.....	28
1.1.6 The Presentation of Heart Failure.....	31
1.1.7 Differential Diagnosis.....	32
1.1.8 Diagnosing in the Emergency Department .....	33
1.1.9 The Pathophysiology of Heart Failure.....	35
1.1.10 Neuroendocrine response.....	37
1.1.11 Remodelling.....	40
1.1.12 Current Guidelines.....	41
1.1.13 Aims of This Thesis.....	42

## **CHAPTER 2 DIAGNOSING HEART FAILURE.....44**

2.1.1 Diagnosing Heart Failure.....	44
2.1.2 Studies on the Presentation of Heart Failure .....	45
2.1.3 Clinical History.....	50
2.1.4 Clinical Examination.....	51
<b>2.2 Investigations.....</b>	<b>53</b>
2.2.1 The Electrocardiograph .....	53
2.2.2 The Chest X-Ray.....	57
2.2.3 Blood Tests .....	60
2.2.4 Echocardiography.....	60

2.2.5 Invasive Techniques .....	63
2.2.6 Magnetic Resonance Imaging.....	64
2.2.7 Nuclear Studies .....	64
2.2.8 Peripheral Temperature.....	65
2.2.9 Response to Treatment .....	65
2.2.10 Scoring Systems for the Diagnosis of Heart Failure.....	65
2.2.11 Assessing the Severity of Heart Failure.....	71
2.2.12 Diagnosing Diastolic Heart Failure.....	73
2.2.13 Emergency Department Echocardiograms.....	75
2.2.14 Summary.....	75
<b>CHAPTER 3 THE NATRIURETIC PEPTIDES .....</b>	<b>77</b>
3.1.1 Introduction .....	77
3.1.2 Characteristics of an ideal marker – from Collinson (2004) .....	77
3.1.3 Atrial Natriuretic Peptide.....	79
3.1.4 Brain Natriuretic Peptide .....	80
3.1.5 The Function of Natriuretic Peptides.....	82
3.1.6 Measuring BNP .....	84
3.1.7 Optimal Utility of BNP.....	87
3.1.8 Comparing BNP with NT-proBNP.....	90
3.1.9 Prognostic value of BNP levels.....	91
3.1.10 Therapeutic use of exogenous BNP.....	94
3.1.11 Other Diagnostic Roles for BNP .....	94
3.1.12 Summary .....	95
<b>CHAPTER 4 DIAGNOSTICS AND DECISION RULES.....</b>	<b>96</b>
4.1.1 Introduction .....	96
4.1.2 Definition of Diagnosis .....	97
4.1.3 The Reference Standard.....	97
4.1.4 The Measurement of Outcomes.....	98
4.1.5 Normality.....	99
4.1.6 The Diagnosis or ‘Not the Diagnosis’ .....	100
4.1.7 Diagnosis Within Emergency Medicine .....	101
4.1.8 Measuring Diagnostic Utility .....	102
4.1.9 Reporting of Significance of Findings .....	105
4.1.10 Models of Decision Making .....	106
4.1.11 Methodology.....	108
4.1.12 The Risks of Error .....	111
4.1.13 The ‘Art’ of Medicine.....	112

4.1.14 Established Clinical Decision Rules.....	112
4.1.15 Mathematical Techniques to Derive the CDR.....	113
4.1.16 Validation of the CDR .....	114
4.1.17 Application to this Study.....	115
<b>CHAPTER 5 META-ANALYSIS OF DIAGNOSTIC STUDIES.....</b>	<b>117</b>
5.1.1 Introduction .....	117
5.1.2 Threshold Effect .....	118
5.1.3 Heterogeneity.....	119
5.1.4 Random Effects Model.....	121
5.1.5 Methods of Meta-analysis of Diagnostic Studies.....	122
5.1.6 Bivariate Analysis .....	125
5.1.7 Hierarchical Summary ROC Curve Method .....	126
5.1.8 Likelihood Ratios .....	126
5.1.9 Selected Methods of Meta-analysis.....	128
5.1.10 Methods Used .....	131
5.1.11 Summary.....	131
<b>CHAPTER 6 SYSTEMATIC REVIEW AND META-ANALYSIS.....</b>	<b>132</b>
6.1.1 Scope.....	132
6.1.2 Literature Search .....	133
6.1.3 Risk of Bias.....	135
6.1.4 Inclusion and Exclusion Criteria .....	139
6.1.5 The Reference Standard.....	139
6.1.6 Methods.....	141
6.1.7 Extraction of Information.....	141
6.1.8 Assessment of Quality.....	142
6.1.9 Data Presentation .....	143
6.1.10 Selection of Variables for Meta-analysis.....	143
6.1.11 Included Studies .....	144
6.1.12 Assessment of Quality.....	154
6.1.13 Excluded Studies .....	156
<b>Systematic Review .....</b>	<b>163</b>
<b>6.2 Variables with Little Diagnostic Value .....</b>	<b>164</b>
6.2.1 Absence of History of Asthma .....	164
6.2.2 Presence of an Abnormal Electrocardiogram .....	165
6.2.3 History of Smoking Tobacco .....	167
6.2.4 Hypertension .....	168

6.2.5 History of a cough .....	170
6.2.6 History of Orthopnoea .....	172
6.2.7 Absence of Fever .....	174
6.2.8 History of Nocturnal Cough.....	175
6.2.9 The Use of Diuretics.....	177
6.2.10 Use of ACE Inhibitors.....	178
6.2.11 Use of ARB Medication.....	180
6.2.12 Presence of Peripheral Oedema .....	181
6.2.13 Presence of Dyslipidaemia .....	183
6.2.14 The Presence of Rales.....	183
6.2.15 Presence of Wheeze .....	186
6.2.16 Patient Taking Beta Antagonist Medication.....	188
6.2.17 History of Chronic Obstructive Pulmonary Disease .....	191
6.2.18 History of Ischaemic Heart Disease .....	194
<b>6.3 Variables with Limited Applicability .....</b>	<b>195</b>
6.3.1 Impedance Cardiography .....	196
6.3.2 Bioelectrical Impedance Analysis.....	196
6.3.3 Acoustic Third Heart Sound .....	199
6.3.4 Mid-regional pro-Atrial Natriuretic Peptide.....	200
6.3.5 Use of Ultrasound .....	201
<b>6.4 Variables with High Specificity .....</b>	<b>203</b>
6.4.1 History of Paroxysmal Nocturnal Dyspnoea.....	204
6.4.2 Presence of diabetes.....	206
6.4.3 Patients with Atrial Fibrillation.....	208
6.4.4 History of Renal Impairment.....	210
6.4.5 Presence of a raised Jugular Venous Pressure.....	212
6.4.6 Presence of a Cardiac Murmur.....	215
6.4.7 Troponin Levels.....	218
6.4.8 Hepato-jugular Reflux.....	221
6.4.9 Presence of a Third Heart Sound .....	223
6.4.10 Presence of a Fourth Heart Sound.....	226
<b>6.5 Variables with Good Diagnostic Value .....</b>	<b>228</b>
6.5.1 History of Heart Failure .....	228
6.5.2 Clinical Opinion.....	231
6.5.3 Presence of Cardiomegaly on the Chest X-ray .....	234
6.5.4 Oedema on Chest X-ray.....	237
6.5.5 Cephalisation of Vessels on Chest X-ray .....	239

6.5.6 Brain Natriuretic Peptide 100 pg/ml.....	243
6.5.7 Brain Natriuretic Peptide 400 pg/ml.....	247
6.5.8 'N' Terminal proBNP (NT-proBNP) .....	249
<b>6.6 Discussion.....</b>	<b>254</b>
6.6.1 Summary of Findings .....	255
6.6.2 Limitations of the Systematic Review.....	256
6.6.3 Limitations of Brain Natriuretic Peptide.....	259
6.6.4 Other Reviews of the Literature.....	262
6.6.5 Application of Study Findings .....	265
6.6.6 The Need For Original Research .....	266
<b>CHAPTER 7 ORIGINAL DIAGNOSTIC STUDY .....</b>	<b>267</b>
<b>7.1 Introduction.....</b>	<b>267</b>
7.1.1 Research Background .....	267
7.1.2 Considerations for this Study .....	267
<b>7.2 Methods .....</b>	<b>268</b>
7.2.1 Setting .....	268
7.2.2 Inclusion/Exclusion criteria .....	269
7.2.3 Informed Consent .....	271
7.2.4 Blood Sampling.....	271
7.2.5 Recording of Patient Data .....	272
7.2.6 Echocardiography.....	273
7.2.7 Brain Natriuretic Peptide .....	274
7.2.8 'N' Terminal pro-Brain Natriuretic Peptide .....	276
7.2.9 High Sensitivity Troponin 'T' .....	277
7.2.10 The Reference Standard.....	277
<b>7.3 Statistical Methods .....</b>	<b>278</b>
7.3.1 Sample Size .....	278
7.3.2 Estimation of Sample Size .....	279
<b>7.4 Selection of Variables.....</b>	<b>279</b>
7.4.1 Random Forests® .....	281
7.4.2 Logistic Regression .....	283
7.4.3 Wald Statistic .....	285
<b>7.5 Creation of the Decision Rule .....</b>	<b>285</b>
7.5.1 Multivariable logistic regression analysis .....	286

7.5.2	Limitations of Logistic Regression .....	287
7.5.3	Assessing the logistic regression model .....	287
7.5.4	Summary of CART analysis.....	288
7.5.5	Growing Methods.....	289
7.5.6	Naïve Bayes Analysis.....	292
7.5.7	Summary.....	299
<b>CHAPTER 8 RESULTS AND ANALYSIS .....</b>		<b>300</b>
<b>8.1</b>	<b>Patient Characteristics and Classification .....</b>	<b>300</b>
8.1.1	Echocardiograms .....	310
<b>8.2</b>	<b>Examination of Data .....</b>	<b>311</b>
8.2.1	Random Forest Variables .....	311
8.2.2	Univariate Analysis .....	314
8.2.3	Correlation Analysis .....	317
8.2.4	Scoring System Derived from the Logistic Regression Equation .....	331
8.2.5	Classification and Regression Tree Analysis .....	333
8.2.6	Manual Classification Tree.....	335
8.2.7	Naïve Bayes Analysis.....	337
8.2.8	Application of the Framingham Criteria.....	344
8.2.9	Performance of Individual Variables .....	348
8.2.10	Summary.....	364
<b>CHAPTER 9 DISCUSSION AND CONCLUSIONS .....</b>		<b>366</b>
9.1.1	Introduction .....	366
9.1.2	Findings of the Thesis .....	367
9.1.3	The Implications of These Findings.....	369
9.1.4	Reliability and validity of the selected variables .....	370
9.1.5	Strengths and Weakness of the Systematic Review .....	370
9.1.6	Meta-analysis of the Selected Variables.....	371
9.1.7	Strengths of the Original Diagnostic Study.....	371
9.1.8	Weaknesses of the Original Diagnostic Study .....	372
9.1.9	Development of the Clinical Decision Rule.....	373
9.1.10	Recommendations for Further Research .....	374
9.1.11	Conclusion.....	375
<b>CHAPTER 10 APPENDICES.....</b>		<b>376</b>
<b>10.1</b>	<b>Appendix I - Literature Search .....</b>	<b>376</b>

10.1.1 EMBASE Search.....	376
10.1.2 Ovid MEDLINE Search.....	377
<b>10.2 Appendix II - QUADAS questions .....</b>	<b>379</b>
<b>10.3 Appendix III – Ethical Approval .....</b>	<b>380</b>
<b>10.4 Appendix IV – Consent Form.....</b>	<b>384</b>
<b>10.5 Appendix V – Clinical Details Form .....</b>	<b>386</b>
<b>10.6 Appendix VI – Characteristics of the Natriuretic Peptides .....</b>	<b>388</b>
<b>10.7 Appendix VI – Summary of Included Studies .....</b>	<b>393</b>
<b>10.8 Appendix VII - References.....</b>	<b>426</b>

LIST OF FIGURES

FIGURE 2-1 ECHOCARDIOGRAPHY ..... 60

FIGURE 2-2 ECHOCARDIOGRAPHY DEMONSTRATING DIASTOLIC DYSFUNCTION  
..... 74

FIGURE 3-1 BRAIN NATRIURETIC PEPTIDE ..... 80

FIGURE 6-1 PRISMA DIAGRAM OF LITERATURE SEARCH ..... 134

FIGURE 6-2 STUDY SIZE VERSUS PREVALENCE FOR INCLUDED STUDIES ..... 145

FIGURE 6-3 FOREST PLOT OF HISTORY OF ASTHMA ..... 164

FIGURE 6-4 SUMMARY ROC CURVE FOR ABSENCE OF HISTORY OF ASTHMA  
..... 165

FIGURE 6-5 FOREST PLOT FOR ABNORMAL ELECTROCARDIOGRAM ..... 166

FIGURE 6-6 SUMMARY ROC CURVE FOR ABNORMAL ELECTROCARDIOGRAM  
..... 166

FIGURE 6-7 - HISTORY OF TOBACCO SMOKING ..... 167

FIGURE 6-8 SUMMARY ROC CURVE - HISTORY OF TOBACCO SMOKING..... 168

FIGURE 6-9 SENSITIVITY AND SPECIFICITY OF HYPERTENSION AS A PREDICTOR  
OF HEART FAILURE ..... 169

FIGURE 6-10 SUMMARY ROC CURVE - HISTORY OF HYPERTENSION..... 170

FIGURE 6-11 SENSITIVITY AND SPECIFICITY OF HISTORY OF COUGH..... 171

FIGURE 6-12 SUMMARY ROC CURVE – HISTORY OF COUGH..... 171

FIGURE 6-13 FOREST PLOT – PRESENCE OF ORTHOPNOEA..... 172

FIGURE 6-14 SUMMARY ROC CURVE – PRESENCE OF ORTHOPNOEA ..... 173

FIGURE 6-15 FOREST PLOT FOR ABSENCE OF FEVER ..... 175

FIGURE 6-16 SUMMARY ROC CURVE FOR ABSENCE OF FEVER ..... 175

FIGURE 6-17 FOREST PLOT OF HISTORY OF NOCTURNAL COUGH ..... 176

FIGURE 6-18 SUMMARY ROC CURVE FOR HISTORY OF NOCTURNAL COUGH  
..... 176

FIGURE 6-19 FOREST PLOT - PATIENTS USING DIURETICS ..... 177

FIGURE 6-20 SUMMARY ROC CURVE - PATIENTS USING DIURETICS ..... 178

FIGURE 6-21 FOREST PLOT FOR THE USE OF ACE INHIBITORS ..... 179

FIGURE 6-22 SUMMARY ROC CURVE FOR THE USE OF ACE INHIBITORS..... 180

FIGURE 6-23 FOREST PLOT FOR USE OF ARB MEDICATION ..... 181

FIGURE 6-24 SENSITIVITY AND SPECIFICITY - PRESENCE OF PERIPHERAL  
OEDEMA..... 181

FIGURE 6-25 SUMMARY ROC CURVE - PERIPHERAL OEDEMA .....	182
FIGURE 6-26 FOREST PLOT - HISTORY OF DYSLIPIDAEMIA.....	183
FIGURE 6-27 SENSITIVITY AND SPECIFICITY OF PRESENCE OF RALES.....	184
FIGURE 6-28 SUMMARY ROC CURVE - PRESENCE OF RALES PRESENCE OF RALES.....	184
FIGURE 6-29 SUMMARY ROC CURVE - PRESENCE OF RALES .....	186
FIGURE 6-30 SENSITIVITY AND SPECIFICITY OF ABSENCE OF WHEEZE.....	187
FIGURE 6-31 SUMMARY ROC CURVE - ABSENCE OF WHEEZE .....	188
FIGURE 6-32 FOREST PLOT FOR PATIENTS TAKING BETA-BLOCKERS .....	189
FIGURE 6-33 SUMMARY ROC CURVE FOR PATIENTS TAKING BETA-BLOCKERS .....	190
FIGURE 6-34 FOREST PLOT - HISTORY OF COPD.....	192
FIGURE 6-35 SUMMARY ROC CURVE NO HISTORY OF COPD.....	193
FIGURE 6-36 SENSITIVITY AND SPECIFICITY OF IHD AS A PREDICTOR OF HEART FAILURE .....	194
FIGURE 6-37 SUMMARY ROC CURVE – IHD .....	195
FIGURE 6-38 FOREST PLOT FOR IMPEDANCE CARDIOGRAPHY .....	196
FIGURE 6-39 FOREST PLOT FOR BIOELECTRICAL IMPEDANCE ANALYSIS .....	197
FIGURE 6-40 ROC CURVES FOR BIOELECTRICAL IMPEDANCE ANALYSIS.....	198
FIGURE 6-41 FOREST PLOT OF ACOUSTIC THIRD HEART SOUND .....	199
FIGURE 6-42 SUMMARY ROC CURVE FOR ACOUSTIC THIRD HEART SOUND	200
FIGURE 6-43 FOREST PLOTS FOR MR-PROANP.....	201
FIGURE 6-44 ROC CURVES FOR MR-PROANP .....	201
FIGURE 6-45 FOREST PLOT FOR PERFORMANCE OF ULTRASOUND.....	202
FIGURE 6-46 SUMMARY ROC CURVE FOR PERFORMANCE OF ULTRASOUND	203
FIGURE 6-47 SENSITIVITY AND SPECIFICITY PND .....	204
FIGURE 6-48 SUMMARY ROC CURVE PND.....	205
FIGURE 6-49 FOREST PLOT OF PRESENCE OF DIABETES.....	206
FIGURE 6-50 SUMMARY ROC CURVE OF PRESENCE OF DIABETES .....	207
FIGURE 6-51 FOREST PLOT FOR PRESENCE OF ATRIAL FIBRILLATION.....	208
FIGURE 6-52 SUMMARY ROC CURVE FOR THE PRESENCE OF ATRIAL FIBRILLATION .....	209
FIGURE 6-53 FOREST PLOT - RENAL IMPAIRMENT .....	210
FIGURE 6-54 SUMMARY ROC CURVE - RENAL IMPAIRMENT.....	211

FIGURE 6-55 FOREST PLOT FOR PRESENCE OF RAISED JVP .....	213
FIGURE 6-56 SUMMARY ROC CURVE FOR PRESENCE OF RAISED JVP .....	214
FIGURE 6-57 FOREST PLOT FOR PRESENCE OF HEART MURMUR.....	216
FIGURE 6-58 SUMMARY ROC CURVE FOR THE PRESENCE OF A HEART MURMUR .....	216
FIGURE 6-59 SUMMARY ROC CURVE HEART MURMUR WITH BIVARIATE META- ANALYSIS DATA.....	218
FIGURE 6-61 SUMMARY ROC CURVE FOR TROPONIN RISE .....	220
FIGURE 6-62 FOREST PLOT FOR HEPATO-JUGULAR REFLUX.....	221
FIGURE 6-63 SUMMARY ROC CURVE FOR HEPATO-JUGULAR REFLUX .....	222
FIGURE 6-64 FOREST PLOT - PRESENCE OF THIRD HEART SOUND.....	223
FIGURE 6-65 SUMMARY ROC CURVE - PRESENCE OF THIRD HEART SOUND	224
FIGURE 6-66 FOREST PLOT FOR PRESENCE OF FOURTH HEART SOUND.....	226
FIGURE 6-67 SUMMARY ROC CURVE FOR THE PRESENCE OF A FOURTH HEART SOUND.....	227
FIGURE 6-68 SENSITIVITY AND SPECIFICITY OF A HISTORY OF HEART FAILURE .....	229
FIGURE 6-69 SUMMARY ROC CURVE - HISTORY OF HEART FAILURE .....	229
FIGURE 6-70 FOREST PLOT OF CLINICAL OPINION .....	232
FIGURE 6-71 SUMMARY ROC CURVE - CLINICAL OPINION.....	232
FIGURE 6-72 SENSITIVITY AND SPECIFICITY PRESENCE OF CARDIOMEGALY .	234
FIGURE 6-73 SUMMARY ROC CURVE - PRESENCE OF CARDIOMEGALY .....	235
FIGURE 6-74 PRESENCE OF PULMONARY OEDEMA ON CHEST X-RAY .....	238
FIGURE 6-75 FOREST PLOT - PRESENCE OF PULMONARY OEDEMA ON CXR	238
FIGURE 6-76 SENSITIVITY AND SPECIFICITY OF CEPHALISATION OF THE VESSELS ON CXR.....	240
FIGURE 6-77 SUMMARY ROC CURVE FOR CEPHALISATION OF VESSELS .....	241
FIGURE 6-78 SUMMARY ROC CURVE WITH BIVARIATE DATA CEPHALISATION VESSELS .....	243
FIGURE 6-80 SUMMARY ROC CURVE - BNP 100 PG/ML .....	244
FIGURE 6-81 SUMMARY ROC CURVE FOR BNP 100PG/ML .....	245
FIGURE 6-82 - PREVALENCE OF HEART FAILURE VERSUS SPECIFICITY FOR BNP STUDIES.....	246
FIGURE 6-83 FOREST PLOT FOR BNP WITH 400 PG/ML CUT-OFF .....	247

FIGURE -6-84 SUMMARY ROC CURVE - BNP 400 PG/ML .....	248
FIGURE 6-85 FOREST PLOT USING NT-PROBNP CUT-OFF VALUE OF 300 PG/ML .....	250
FIGURE 6-86 SUMMARY ROC CURVE FOR NT-PROBNP WITH CUT-OFF VALUE OF 300 PG/ML .....	251
FIGURE 6-87 FOREST PLOT USING NT-PROBNP WITH CUT-OFF VALUE OF 2000 PG/ML .....	252
FIGURE 6-88 SUMMARY ROC CURVE USING NT-PROBNP WITH CUT-OFF VALUE OF 2000 PG/ML .....	253
FIGURE 6-89 ROC CURVE AND DATA FOR BNP TAKEN FROM MAISEL 2002	260
FIGURE 6-90 SUMMARY ROC CURVE COMPARING BNP 100 PG/ML DATA AGAINST NT-PROBNP 300 PG/ML DATA .....	262
FIGURE 7-1 BINARY OUTCOME ASSOCIATED WITH A CONTINUOUS VARIABLE..	284
FIGURE 7-2 PROBABILITY OF A BINARY OUTCOME ASSOCIATED WITH A CONTINUOUS VARIABLE .....	284
FIGURE 7-3 EXAMPLE USING A NOMOGRAM FOR NAIVE BAYES' ANALYSIS ...	295
FIGURE 8-1 REFERENCE STANDARD AND DISCHARGE DIAGNOSES .....	303
FIGURE 8-2 - OVERALL ERROR RATE FOR RANDOM FOREST MODEL .....	314
FIGURE 8-3 - ROC CURVE LOGISTIC REGRESSION EQUATION .....	330
FIGURE 8-4 - AUC LOGISTIC REGRESSION .....	331
FIGURE 8-5 LR MODEL COMPARED AGAINST SCORING SYSTEM .....	333
FIGURE 8-6 EXAMPLE OF MANUAL DECISION TREE CREATED USING ORANGE SOFTWARE (DEMŠAR ET AL., 2013).....	337
FIGURE 8-7 ROC CURVE FOR NAIVE BAYES ANALYSIS OF VARIABLES SIGNIFICANT ON UNIVARIATE ANALYSIS.....	339
FIGURE 8-8 ROC CURVE FOR BAYES ANALYSIS USING FIRST TEN VARIABLES FROM RANDOMFOREST ANALYSIS .....	340
FIGURE 8-9 NAIVE BAYES ANALYSIS NOMOGRAM APPLIED TO PATIENT 1 RESULTS .....	341
FIGURE 8-10 ROC CURVE FOR VARIABLES SELECTED BY LOGISTIC REGRESSION WHEN NAIVE BAYES ANALYSIS PERFORMED USING 'MISS- ONE-OUT' ANALYSIS.....	342
FIGURE 8-11 RECEIVER OPERATING CHARACTERISTIC (ROC) CURVE FOR CARDIOTHORACIC RATIO FOR POSSIBLE/DEFINITE HEART FAILURE .....	349

FIGURE 8-12 AUC FOR CT RATIO.....	350
FIGURE 8-13 RECEIVER OPERATING CHARACTERISTIC CURVE (ROC) BNP FOR DEFINITE / PROBABLE HEART FAILURE.....	352
FIGURE 8-14 – RECEIVER OPERATING CHARACTERISTIC CURVE FOR UREA FOR DIAGNOSIS OF HF .....	354
FIGURE 8-15 AREA UNDER THE CURVE FOR UREA.....	355
FIGURE 8-16- RECEIVER OPERATING CHARACTERISTIC CURVE FOR NT- PROBNP .....	357
FIGURE 8-17 STARD DIAGRAM FOR PATIENTS WITH VARIABLES SPECIFIC FOR THE DIAGNOSIS OF HEART FAILURE .....	361
FIGURE 8-18 STARD DIAGRAM FOR PATIENTS WITH VARIABLES SPECIFIC FOR THE DIAGNOSIS OF NOT HEART FAILURE.....	364
FIGURE 10-1 FREQUENCY HISTOGRAM OF BNP .....	388
FIGURE 10-2 FREQUENCY HISTOGRAM OF LOG BNP .....	389
FIGURE 10-3 NORMAL P-P PLOT FOR LOG BNP .....	390
FIGURE 10-4 BNP VALUES BY SEX OF PATIENTS .....	391
FIGURE 10-5 NT-PROBNP VALUES BY PATIENT SEX .....	392

## LIST OF TABLES

TABLE 2-1 LIKELIHOOD RATIOS OF DIAGNOSTIC FINDINGS FROM FONSECA ET AL (2004) .....	46
TABLE 2-2 RESULTS FROM SR & META-ANALYSIS FROM WANG ET AL (2005). 47	
TABLE 2-3 RESULTS OF META-ANALYSIS MANT (2010).....	48
TABLE 2-4 RESULTS OF META-ANALYSIS MADHOK (2008) .....	49
TABLE 2-5 DIAGNOSTIC VALUE OF CLINICAL FINDINGS, KNUDSEN (2004) .....	53
TABLE 2-6 ABNORMAL ECHOCARDIOGRAM FINDINGS.....	63
TABLE 2-7 FRAMINGHAM CRITERIA FOR HEART FAILURE .....	67
TABLE 2-8 KILLIP CLASSIFICATION OF SEVERITY OF HEART FAILURE .....	68
TABLE 2-9 BOSTON CRITERIA FOR HEART FAILURE.....	69
TABLE 2-10 NYHA CLASSIFICATION OF HEART FAILURE .....	72
TABLE 3-1 SUMMARY OF EFFECTS OF BNP .....	84
TABLE 3-2 SUMMARY OF COMMERCIALY AVAILABLE ASSAYS (BASED ON TABLE FROM SILVER ET AL. (2002)).....	87
TABLE 3-3 FACTORS AFFECTING NATRIURETIC PEPTIDE LEVELS.....	90
TABLE 3-4 - CHARACTERISTICS OF BNP / NT-PROBNP.....	91
TABLE 6-1 DIFFERENT FORMS OF BIAS .....	137
TABLE 6-2 INCLUSION CRITERIA.....	139
TABLE 6-3 EXCLUSION CRITERIA .....	139
TABLE 6-4 – INCLUDED PAPERS .....	146
TABLE 6-5 QUALITY ASSESSMENT OF INCLUDED PAPERS .....	154
TABLE 6-6 EXCLUDED PAPERS .....	156
TABLE 6-7 - EXAMPLE DATA FOR THIRD HEART SOUND -1 .....	225
TABLE 6-8 EXAMPLE DATA FOR THIRD HEART SOUND - 2 .....	225
TABLE 6-9 SUMMARY TABLE - DIAGNOSTIC UTILITY OF SELECTED VARIABLES .....	255
TABLE 7-1 INCLUSION / EXCLUSION CRITERIA .....	270
TABLE 8-1 - PATIENT CHARACTERISTICS AND CLASSIFICATION.....	301
TABLE 8-2 REPORTED SYMPTOMS .....	304
TABLE 8-3 – PATIENT HISTORICAL FEATURES.....	305
TABLE 8-4 CURRENT MEDICATIONS .....	307
TABLE 8-5 CLINICAL SIGNS AND OBSERVATIONS.....	308
TABLE 8-6 RESULTS OF INVESTIGATIONS.....	309

TABLE 8-7 CLINICAL OPINION .....	310
TABLE 8-8 VARIABLES RANKED BY RANDOMFOREST .....	313
TABLE 8-9 - RESULTS OF UNIVARIATE ANALYSIS .....	315
TABLE 8-10 - CORRELATIONS OF SELECTED CONTINUOUS VARIABLES.....	319
TABLE 8-11 - THE LOGISTIC REGRESSION MODEL.....	328
TABLE 8-12 - PROBABILITY OF HEART FAILURE FOR FIVE PATIENTS CALCULATED USING THE LOGISTIC REGRESSION EQUATION .....	330
TABLE 8-13 COORDINATES OF ROC CURVE FOR LR EQUATION.....	331
TABLE 8-14 - SCORING SYSTEM DERIVED FROM LOGISTIC REGRESSION EQUATION .....	332
TABLE 8-15 - DIAGNOSTIC PERFORMANCE OF MANUAL DECISION TREE.....	336
TABLE 8-16 RESULTS OF DECISION TOOL ON VALIDATION GROUP.....	343
TABLE 8-17 FRAMINGHAM CRITERIA FOR HEART FAILURE.....	345
TABLE 8-18 APPLICATION OF FRAMINGHAM CRITERIA TO INCLUDED PATIENTS .....	346
TABLE 8-19 BOSTON CRITERIA FOR HEART FAILURE .....	347
TABLE 8-20 APPLICATION OF BOSTON CRITERIA TO STUDY PATIENTS .....	348
TABLE 8-21 COORDINATES FOR CARDIOTHORACIC RATIO ROC CURVE.....	350
TABLE 8-22 - DIAGNOSTIC PERFORMANCE OF BNP .....	352
TABLE 8-23 COORDINATES FOR ROC CURVE FOR BNP .....	352
TABLE 8-24 - DIAGNOSTIC UTILITY OF UREA AS A SINGLE VARIABLE.....	353
TABLE 8-25 COORDINATES FOR ROC CURVE FOR UREA.....	355
TABLE 8-26 - DIAGNOSTIC VALUE OF NT-PROBNP .....	357
TABLE 8-27 COORDINATES FOR ROC CURVE FOR NT-PROBNP .....	358
TABLE 8-28 - SUMMARY TABLE FOR INDIVIDUAL VARIABLES WITH HIGH SPECIFICITY FOR HEART FAILURE.....	359
TABLE 8-29 COMBINING SPECIFIC VARIABLES (NUMBER OF PATIENTS WITH BOTH VARIABLES POSITIVE / NUMBER OF PATIENTS DIAGNOSED AS NOT HAVING HEART FAILURE) .....	360
TABLE 8-30 SUMMARY OF VARIABLES WITH HIGH SENSITIVITY FOR THE ABSENCE OF HEART FAILURE.....	362
TABLE 8-31 COMBINING VARIABLES SENSITIVE FOR RULING OUT HEART FAILURE (NUMBER OF PATIENTS WITH BOTH VARIABLES POSITIVE / NUMBER OF PATIENTS DIAGNOSED WITH HEART FAILURE) .....	363

## *Abbreviations*

aa –	Amino acid
ACC –	The American College of Cardiology
ACE –	Angiotensin Converting Enzyme
AHA -	American Heart Association.
AMI -	Acute myocardial Infarction
ANF -	Atrial Natriuretic Factor
ANP -	Atrial Natriuretic Peptide
AP -	Antero-posterior
ARB -	Angiotensin Receptor Blocker
AUC -	Area Under the Curve
B Blocker -	Beta Receptor Antagonist
BHF -	British Heart Foundation
BMI -	Body Mass Index
BNP -	Brain Natriuretic Peptide
BP -	Blood pressure
BPM -	Beats per minute
Ca -	Calcium
CABG -	Coronary Arterial By-pass Graft
CDR -	Clinical Decision Rule
CI -	Confidence Interval
CO2 -	Carbon dioxide
COAD -	Chronic Obstructive Airway Disease
COPD -	Chronic obstructive pulmonary disease
Creps -	Crepitations
CRT-	Classification and Regression Trees
CTR -	Cardio-thoracic Ratio
DBP -	Diastolic Blood Pressure
DOR -	Diagnostic odds ratio
ECG -	Electrocardiogram
Echo -	Echocardiogram
ED -	Emergency Department

ESC -	European Society of Cardiology
Excl. -	Excluded
Exp -	Exponential function
FN -	False Negative
FP -	False Positive
GFR -	Glomerular filtration rate
Hb -	Haemoglobin
HF -	Heart Failure
Hx -	History
HS-TNT -	High Sensitivity Troponin-T
IHD -	Ischaemic Heart Disease
JVP -	Jugular Venous Pulse or Pressure
LA -	Left Atrium
LAD -	Left axis deviation
LBBB -	Left bundle branch block
Ln -	natural logarithm
LOS -	length of stay
LR -	Likelihood ratio, also logistic regression
LV -	Left ventricle
LVD -	Left ventricular dysfunction
LVF -	Left ventricular failure
LVSD -	Left ventricular Systolic Dysfunction
MR-proANP -	Mid-range pro-atrial natriuretic peptide
NICE -	National Institute for Health and Clinical Excellence
NSAID -	Non-steroidal Anti-inflammatory Drugs
NT-proBNP -	Amino-terminal pro-brain natriuretic peptide
NYHA -	New York Heart Association
OR-	Odds Ratio
PA -	Postero-anterior
PND -	Paroxysmal Nocturnal Dyspnoea
PRISMA -	Preferred Reporting Items for Systematic Reviews and Meta-analyses
QUADAS -	Quality Assessment Tool for Diagnostic Accuracy studies

RA -	Right Atrium
RBBB -	Right Bundle Branch Block
ROC -	Receiver Operating Characteristic
RV -	Right ventricle
Sats -	Oxygen saturation
SBP -	Systolic Blood Pressure
SE -	Standard Error
Sig. -	Significance
STARD -	STAndards for the Reporting of Diagnostic accuracy studies
TN -	True Negative
TP -	True Positive
ULD -	Upper lobe diversion
Ur -	Urea
Vs -	Versus
XR -	X-ray

## Abstract

### Background

Acute, decompensated heart failure is a serious and common presentation in patients attending Emergency Departments. Diagnosis of this condition in this environment can be challenging.

### Main Objective:

- To assess the diagnostic utility of variables related to the diagnosis of decompensated heart failure in the Emergency Department setting.
- To create a Clinical Decision Rule to facilitate the diagnosis of decompensated heart failure in the Emergency Department setting.

### Methods:

- A systematic review and meta-analysis of literature related to the diagnosis of acute heart failure in the Emergency Department was performed to assess and rank the diagnostic utility of all potential predictor variables. Bivariate meta-analysis was performed where appropriate to provide summary statistics for variable utility.
- A diagnostic study performed in a single, urban Emergency Department to allow multivariable analysis of the data to derive a Clinical Decision Rule. Logistic regression, Random Forest analysis and CART analysis were used in the analysis.

### Results:

44 papers were included in the systematic review providing data on 41 potential diagnostic variables. A history of heart failure, clinical opinion of heart failure, natriuretic peptide levels and the findings of cardiomegaly, pulmonary oedema or cephalisation of vessels on the chest x-ray had the greatest diagnostic utility.

- 105 patients were recruited in the diagnostic study and 62 potential variables were assessed against a reference standard of two cardiologists opinions. Cardiothoracic ratio, natriuretic peptide levels, the presence of pulmonary oedema, clinical impression of heart failure and urea levels were found to be the best performing variables.
- Two or more of these variables, using selected cut-off values for continuous variables, had a specificity of nearly 100% for definite or

possible heart failure. Different cut-off values could be used with these variables to provide high sensitivity for this condition.

**Conclusion:** The clinical and investigative findings with the greatest diagnostic utility for the acutely dyspnoeic patient with heart failure have been identified and can be used to rule in heart failure.

## Declaration

No part of this thesis has been submitted in support of an application for another degree or qualification of this or any other university or other institute of learning.

## Acknowledgements

Professor Kevin Mackway-Jones, Consultant in Emergency Medicine at Manchester Royal Infirmary was the instigator of this research project. In addition, Professor Mackway-Jones was the second appraiser of the literature search and a second reader for several of selected papers. I am indebted to his support, advice, encouragement and, at times cajoling, in all stages of this research project.

I would also like to thank Dr. Francis Fatoye, Dr. Chris Wibberley and Dr. Abebaw Yohannes from the Manchester Metropolitan University for their support, time, patience and learned wisdom in guiding me through this process.

I am also grateful to Dr. Janos Baombe, Dr. Tom Bartram, Dr. Richard Body, Professor Simon Carley, Dr. Bernard Föex and Dr. Daniel Horner who acted as second readers and assessors for papers selected for the systematic review.

Katharine Wylie, Senior Informationist at Manchester Royal Infirmary was invaluable as a source of guidance and advice in my forays into the world of medical literature.

This research project would not have been possible without the assistance and cooperation of the medical, nursing, administrative and secretarial staff in the Emergency Department and I would like to express my gratitude to the entire department for this.

## Dedication

This thesis is dedicated to my beloved wife, Claire, and my son, James, who have provided me with constant support, tolerance and love over the many years that this manuscript has taken to complete.

## Chapter 1 Background

### 1.1 Heart Failure

#### 1.1.1 The Burden of Heart Failure

Heart failure is a condition affecting the lives of millions of people throughout the world. It is currently incurable and it shortens and impinges on the lives of those afflicted and their carers, impairing their ability to perform their daily activities due to the symptoms of breathlessness and fatigue. There is a pattern of gradual decline with terrifying acute exacerbations. In the patient with pulmonary symptoms, they are unable to breathe adequately due to the collection of fluid within their lungs. It is possible to provide treatment for patients during these episodes of acute decompensation, to relieve symptoms and improve survival, but this relies on the accurate and early diagnosis of this condition. Acute decompensation of cardiac failure causes severe breathlessness but so do many other conditions; the treatments for the other conditions may be radically different. Rapid and accurate diagnosis allows the correct treatment can be started as early as possible.

#### 1.1.2 The Definition of Heart Failure

There is no universally agreed definition for the diagnosis of heart failure; a fact that makes it difficult to compare epidemiological, diagnostic and intervention research in this area. The diagnosis is made on clinical grounds based on the interpretation of numerous historical, clinical and investigative findings. There is general agreement that it is a syndrome rather than a single entity and there are numerous underlying aetiologies and clinical symptom presentations.

Heart failure may be classified as right heart failure causing peripheral symptoms and left heart failure causing pulmonary oedema. The term

'congestive' is added if there is a suggestion of oedema present due to vascular 'congestion'.

The lack of a clear definition of heart failure hampers the research into this subject. For therapeutic trials there are few objective criteria to define the target cohort of patients are recruited to ensure that this is consistent, and the measured outcomes may be controversial, aside from all-cause-mortality (Yasue et al., 1994; Allen et al., 2009). Even the assessment of dyspnoea is subjective with no universally agreed and validated instrument, and it may be hard to separate symptoms due to a pathological process from those due to the physiological process of ageing (Eriksson, Svärdsudd, et al., 1987; Baxter, 2004).

Even with an individual patient, the degree of decompensation is not constant with variation in the clinical and diagnostic findings over a period of time. The diagnostic findings from clinical examination and investigations can also be subjectively interpreted. For these reasons, it can be difficult to perform research comparing the effectiveness of treatments or even diagnostic tests for this condition.

The Oxford Textbook of Medicine defines heart failure as, "a clinical syndrome that results from any structural or functional cardiac disorder that reduces the ability of the heart to function as a pump" (Cowie & Chandrasekaran, 2011;1).

The NHS National Institute for Health and Clinical Evidence (NICE) defines the condition as the following: "Heart failure is a complex clinical syndrome of symptoms and signs that suggest impairment of the heart as a pump supporting physiological circulation. It is caused by structural or functional abnormalities of the heart. The demonstration of objective evidence of these cardiac abnormalities is necessary for the diagnosis of heart failure to be made" (NICE, 2010;19).

The American Heart Association (AHA) describes heart failure as a complex condition resulting from any structural or functional cardiac disorder that impairs the ability of the ventricle to fill with or eject blood(AHA, 2002). The European Society of Cardiology (ESC) defines heart failure as an abnormality of cardiac structure or function leading to failure of the heart to deliver oxygen at a rate commensurate with the requirements of metabolising tissues(McMurray et al., 2012).

The definition of acute, decompensated heart failure used for the purposes of this study is where “the symptom of acute dyspnoea is caused in whole or mainly by the impaired ability of the heart to function as a pump sufficient to support a physiological circulation whether or not it is accompanied by other symptoms or signs”. In the absence of any more closely defined parameters for this condition, the clinical impression of heart failure, based on the assessment of clinical signs, the symptoms expressed by the patient and their progress and response to treatment, as assessed by at least one cardiologist, has been accepted as a reasonable reference standard for the studies included in this research (Felker et al., 2003).

### **1.1.3 Demographics**

In Western society, improved survival from traditional causes of mortality such as infectious disease has improved general longevity to the point that many more people live long enough to develop ischaemic heart disease. There has also been increasing prevalence of some of the known risk factors for heart disease such as obesity and diabetes. These factors combined are thought to be responsible for the increasing incidence of heart failure in the UK. Overall, the number and proportion of elderly people in the community continues to rise, and advancing medical care has improved survival from cardiac events leading to more patients living with a damaged heart.

The prevalence of heart failure increases with increasing age (Eriksson, Svärdsudd, et al., 1987; Lloyd-Jones, 2002). In the UK, the British Heart Foundation estimates that 94,200 men and 105,400 women over the age of forty-five years have definite heart failure (Townsend et al., 2012).

In the US, current estimates are that 2.4% of the population has heart failure, with 12% of people over the age of 80 years affected by this condition (Roger et al., 2012). The estimated annual total cost of treating patients with heart failure in the US is around \$32 billion (Go et al., 2013).

Heart failure is generally described as more common in men than women in all ages, and the higher incidence of myocardial infarction in men has been suggested as a possible reason. (Tudor et al., 1997; Nieminen et al., 2006; BHF, 2008). Large registry studies such as AHERE in the US show a slightly higher incidence in females so this belief may be due to a bias in recruitment of males for heart failure studies (Adams et al., 2005). The overall lifetime risk of developing heart failure is thought to be twenty per cent for both men and women (Struthers, 2000; Lloyd-Jones, 2002; Heidenreich et al., 2013).

Data from the latest national heart failure audit for England and Wales showed that the mean age of patients on their first admission for heart failure in 2011 was 77.7 years, with a median age of 80.1 years. In patients up to 85 years of age, the majority were men (61.1%) while in patients over 85 years there were more women (57.9%) (Cleland et al., 2012).

38,973 men and 39,017 women over the age of 45 years were admitted to English hospitals with a main diagnosis of heart failure in 1999/2000. This represents a five per cent increase in the number of admissions for men and a four per cent increase in admissions for women over the previous ten years (Drazner et al., 2001; Gnani & Ellis, 2002). There are

significant costs associated with the increased level of morbidity in terms of society and direct financial cost to the NHS.

The National Heart Failure Audit (2010) estimates that providing services to patients with heart failure costs the NHS £625 million per year (NHSIC, 2010). Almost 90% of heart failure admissions are emergency admissions and overall they account for 5% of all emergency medical admissions. The figures for 2008-2009 were of 58,164 admissions with 740,697 bed days for patients with heart failure (NHSIC, 2010).

#### **1.1.4 The Consequences of Heart Failure**

The British Heart Foundation (BHF) suggests that the one-year survival for patients following initial diagnosis of heart failure is around 60% (Townsend et al., 2012). This is worse than the one-year survival rates for patients diagnosed with cancer of the breast or prostate and similar to the survival rate for patients with cancer of the colon (ONS, 2012). Data from the Framingham study shows five-year survival rates of 25% in males and 38% in females following the diagnosis of heart failure (Kannel, 1999; Fonseca et al., 2004). Other data from the United States suggests a 5-year mortality rate of 50% (Levy et al., 2002; Roger et al., 2004).

The overall in-patient mortality rate for patients admitted to hospital with heart failure is around fifteen per cent. This figure rises to twenty-three per cent for the patients over the age of eighty-five years (Henriksson et al., 1990; Nicol et al., 2008). In 2011/2012 the National Heart Failure Audit for England and Wales showed an 11.1% overall in-hospital mortality rate for patients admitted with a primary diagnosis of heart failure. The in-hospital mortality rate for men was 10.2% and for women was 12.1%. The mortality rates increased with age with a 2.5% mortality rate for patients aged less than forty-five years, compared with a 16.8% mortality rate for patients over 85 years. The overall one-year mortality for

this audit group was 26.2%, with a similar distribution for men and women (Cleland et al., 2012). This data was derived from the 24,724 patients on whom a mortality status could be obtained.

There was a significant difference in the in-patient mortality rates within this audit group depending on where patients were treated: the mortality rate for patients treated on cardiology wards was 7.8%; for general medical wards was 13.2%; and for patients admitted to other wards 17.4%. These differences persisted in the data for one-year survival (Cleland et al., 2012). Data from the ADHERE study, a multicentre, prospective registry for patients hospitalised with acute decompensated heart failure in the US showed an overall mortality of 4.2% for patients admitted between 2001 and 2003 (Abraham et al., 2005).

The diagnosis of heart failure is often associated with numerous hospital admissions and poor quality of life. Caring for a person with heart failure is also associated with psychological and physical problems for family members and caregivers (Pattenden et al., 2007). Patients with end-stage heart failure receive less health and palliative care services and have a poorer understanding of their condition compared with patients with lung cancer despite sharing a similar prognosis (Murray et al., 2002).

### **1.1.5 The Aetiology of Heart Failure**

There have been several studies examining the aetiology of heart failure and a wide variety of causes for the syndrome of heart failure have been found; the commonest cause of heart failure in the UK, Europe and the United States is coronary artery disease, either due to ischaemia or as a result of a myocardial infarction (Kannel, 1999; Zannad et al., 1999; McMurray & Stewart, 2000; Fox et al., 2001; He et al., 2001).

Other causes include cardiomyopathies and valvular heart disease. Valvular heart disease is less common now in Western Europe due to the

decreased incidence of Rheumatic heart disease although this remains a significant condition in other countries. Mitral incompetence and aortic stenosis are the commonest valvular defects leading to heart failure. Cardiomyopathies and congenital anomalies may also lead to heart failure. Cor pulmonale, where a primary disorder of the respiratory system (e.g. severe COPD) results in right ventricular hypertrophy and eventually progresses to right ventricular failure is less common but remains a significant cause. Abdelhafiz (2002) suggests that although there is a higher prevalence of heart failure in older patients, the underlying causes are similar.

Ho et al. (1993), examining the data from the Framingham study found hypertension and coronary heart disease to be the commonest underlying causes. A history of cigarette smoking, diabetes and changes of left ventricular hypertrophy on the ECG were found to be significant risk factors for the subsequent development of heart failure.

Cowie et al. (1999) looked at patients presenting with a new diagnosis of heart failure over the period of one year, in the area of Hillingdon, which had a population of 150,852. Of the 220 patients who were diagnosed with heart failure the commonest single aetiology was coronary heart disease. This made up 36% of the cases. Many of the patients also had a history of hypertension although this was not usually attributed as the main cause of heart failure. 31% of the patients with heart failure had atrial fibrillation at the time of presentation.

Zannad et al. (1999) examined a population in Northern France and found an overall incidence for heart failure of 225 patients per million. The underlying aetiology for the 499 patients diagnosed with heart failure, during the time of the study, was considered to be coronary heart disease in 46% of the patients while 15% of patients had the condition attributed to cardiomyopathy.

Fox et al. (2001), examined a population in southern London and found that there was an incidence of heart failure of 0.9 cases per 1,000 population. The majority of the patients who were diagnosed with heart failure had unknown aetiology (42%). The commonest identified causes were ischaemic heart disease (29%) and hypertension (9%).

He et al. (2001) describe the National Health and Nutritional Examination Survey (NHANES) epidemiological cohort study where 5545 men and 8098 women, with no known history of heart disease, were followed up for an average of nineteen years. 1388 people developed heart failure during this period. The commonest aetiology was ischaemic heart disease, accounting for 61.6% of the patients while 17.1% were attributed to cigarette smoking and 10.1% to hypertension.

Rudiger et al. (2005) looked at patients presenting with acute onset of dyspnoea at two hospitals, one in Zurich, Switzerland and the other in Helsinki, Finland. The diagnosis of heart failure was made on the basis of a history of heart failure, symptoms consistent with heart failure and a characteristic chest x-ray. These researchers found that ischaemic heart disease; dilated cardiomyopathy and valvular heart disease were the three commonest underlying aetiologies for the heart failure.

The national heart failure audit for England and Wales 2011-2012 found hypertension (54%) and ischaemic heart disease (46%) to be the commonest contributory causes of heart failure with 26% of admitted patients having a history of both. 31% of patients had a history of previous myocardial infarction and 22% had a history of valvular heart disease (Cleland et al., 2012).

The 2013 Annual Report from the American Heart Association stated that a history of hypertension or previous myocardial infarction were the risk factors most closely associated with the presence of heart failure (Go et al., 2013).

### 1.1.6 The Presentation of Heart Failure

Heart failure is usually associated with dyspnoea, fatigue, and fluid retention. Other symptoms may include nocturia, anorexia, abdominal bloating and discomfort, constipation, and cerebral symptoms such as confusion, dizziness, and memory impairment (Fromm et al., 1995; Warrell et al., 2010).

One of the cardinal symptoms of heart failure is breathlessness. This is usually in the form of dyspnoea on mild exertion but may also present as a cough, which may be nocturnal (Rudiger et al., 2005). Extreme fatigue, poor exercise tolerance, sleep disturbance and anorexia all combine to provide a very poor quality of life for patients with heart failure (Boyd et al., 2004).

The jugular venous pressure may be raised, reflecting the high filling pressure of the heart. This may also be related to an increase in total body water as a result of the physiological changes required to maintain cardiac filling. A gallop rhythm may be present on palpation of the pulse, and a third or fourth heart sound can be present due to cardiac strain and dilatation. Peripheral oedema may be present in gravity dependent areas; in severe cases the oedema may be present as far up as the sacrum (anosarca).

Paroxysmal nocturnal dyspnoea (PND) is another common complaint, which is likely to be caused by the natural circadian narrowing of airways during the night in combination with the redistribution of the fluid due to the change in physical position. Patients may also complain of cough or wheeze and may mention white, frothy sputum. The patient with acute decompensation is usually tachypnoeic and tachycardic (Fromm et al., 1995). In some cases, especially during severe decompensation, the patient may be normocardiac or even bradycardic.

The National Heart Failure Audit for England and Wales found that 40% of patients were in NYHA Class III at first admission, with breathlessness on mild exertion while 32 % were in Class IV with breathlessness at rest. Other patients were admitted due to peripheral oedema. The proportion of patients with Class IV symptoms in patients who were readmitted to hospital was higher at 78% and 52% of these patients had moderate or severe oedema (Cleland et al., 2012). In the large ADHERE registry study, recording patients admitted to hospital due to heart failure in the US, 87% of patients presented in NHYA Class III or IV (Gheorghide, 2005).

### **1.1.7 Differential Diagnosis**

Why is the diagnosis of heart failure so difficult to make?

Breathlessness, the subjective sensation of inadequate breathing, is a very common presentation to the Emergency Department. There are numerous causes for this symptom that may be pulmonary in origin, for example asthma, extra-pulmonary, for example anaemia or even psychological. DiagnosisPro, a website providing an online diagnosis tool has a list of 518 potential causes, including 110 due to specific agent poisoning (DiagnosisPro, 2014).

Even in the patient who has a confirmed diagnosis of heart failure, there are numerous aetiologies for developing this syndrome. Although there are associations between numerous conditions and the syndrome of heart failure the relationship between the cause and effect is not always clear. For example, patients may develop heart failure after a myocardial infarction but not every patient who has a myocardial infarction will develop heart failure.

The signs and symptoms associated with heart failure are wide-ranging and often open to subjective assessment; their severity, or even presence, varies between patients and even within the same patient at different times.

Pathological processes may occur simultaneously and it is not unusual for a patient to have a diagnosis of both Chronic Obstructive Pulmonary Disease (COPD) and heart failure. Although the main presenting symptom is the same, the underlying pathology and therefore the treatments are very different. Not only will incorrect treatment fail to benefit the patient, inappropriate medication may cause harm to the patient, especially in the patient who is already significantly compromised by their acute condition.

There is some evidence to suggest that improving the diagnostic process can have positive effects on patient care. Mueller et al. (2004) followed 452 patients who presented to their hospital with acute dyspnoea. The patients were randomised as to whether or not to have a Brain Natriuretic Peptide (BNP) level measured. The patient group who had a BNP level measured to aid the diagnosis of the attending clinician, were less likely to be admitted to hospital, had a lower length of stay and a non-significant trend towards a lower mortality rate.

### **1.1.8 Diagnosing in the Emergency Department**

The Emergency Department can be a challenging environment in which to assess the patient who may have acute heart failure: the patient is often too unwell to give a clear history of events; relatives may not be available to provide a corroborative account of events; the patient's hospital notes or results of previous investigations may not be available; there is a limited repertoire of investigations available in the Emergency Department setting; and even positive findings on examination or from investigation are not definitive of heart failure.

Most Emergency Departments have a limited array of tests that they will perform. The majority of tests are intended to rapidly aid the clinician in confirming or refuting a particular diagnosis. Patients requiring subtler, more time-consuming, more invasive or more specialised tests will either have these as following admission to hospital or on an elective basis as an out-patient. Specialised tests also require specialist interpretation in order to be of greatest utility. There has to be a pragmatic approach to produce a balance of appropriate and efficient investigation; it has been shown that routinely applying screening tests to attending patient's increases costs without demonstrating improvements in patient outcomes (Durbridge et al., 1976).

Diagnostic testing within the emergency department is designed to risk-manage the patient until there is an acceptably probability of disease presence, at which time appropriate treatment can be commenced, or an acceptably low risk of disease at which point other conditions can be considered or the patient reassured and discharged.

The time taken to make a diagnosis in a patient is limited, in the management of an acutely ill patient one would wish to commence the appropriate treatment as quickly as possible. In the UK setting, Emergency Department staff are required to see, diagnose, treat and either discharge or admit to an inpatient speciality, 95% of the attending patients within four hours of the patient's arrival (DOH, 2014). The restricted time frame limits the diagnostic processes that can be facilitated within this period. For the most gravely ill patients treatment must often be started on an empirical basis for likeliest diagnosis based on the immediately available information.

The limitations of working within this environment might be expected to produce less than perfect diagnostic accuracy and this is reflected in the available data. Collins, Lindsell, Peacock, Eckert, et al. (2006) showed that 14.3% of patients with an end-diagnosis of heart failure following a

period of hospital admission did not have heart failure diagnosed as the underlying cause within the Emergency Department.

### **1.1.9 The Pathophysiology of Heart Failure**

Heart failure exists when the cardiac pump fails to provide adequate circulation for the respiratory needs of the body and is usually primarily due to a problem of the cardiac pump. The dysfunction may be due to myocardial dysfunction, valvular dysfunction, an abnormality of the pericardium, an arrhythmia or a pre-load / after-load mismatch. Examples of extra-cardiac causes of failure are the increase in after-load due to a massive pulmonary embolism, increased pre-load due to increased volume intake or decrease in excretion, or the high-output states seen in thyrotoxicosis or sepsis.

The underlying pathology may be sudden in onset, for example a sudden decrease in function due to myocardial infarction or may develop over a long period of time in conditions such as persistent hypertension or a cardiomyopathy. The ability of the heart to compensate for some loss in function means that there is often a latent period of asymptomatic ventricular dysfunction.

The initial compensation occurs by allowing the left ventricular end diastolic volume to increase, which in turn, increases the cardiac output. This mechanism works by Starling's principal that the force of myocardial contraction is increased in response to stretch (Starling, 1896). This mechanism assists cardiac function initially but is often overwhelmed, and increasing the end diastolic volume further then leads to decreased efficiency of the heart function.

Hypertrophy of the myocardium decreases the load on individual cells but leads to increased distance between capillaries, impaired diffusion of substrates and increases in the cell volume. There are molecular-level

changes in protein synthesis including reversion to foetal iso-forms of proteins.

Cardiac function tends to diminish incrementally over time until a certain tipping point is reached and the patient becomes symptomatic. This progression is accompanied by changes within the myocardium and further activation of the neuro-hormonal and cytokine systems. It has been suggested that this progression is independent of physical haemodynamics and is dependent on the hormonal effects for this remodelling. This raises the possibility of a mechanism for drugs to prevent this remodelling (Yan et al., 2008).

Compensatory mechanisms for evolving heart failure involve the stimulation of the sympathetic nervous system. This utilises a variety of methods to improve cardiac output. It is part of the 'flight or fight' response and works most effectively as a short-term response to acute haemorrhage. Hormones such as adrenaline and noradrenaline are released, causing constriction of the blood vessels to increase their resistance and an increase in the force and rate of the heart contractions. The renin-angiotensin pathway is also activated, promoting salt and fluid retention by hormonal action on the kidney and other tissues. There is also increased excretion of vasodilatory molecules including natriuretic peptides, prostaglandins (PGE<sub>2</sub> and PGEI<sub>2</sub>) and nitric oxide (Fletcher & Thomas, 2001).

Heart failure can broadly be split into systolic and diastolic dysfunction. Systolic dysfunction is the aspect that is most often considered when using the term cardiac failure, and is somewhat easier to diagnose than diastolic failure. The failure results from the heart's inability to contract effectively, most commonly due to the results of ischaemic heart disease. The reduced coronary artery blood flow to the cardiac tissue either causes impaired contraction directly due to the reduced delivery of nutrients and oxygen or has caused an area of infarction at some point in the past where the cardiac muscle is permanently damaged. The

damaged area will not contract with the rest of the ventricle and there may even be paradoxical movement further diminishing the efficiency of the contraction (Dhalla et al., 1993).

Peripheral oedema is caused by an increase in cardiac preload due to backpressure secondary to pump failure. The increased orthostatic pressure drives fluid out of the vessels into the tissues, particularly in gravity-dependent areas.

### **1.1.10 Neuroendocrine response**

In response to the cardiac insufficiency there are several compensatory mechanisms. Mechanoreceptors sense arterial under-filling and are involved in the regulation of fluid volume. These receptors are present in the left ventricle, carotid sinus, aortic arch and renal afferent arterioles. A reduction in the stimulation of these receptors results in increased sympathetic activity, activation of the renin-angiotensin system, the release of arginine vasopressin (anti-diuretic hormone) and increased thirst sensation. The renin-angiotensin system increases the reabsorption of sodium by the renal system with the accompaniment of water. The angiotensin released is also a powerful vasoconstrictor.

The increase in sympathetic tone that results from baroreceptor stimulation results in increased myocardial contractility, tachycardia and arterial vasoconstriction, and therefore, increased cardiac after-load. It also stimulates veno-constriction resulting in increased preload. Increased sympathetic activity is beneficial in the short-term for patients with cardiac failure as a result of positive chronotropic and inotropic effects, but in the long-term the compensatory effects are overcome and damage to the cardiac muscle occurs, with remodelling of the myocardial tissue over time (Packer, 1995a).

Renal vasoconstriction, stimulation of the renin-angiotensin system and direct effects on the proximal convoluted tubule lead to renal sodium and water retention. Not only is the renin-angiotensin system activity increased in patients with heart failure, but also the action of aldosterone is more persistent in these patients than in normal controls. If aldosterone levels rise in an individual who does not have heart failure, after an initial increase in sodium retention the hormonal effect plateaus and sodium levels do not increase further even if a high aldosterone is maintained. This seems to result from an increased sodium concentration in the collecting ducts where the aldosterone causes its main action (Kjaer & Hesse, 2001). In patients with heart failure this tolerance of aldosterone does not seem to happen and there is continuous stimulation promoting sodium retention. These patients have a substantial diuresis in response to spironolactone. The use of spironolactone in patients with heart failure has been shown to substantially decrease mortality (Pitt et al., 1999).

Progressively deteriorating cardiac function is often associated with impaired renal function, described as cardio-renal syndrome. The exact mechanism of this relationship remains unknown but there is a correlation between decompensated heart failure and a decreased glomerular filtration rate (GFR). Impaired renal function in patients admitted to hospital with acute heart failure is associated with longer periods of hospital stay and a higher mortality (Ronco et al., 2008; Tang & Mullens, 2010)

A diagram representing the main neurological and hormonal factors involved in the regulation of the circulation is presented in Figure 1-1.



### 1.1.11 Remodelling

Both the structure and the composition of the cardiac walls change in patients with heart failure. There is a reduction in the number of myocytes, hypertrophy of the remaining myocytes and changes to the make up of the extracellular matrix with increased fibrosis. The changes to the myocardium as a result of the remodelling produce a ventricle that is larger, more spherical and has thinner walls; this results in a less efficient pump with a greater oxygen demand.

Norepinephrine, angiotensin II, endothelin, aldosterone and tumour necrosis factor have all been suggested as having a role in the progression of this condition. An increased level of epinephrine is thought to contribute to myocardial cell hypertrophy. Norepinephrine is toxic to cardiac cells causing calcium overload or apoptosis although the norepinephrine-induced cell death can be prevented by the use of beta-blockers (Mann, 1999).

Although these changes, termed remodelling, are agreed to be associated with heart failure the level to which they are a result of cardiac dysfunction and the degree to which they actually contribute to further impairment of cardiac function remains a subject of debate. It has been noted that the heart failure phenotype is very similar in all patients with heart failure despite the failure being caused by a variety of different pathologies (Mann, 1999). There are on-going attempts to synthesise pharmacological products that modify the remodelling of the left ventricular myocardium. Brain natriuretic peptide may also have some anti-fibrotic properties (Ogawa et al., 2001).

Although the pathological processes underlying the common concept of heart failure are complex and intertwined, this also presents opportunities in terms of therapeutic interventions and diagnostic modalities.

### 1.1.12 Current Guidelines

The American College of Cardiology Foundation / American Heart Association (ACC/AHA) guidelines were published in 2009 (Jessup et al., 2009). The guidelines are mainly directed towards the diagnosis and management of chronic heart failure although there is some guidance provided regarding patients who have been acutely admitted to hospital. The guidelines state that the diagnosis should be based on the signs and symptoms suggestive of heart failure but that natriuretic peptide levels may be considered if the diagnosis remains unclear. The authors stress in the text that the levels of natriuretic peptides should be interpreted with reference to the clinical picture and not in isolation.

NICE (2010) has published guidelines for diagnosis and management of chronic heart failure. NICE recommends referring patients with suspected heart failure who have a history of myocardial infarction for an urgent echocardiogram (within two weeks). For patients with no history of myocardial infarction, NICE recommends checking BNP or NT-proBNP. For those patients with BNP levels greater than 400 pg/ml (NT-proBNP > 2000 pg/ml) NICE recommend an echocardiogram within two weeks, and for patients with BNP levels between 100 and 400 pg/ml (NT-proBNP 400 – 2000 pg/ml) the recommendation is for an echocardiogram within six weeks. The authors state that it is very unlikely for a patient to have heart failure if there is no history of MI and natriuretic peptides levels are low. The guidelines also recommend obtaining an ECG and considering the following tests: chest x-ray; urea and electrolytes; estimated GFR; TFTs; LFTs; lipids; glucose; FBC; urinalysis; peak flow or spirometry.

The European Society of Cardiology (ESC) has also published guidelines for the diagnosis and management of heart failure (McMurray et al., 2012). These guidelines suggest that a chest x-ray and ECG be

performed initially and then the patient should either undergo echocardiography or have BNP (or NT-proBNP) levels checked. The guidelines state that heart failure will be the underlying cause in less than 2% of patients with both acute symptoms and a normal ECG. The ESC recommends that patients with a normal ECG and a BNP level <100 pg/ml should not be diagnosed with heart failure, while patients with either an abnormal ECG or a BNP value  $\geq 100$  pg/ml, should undergo echocardiography as further investigation. In the presence of shock or haemodynamic compromise, immediate echocardiography is recommended.

These institutional guidelines focus on the diagnosis of patients with chronic heart failure rather than the acute and possibly primary presentation of the patient in extremis. While echocardiography provides the closest thing that there is to a gold standard in terms of assessing cardiac structure and function, numerous hurdles exist to the provision of this service in the average Emergency Department. Other investigations are more likely to be performed but the optimum combination of investigative and clinical findings to robustly confirm or refute the diagnosis of heart failure in the breathless patient presenting to the Emergency Department has yet to be defined.

### **1.1.13 Aims of This Thesis**

The intention of this study was to look at the myriad of signs, symptoms, historical findings and investigative results related to the condition of heart failure in order to quantify the diagnostic utility of each within the Emergency Department environment. A systematic review was performed to examine all diagnostic studies carried out in this setting, with a reasonable reference standard diagnosis of heart failure as the end-point of the study. The intention was to synthesise a list of practical diagnostic criteria and quantify these in terms of sensitivity and specificity based on meta-analysis of data from the available literature.

The next step was the execution of original research in this diagnostic area in order to allow assessment of the independence of the suggested variables.

Using the results of this research, various techniques of combining the selected variables were examined in order to determine the optimal method in terms of sensitivity and specificity. The ultimate aim was to produce a clinical rule to allow the Emergency Physicians to confirm or refute the diagnosis of heart failure in the acutely dyspnoeic patient in the most reliable and timely fashion.

The aims this research are summarised thus:

Primary Aim: Develop and assess a clinical decision rule for patients with heart failure.

In [adult patients with suspected heart failure] what [diagnostic variables] can be combined to [confirm or refute the diagnosis]?

Secondary Aims:

1. To conduct a systematic review of medical literature relating to the diagnosis of heart failure
2. To carry out a meta-analysis of all the potential diagnostic variables for heart failure
3. To assess whether the selected variables were independent predictors of the diagnosis of heart failure
4. To explore the different methods for synthesizing a clinical decision rule and the assess the optimum rule

## Chapter 2 Diagnosing Heart Failure

### 2.1.1 Diagnosing Heart Failure

The diagnosis of heart failure is difficult, although interestingly, this only appears to have been acknowledged in medical literature following the discovery of the natriuretic peptides. Patients with heart failure tend to experience waves of exacerbation and remission in terms of their symptoms and so findings may not be consistent even within the same individual. The associations between smoking and the development both heart disease and lung disease is particularly problematic; not only can it be difficult to differentiate between chronic obstructive pulmonary disease and heart failure, but the patient may well have some elements of both present at the time of assessment.

Even in stable patients who have been investigated over a period of time the diagnosis may only be made from a combination of signs, symptoms, clinical history, investigations and progress over time, including response to treatment. More definitive investigations such as echocardiography or radionuclide angiography can be difficult to perform, are relatively expensive, take time to arrange and may carry some risk to the patient.

The presence of an acute exacerbation of one disease process in a patient with multiple pathologies can also cause worsening of the patient's condition with regard to their other conditions. It can be very difficult to decide which condition is primarily responsible for the patient's illness in a patient with multiple pathologies.

In attempting to differentiate which patients have presented due to acute, decompensated heart failure and which patients have presented due to another cause, it is important to recognise that the 'non-heart failure' group are not a homogeneous group. These patients will all have acute pathology causing their dyspnoea and the diagnostic strategies applied to

these patients will all have sensitivities and specificities that may be affected by their underlying condition.

In modern medical practice, the traditional diagnostic tools of careful history taking and exhaustive clinical examination, often using eponymous tests to skilfully elicit clinical signs of disease, have been supplemented and to some degree replaced by medical investigations. The skill of selectively applying these investigations in order to achieve the greatest utility, without subjecting the patient to unnecessary procedures or even risk of harm, has not always been appreciated (Moynihan et al., 2012).

### **2.1.2 Studies on the Presentation of Heart Failure**

Fonseca et al. (2004) recruited 5434 patients who had attended their GP for unrelated complaints and assessed the patients to decide if they had heart failure. Patients were considered to have possible heart failure if they scored 3 or more on the Boston questionnaire or were on diuretics for possible heart failure. The Boston criteria are described in a later section. Of the attending patients, 1058 patients were selected for further investigation and a further 174 of these patients were excluded due to technical difficulties in performing an echocardiogram. ECG, chest x-ray, laboratory tests and echocardiograms were performed on the remaining patients. Heart failure was considered the likeliest diagnosis in 551 of the patients. Paroxysmal nocturnal dyspnoea (LR 35.5), orthopnoea (LR 39.1) and breathlessness walking on flat (LR 25.8) were associated with having heart failure but the overall sensitivity was less than 36% as these were relatively uncommon findings in the patients who were considered to have heart failure.

Certain clinical findings were found to be specific for heart failure but were relatively uncommon findings in the patients with heart failure and so lacked sensitivity for this condition; the overall sensitivity was around 10%. The best performing variables are recorded in Table 2-1.

Table 2-1 Likelihood Ratios of Diagnostic Findings from Fonseca et al (2004)

Historical or clinical finding	Likelihood Ratio
Paroxysmal Nocturnal Dyspnoea	35.5
Orthopnoea	39.1
Breathlessness on walking on the flat	25.8
JVP >6cm and lower leg oedema	130.3
Rales on auscultation	23.3
Ventricular gallop	30.0
Heart rate >110	26.7

Wang et al. (2005) performed a systematic review and meta-analysis on studies that examined the diagnosis of heart failure in the Emergency Department setting. The methods for the meta-analysis are not described in detail but the likelihood ratios and confidence intervals were performed using random-effects models based on the delta method. The features that best predicted the diagnosis of heart failure were a previous history of heart failure (LR 5.8 95% CI 4.1-4.8), the symptom of paroxysmal nocturnal dyspnoea (LR 2.6 95% CI 1.5-4.5) and the presence of a third heart sound (LR 11, 95% CI 4.9 – 25.0). The features that best predicted a cause other than heart failure were the absence of a history of heart failure (LR .45 95% CI .38 - .53), the absence of dyspnoea on exertion (LR .48 95% CI .35 - .67) and a BNP level less than 100 pg/ml (LR .11 95% CI .07 - .16). A summary of the findings of this study is provided in Table 2- 2.

Table 2-2 Results from SR & Meta-analysis from Wang et al (2005)

Results from the Systematic Review and Meta-analysis by Wang et al.	
Features Predictive of Heart Failure	Likelihood Ratio
Previous history of heart failure	LR 5.8 (95% CI 4.1- 4.8)
Paroxysmal Nocturnal Dyspnoea	LR 2.6 (95% CI 1.5 - 4.5)
Presence of Third Heart Sound	LR 11.0 (95% CI 4.9 – 25.0)
Pulmonary Venous Congestion	LR 12.0 (95% CI 6.8 – 21.0)
Atrial Fibrillation	LR 3.8 (95% CI 1.7 – 88.8)
Features Predicting Absence of HF	Likelihood Ratio
BNP < 100 pg/ml	LR 0.11 (95% CI 0.07 -0.16)
Absence of History of Heart Failure	LR 0.45 (95% CI 0.38 - 0.53)
Absence of Dyspnoea on Exertion	LR 0.48 (95% CI 0.35 - 0.67)
Absence of Rales	LR 0.51 (95% CI 0.37 – 0.71)
Absence of Cardiomegaly on XR	LR 0.33 (95% CI 0.23 - .0.48)

Mant et al. (2010) performed a systematic review and meta-analysis of the diagnostic features and investigations involved in the diagnosis of heart failure in patients in primary care. The authors used the results of this study to develop a clinical rule to identify which patients should be referred for echocardiography as investigation for possible heart failure. The following features had diagnostic utility for this condition: dyspnoea; history of myocardial infarction; orthopnoea; oedema; elevated jugular venous pressure; cardiomegaly; added heart sounds; lung crepitations; hepatomegaly; ECG; and natriuretic peptides. The sensitivity and specificity of each of these findings is listed in along with the 95% confidence intervals (CI). The clinical decision rule derived was that in patients with suspected heart failure and breathlessness, the patient should be sent directly for echocardiography if there was a history of myocardial infarction, crepitations on auscultation or if the patient was a male with ankle oedema. In the absence of any of these features the recommendation was to measure the BNP or NT-proBNP level and refer patients with a raised level.

Table 2-3 Results of Meta-analysis Mant (2010)

Feature	Sensitivity (95% CI)	Specificity (95% CI)	Diagnostic Odds Ratio (95% CI)
Hx of MI	0.26 (0.19-0.37)	0.89 (0.85-0.91)	2.87 (1.71-4.82)
Dyspnoea	0.83 (0.62 – 0.94)	0.54 (0.4 – 0.67)	5.71 (1.78 – 18.31)
Orthopnoea	0.44 (0.33 – 0.56)	0.89 (0.69 – 0.96)	6.23 (2.3 – 16.92)
Oedema	0.53 (0.44 – 0.62)	0.72 (0.62 – 0.8)	2.91 (1.89 - 4.49)
Raised JVP	0.52 (0.41 – 0.63)	0.7 (0.56 – 0.80)	2.52 (1.51 4.22)
Added HS	0.11 (0.04 - 0.24)	0.99 (0.97 -1.0)	13.4 (6.58 – 27.3)
Crepitations	0.51 (0.44 – 0.58)	0.81 (0.71 – 0.88)	4.34 (2.91 – 6.47)
Abnormal ECG	0.89 (0.77 – 0.95)	0.56 (0.46 – 0.66)	4.8 (4.36 – 25.7)
Abnormal CXR	0.68 (0.40 – 0.88)	0.83 (0.66 – 0.93)	10.7 (4.45 – 25.5)
CTR	0.67 (0.53 – 0.78)	0.76 (0.65 – 0.84)	6.25 (3.6 – 10.8)
BNP	0.93 (0.91 – 0.95)	0.74 (0.63 – 0.83)	39.5 (21.44 – 72.6)
NT-proBNP	0.93 (0.88 – 0.96)	0.65 (0.56 – 0.74)	24.6 (14.4 – 42.2)

Madhok et al. (2008) performed a similar systematic review and meta-analysis in 2008 looking at the diagnosis of heart failure in primary care.

There was a substantial overlap with studies included in the Mant paper although some differences in perceived utility of the diagnostic features as shown in Table 2-4.

Table 2-4 Results of Meta-analysis Madhok (2008)

Diagnostic Test	Pooled PLR	(CI) or Range	Pooled NLR	(CI) or Range
Clinical History				
History of MI	2.86	1.37 - 4.40	0.69	0.48 - 0.89
Diabetes	2.29	0.86 - 6.65	0.95	0.89 - 1.02
Hypertension	0.58	(0.39 - 0.87)	1.30	(1.13 - 1.50)
Male	1.61	(1.41 - 1.84)	0.68	(0.60 - 0.77)
Symptoms				
Dyspnoea	1.15	(1.09 - 1.21)	0.50	(0.20 - 1.26)
Orthopnoea	1.59	0.89 - 3.58	0.89	0.77 - 1.04
PND	1.71	1.12 - 2.23	0.87	0.75 - 0.99
Signs				
Peripheral oedema	1.18	0.96 - 1.48	0.92	0.74 - 1.05
Abnormal Breath Sounds	1.53	(1.17 - 1.19)	0.85	0.64 - 0.94
Raised JVP	4.36	2.66 - 7.44	0.88	(0.83 - 0.91)
Displaced Apex Beat	15.96	(8.24 - 30.93)	0.58	0.35 - 0.93

3rd Heart Sound	7.34	1.56 - 32.37	0.92	0.77 - 0.96
<b>Tests</b>				
CXR	1.47	1.19 - 1.71	0.58	0.49 - 0.77
ECG	2.13	(1.95 - 2.33)	0.27	0.06 - 0.76
BNP (cut - offs nearest 15pmol/L)	1.90	1.26 – 6.20	0.30	0.02 - 0.80
NT-proANP	2.98	(2.24 - 3.97)	0.18	0.04 - 0.64
NT-proBNP	1.39	1.09 - 1.80	0.26	0.02 - 0.80
<b>Combined</b>				
ECG and Hx of MI	2.80	(2.43 - 3.24)	0.13	0.03 - 0.48

These studies show that while there is a wide variety of signs and symptoms that are associated with heart failure, there are no clinical variables that either define heart failure or rule it out completely. The presence of a displaced apex beat was strongly predictive of heart failure although was only assessed in two studies and was found in a very small number of patients. Natriuretic peptide concentrations and the presence of abnormal ECGs had a similar value in terms of ruling out this syndrome. Although there are common themes within the findings of these studies there is also significant variation within the results and the reasons for this are not clear. It is not clear if the best performing variables are independent of each other.

### 2.1.3 Clinical History

In keeping with the fluctuating nature of this condition, patients with heart failure will often have a history of previous attendances with the same condition. The majority of patients with heart failure also have a history of ischaemic heart disease with angina or a previous myocardial infarction. A history of hypertension is also a significant factor, as is diabetes. Other

associated conditions include peripheral vascular disease, valvular heart disease and impaired renal function (Acanfora et al., 1998).

Davie et al. (1997) looked at 259 patients referred by their General Practitioner to an outpatient echocardiography service with suspected chronic heart failure. 41 patients were found to have significant left ventricular systolic dysfunction on echocardiography. The positive predictive factors were past medical history of myocardial infarction (sensitivity 59%, specificity 86%, PPV 44%, NPV 92%) and hypertension (sensitivity 20%, specificity 65%, PPV 10%, NPV 81%). Cigarette smoking was sensitive but had poor predictive value due to high prevalence (sensitivity 73%, specificity 41%, PPV 19%, NPV 89%). The use of diuretic as a discriminator performed in a similar manner (sensitivity 73%, specificity 41%, PPV 19%, NPV 89%).

In the study by Wang et al. (2005) the historical features that contributed to the diagnosis of heart failure were examined. A previous history of heart failure (LR 5.8; 95% CI 4.1-8.0), myocardial infarction (LR 3.1; 95% CI 2.0-4.9) or coronary artery disease (LR 1.8; 95% CI 1.1-2.8) was predictive of heart failure. Absence of any of these conditions also reduced the likelihood of heart failure being the cause of the patient's dyspnoea.

#### **2.1.4 Clinical Examination**

A variety of clinical findings are associated with the condition of heart failure and form part of the standard cardiovascular examination as described in any medical textbook. The diagnostic value of the presence or absence of these disease markers is less easy to ascertain from these texts.

A review article by Cleland and Habib (1996) on the assessment and diagnosis of heart failure suggests that pulse, blood pressure, jugular

venous pressure, the presence of pitting oedema, the apex beat and parasternal heave, the presence of a third heart sound and cardiac murmurs all have some diagnostic value.

The study by Davie et al. (1997) found a displaced apex beat to be the best clinical sign (sensitivity 66%, specificity 96%, PPV 75%, NPV 94%). Orthopnoea (sensitivity 22%, specificity 74%, PPV 14%, NPV 83%) and paroxysmal nocturnal dyspnoea (sensitivity 39%, specificity 80%, PPV 27%, NPV 87%) were of limited value.

You need to make a comment on some of the figures e.g. 14NPV etc.

Drazner et al. (2001) performed a retrospective analysis of the data from a large study looking at the benefit of enalapril in the treatment of heart failure. Multivariate analysis of this data in patients with a diagnosis of heart failure suggested that the presence of a raised JVP or a third heart sound are associated with an increased risk of death or hospitalisation for heart failure.

Knudsen, Riis, et al. (2004) examined a subgroup of patients originally recruited to the Breathing Not Properly study, by only including patients who were not missing any of their clinical data (Maisel et al., 2002). This limited the new data set to 880 patients out of the original 1586 patients. The diagnostic value of the following clinical features was calculated from this subset of patients as shown in Table 2-6.

Table 2-5 Diagnostic Value of Clinical Findings, Knudsen (2004)

	Sensitivity	Specificity
Hx of HF	62	87
Previous MI	43	87
Orthopnoea	66	57
JVP	38	90
S3 Gallop	38	90
Rales	59	77

## 2.2 Investigations

### 2.2.1 The Electrocardiograph

Electrocardiograms (ECG) represent the flow of electrical energy through the cardiac tissue as observed from specified points on the surface of the body. As a diagnostic test they are superb in providing diagnosis-confirming findings, while being cheap, fast, safe, portable and non-invasive. ECGs are often abnormal in patients with heart failure, which reflects the fact that the conducting system is traversing diseased or damaged tissue. Although it is uncommon to have a completely normal ECG in a patient with heart failure, there are no findings that are specific to heart failure (Davie et al., 1997). The patient with heart failure is usually tachycardic and may be in atrial fibrillation or have another rhythmic disturbance. Changes associated with left ventricular hypertrophy are commonly present and changes suggestive of ventricular strain may also be present. The ECG may also show inter-ventricular

conduction delay, which can be associated with ventricular dyssynchrony. The diagnosis of acute myocardial infarction must be considered in the patient with a sudden onset of cardiogenic symptoms.

Davie et al. (1996) investigated 534 patients referred by GPs to an open-access echocardiography service and found that an ECG was a very good predictor of LV dysfunction. When interpreted by the cardiologists this test had a sensitivity of 93.6% and a specificity of 61.4%.

Houghton et al. (1997) examined the ECGs from 200 patients with suspected heart failure who had been referred to a heart failure clinic. The ECGs were read and reported by a cardiologist and two General Practitioners. The patients all underwent echocardiography and were reported as having normal or impaired LV function. The ECGs were reported simply as normal or abnormal. There was agreement between the cardiologists report and that of the two GPs, in 165 out of 200 ECGs although no Kappa scores were provided. The presence of an abnormal ECG, when interpreted by the cardiologist, had a sensitivity of 89.1% and a specificity of 45.7% in predicting LVSD.

Brooksby et al. (1999) looked at the QT parameters on ECGs in 554 patients with known heart failure who were followed up for over one year. The QT interval length was significantly associated with all-cause mortality although this was not demonstrated to be an independent predictor on multivariable analysis.

Shamim et al. (1999) retrospectively looked at data from 241 patients with heart failure. On univariate analysis by the Cox proportional Hazard method both inter-ventricular conduction delay (IVCD) and a prolonged, corrected QT interval were shown to be independent predictors of mortality over a mean follow-up period of 31 months. IVCD had better predictive power than the corrected QT interval.

Baldasseroni et al. (2003) analysed primary follow-up data from the Italian Registry of CHF. They found that the presence of left bundle branch block (LBBB) on the ECG was associated with a significantly higher one-year-mortality rate than those patients without (16.1% vs. 11.9%). Right bundle branch block (RBBB) was not associated with a higher mortality rate.

Ng et al. (2003) looked at a random selection of 1360 patients from Leicestershire in the UK who underwent investigation for possible left ventricular dysfunction. The patients were stratified according to GP list size and deprivation scores. All of the recruited patients had echocardiography, resting ECG, assessment of ponderal index and blood sampling for natriuretic peptide levels. Significant factors on univariate analysis were BNP levels, gender, major ECG abnormality (Q waves, LVH, AF, LBBB), a history of ischaemic heart disease and plasma creatinine levels. Sensitivity of the presence of major ECG abnormalities as a predictor of left ventricular dysfunction was 88%.

Khunti et al. (2004) performed a systematic review published in 2004 and found four studies examining the value of a normal or abnormal ECG in determining which primary care patients had heart failure. The sensitivity from the included studies varied from 73% to 91% while the specificity ranged from 20% to 65%.

Kruger et al. (2004) looked at 128 patients referred for an echocardiogram due to suspected heart failure. Left ventricular systolic dysfunction (LVSD) was defined as an ejection fraction less than 50%. Cardiac catheterisation was carried out in all of the patients. They found that a cut-off for the QRS complex of greater than 0.12s had a sensitivity of 76% and a specificity of 95%. Sixty-six of the one hundred and twenty-eight patients referred were diagnosed with LVSD.

Fonseca et al. (2004) explored a random selection of 6300 patients over the age of 25 years to examine the prevalence of heart failure within the

population of Portugal. Patients who were suspected of having heart failure were underwent chest radiography, ECG, blood tests and an echocardiogram. Seventeen per cent of the patients were excluded from the study, as they did not have an echocardiogram performed. In the selected group of patients, the finding of an abnormal ECG had a sensitivity of 80% and a specificity of 40% for heart failure.

Knudsen, Omland, Clopton, et al. (2004) examined a subset of the data from the Breathing Not Properly study where more complete data was available to determine the value of the chest x-ray, ECG and clinical findings. These authors only examined the findings for 880 patients with complete data out of the 1580 patients recruited for the larger study. The gold standard was the consensus opinion of two cardiologists who reviewed all the results and responses to treatment. The presence of an abnormal ECG was associated with a sensitivity of 58% and a specificity of 78% for the end-diagnosis of heart failure.

Murkofsky et al. (2005) found that the presence of a prolonged QRS interval suggests LV dysfunction. A QRS segment duration greater than 0.10s had specificity of 83.6% but a sensitivity of 43.8%. Increasing the cut-off value of the QRS duration up to 0.12s increased specificity to 99.3% but decreased sensitivity to 13.8%.

In the meta-analysis by Wang et al. (2005), the presence of AF was the most important factor with a LR of 3.8 (95% CI 1.7-8.8). The presence of a completely normal ECG decreased the likelihood of heart failure with a LR of 0.6 (95% CI 0.47-0.88) but there were no specific findings associated with the condition.

The variability demonstrated by the findings of these studies is likely to reflect the diversity of the aetiology of heart. Although it is logical to assume that heart failure is unlikely to occur in a non-diseased heart and that cardiac disease will often cause ECG changes it is difficult to define anything stronger than a positive correlation.

### 2.2.2 The Chest X-Ray

Pulmonary vascular congestion occurs as a result of impaired pumping by the left ventricle resulting in an increased left atrial pressure. This leads, in turn, to pulmonary hypertension. In the absence of heart failure, more blood is distributed to the lower parts of the pulmonary circulation and less to the upper vessels due to gravitational pull. In pulmonary hypertension this pattern of distribution is lost as the increased pressure causes the upper lobe vessels to dilate. Once the mechanisms to compensate for increased hydrostatic pressure are overcome, excess fluid accumulates in the interstitium before eventually accumulating within the alveoli and pleural spaces. This accumulation of the excess fluid causes distinctive patterns on the chest radiograph (Chen, 1992).

Madsen et al. (1984) looked at 229 patients who were discharged following a myocardial infarction, defined as a history of chest pain and a rise in creatinine kinase. All patients had a chest x-ray and radionuclide ventriculography done prior to discharge. An ejection fraction less than or equal to 0.51 was considered abnormal. The presence of pulmonary venous congestion on chest x-ray was found to have a sensitivity of 20.1% and a specificity of 91.6% for predicting a reduced ejection fraction. Cardiothoracic ratio (CTR), the relationship between the internal thoracic diameter and the diameter of the heart at its widest point, was also recorded for each patient. The internal diameter of the chest was measured at the highest point of the left hemi-diaphragm. The presence of a CTR greater than, or equal to 0.50, was considered abnormal. This had a sensitivity of 47.3% and a specificity of 65.6%.

A paper by Henriksson et al. (1990) examined the level of agreement between three radiologists in reporting chest x-rays performed on 27 patients with suspected heart failure. The paper was originally designed to look at differences in two different types of film-screen for viewing the x-rays so each patient had two x-rays taken, a few minutes apart, with

each film system. The radiologists were asked to grade the presence of congestive heart failure from 0 to 2 for each x-ray. In fact, the study showed poor correlation between even the same radiologists reporting on the same patient with Kappa scores of 0.6, 0.42 and 0.1 for three radiologists. There was no consistent difference based on the type of film used. Comparing the three radiologists results with each other gave Kappa scores of 0.26, 0.18 and 0.02.

Knudsen, Omland, Clopton, et al. (2004) looked at the data from the Breathing Not Properly study to determine the value of the chest x-ray. This article only included 880 patients with complete data out of the 1580 patients recruited for the main study. A radiologist from the recruiting hospital reported the x-rays. The gold standard was the consensus opinion of two cardiologists who reviewed all the results and responses to treatment. They found that the presence of cardiomegaly had a sensitivity of 79% and a specificity of 80% while cephalisation of the vessels had a sensitivity of 41% but a specificity of 96%.

In a study by Fonseca et al. (2004), 6300 patients over the age of twenty-five years were assessed as to whether they may have heart failure, the finding of an abnormal chest x-ray had sensitivity of 57% and a specificity of 78%. The most sensitive component of chest x-ray abnormality was the presence of cardiomegaly with an enlarged cardio-thoracic ratio greater than 0.50, which provided a sensitivity of 54% with a specificity of 79%. Pulmonary vessel cephalisation and lung interstitial oedema both had had a sensitivity of 18% with specificities of 94% and 95% respectively.

Wang et al. (2005) also evaluated the diagnostic use of chest x-rays. The authors included data from seven studies and concluded that the finding of pulmonary venous congestion (LR 12.0, 95% CI 6.8-21.0) and cardiomegaly (LR 3.3, 95%CI 2.4-4.7) increased the chances of the patient having heart failure. Conversely the absence of cardiomegaly had a negative LR of 0.33 (95% CI 0.23-0.48) as did the presence of

pneumonia or hyperinflation although these were only mentioned in one of the included studies.

Høilund-Carlsen et al. (2005) also looked at patients who had suffered a myocardial infarction and who underwent both chest x-ray and radionuclide ventriculography on the same day, during the second week following their infarction. The presence of an enlarged left ventricle, as defined by a single radiologist on the basis of their chest x-ray had a sensitivity of 63% and a specificity of 68% for predicting a LVEF of <0.53%.

There are two issues associated with determining the presence of pulmonary oedema on chest x-ray in order to diagnose heart failure; the first is the degree of subjectivity in interpreting the film which is demonstrated by the relatively poor level of agreement reported. These findings are consistent with other studies examining the level of agreement between radiologists reporting on chest x-ray findings for various different conditions. (Badgett et al., 1996; Tudor et al., 1997; Robinson et al., 1999). It is reasonable to assume that similar levels of disagreement will exist if one were to assess the interpretation of chest x-rays by Emergency Physicians rather than radiologists. Although there is some suggestion that training can improve consistency in the reporting of the presence of the signs of cardiac failure this would limit the applicability of any decision pathway that relied upon it.

The second factor is that retrospective studies suggest that up to 20% of patients with a discharge diagnosis of heart failure do not have pulmonary oedema on their initial chest x-ray (Collins, Lindsell, Storrow, et al., 2006).

### 2.2.3 Blood Tests

Several blood tests may be abnormal in the patient with heart failure, and this may provide useful information suggestive of the underlying aetiology although none are pathognomonic. For example, patients may present with an exacerbation of heart failure due to anaemia or hyperthyroidism (Struthers, 2000). There is also a strong, though incompletely understood, association between renal impairment and heart failure. Other compounds levels, such as cholesterol levels, may not aid in the acute diagnosis but may be relevant in this patient group.

### 2.2.4 Echocardiography

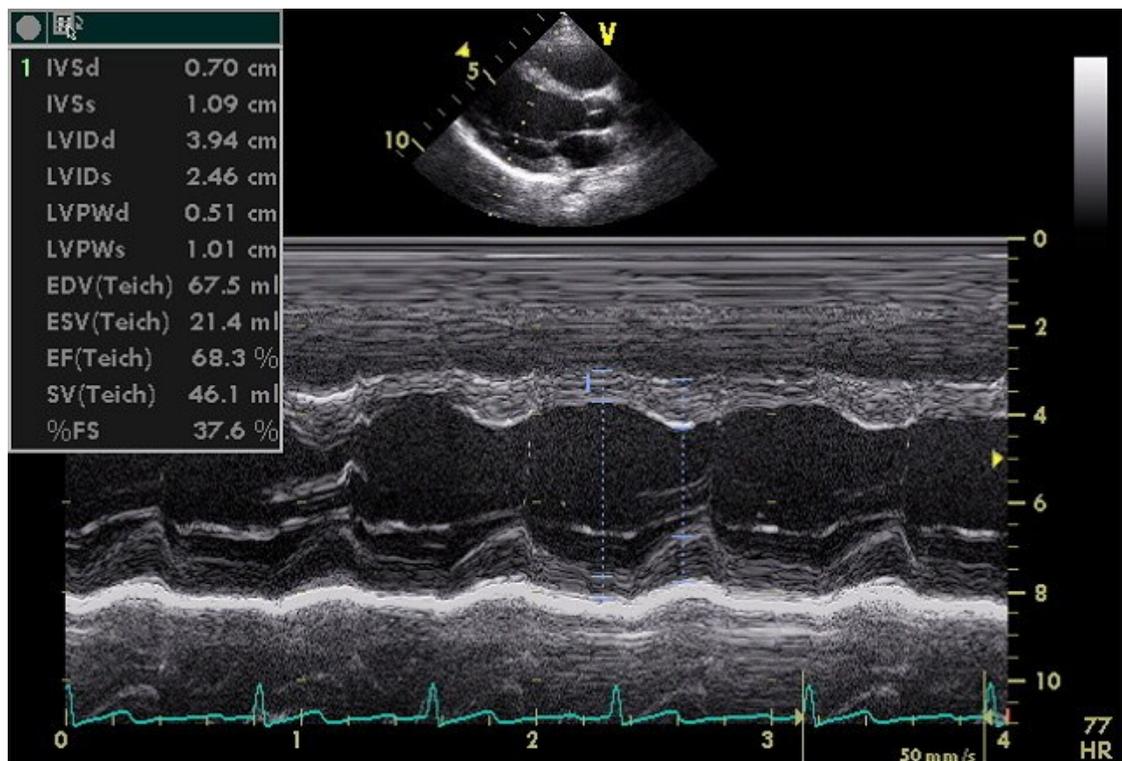


Figure 2-1 Echocardiography

The AHA/ ACC Guidelines (Jessup et al., 2009) describe echocardiography as the single most useful diagnostic test in the

evaluation of patients with heart failure. Echocardiography uses ultrasound technology to allow an evaluation of the heart in terms of structure and function. The reflection of high-frequency sound waves is used to make various measurements during the contraction cycle. Two-dimensional measurements of the heart walls and chambers can be recorded at different stages in the cardiac cycle and the rate of flow of blood during the contraction can be measured using the Doppler phenomenon. This provides information throughout the cardiac cycle and presents both systolic and diastolic function (Porter et al., 1990; Kirkpatrick et al., 2007).

Although standard echocardiography displays a series of two-dimensional images to represent a three-dimensional object, the combination of different views and dynamic imaging is the best representation available of cardiac function, in life, to allow estimations of the cardiac size and function. Various methods are used to quantify the ejection fraction, although these are inherently inaccurate, as assumptions have to be made regarding the shape of the ventricle; it is necessary to represent a three-dimensional form by a series of two-dimensional measurements (Bellenger et al., 2000). It is also possible to obtain a measurement of the degree of cardiac contraction in various regions of the cardiac muscle wall; identifying localised areas abnormal contraction can identify the area of pathology (Zimmerman & Vacek, 2011).

The quality of the images recorded depends on the patient body habitus, the quality of the equipment and the operator experience and skill: there is also an element of operator-dependent variability in the reporting of the results (Bellenger et al., 2000). Patient cooperation is required to acquire optimal images and this may be difficult to obtain in the acutely ill patient.

Trans-thoracic echocardiography is performed through the anterior chest wall. This is non-invasive but depends on obtaining images through echocardiographic 'windows' between ribs and through the patient's lung tissue. Trans-oesophageal echocardiography, in which the probe is

passed down the patient's oesophagus, produces better views from the other side of the heart but usually requires the administration of sedation to be tolerable.

Techniques allow the use of echo to look at myocardial contraction, valvular structure and function, pericardial pathology and pathological consequences of myocardial infarction or ischaemia. Measurements obtained allow an estimate of cardiac output and pulmonary artery pressure. The investigation can also be repeated in the same patient over time to assess progression of the disease process or response to treatment. The echocardiogram findings suggestive of cardiac disease are presented in Table 2-6. This is based upon a table from the ESC Heart Failure Guidelines (McMurray et al., 2012).

Diastolic failure is due to the inability of the heart muscle to relax fully to allow diastolic filling of the heart. This form of disease is more common in elderly patients and, although it presents with similar symptoms, it can be more difficult to diagnose, as the echocardiogram findings associated with this variation are subtler (Aurigemma et al., 2001).

Table 2-6 Abnormal Echocardiogram Findings

Common Echocardiographic Abnormalities	
Parameters related to systolic function	
LV ejection fraction	Decreased
LV fractional shortening	Decreased
LV regional function	Dyskinesia, akinesia
LV end-diastolic size	Increased
LV end-systolic size	Increased
Parameters related to diastolic function	
LV diastolic dysfunction parameters	Abnormal mitral inflow pattern
LA volume index	Increased
LV mass index	Increased
<ul style="list-style-type: none"> <li>▪ Parameters related to valvular function</li> </ul>	
Valvular structure / function	Stenosis or regurgitation
Other parameters	
RV function	Decreased
Tricuspid regurgitation	Increased
Systolic pulmonary artery pressure	Increased
<ul style="list-style-type: none"> <li>▪ Inferior vena cava</li> </ul>	Dilated, no respiratory collapse
Pericardium	Effusion, haemopericardium, calcification

## 2.2.5 Invasive Techniques

Cardiac output can also be measured by invasive techniques although these methods have the potential disadvantages; the insertion of a central line exposes the patient to risks of harm such as bleeding or infection and it is possible that the presence of line within the vessel or chamber lumen may itself affect the measurements (Binanay et al., 2005; Lavdaniti, 2008; Mathews & Singh, 2008).

Pulmonary Artery Balloon Flotation Catheterisation is an invasive examination that can confirm the diagnosis of heart failure, and sometimes define the exact cause. It is possible to measure the pulmonary artery occlusion (wedge) pressure, central venous pressure, right ventricular end-diastolic pressure and cardiac output. Although there is a theoretical benefit of using PAC monitoring to confirm the diagnosis and guide therapy there are also risks associated with the invasive nature of this technique. Overall no survival or outcome benefit was demonstrated when this monitoring was applied to patients with heart failure in a randomised controlled trial (Vesely, 2003; Binanay et al., 2005).

### **2.2.6 Magnetic Resonance Imaging**

Magnetic resonance imaging can provide still and even moving images of the heart as an assessment of cardiac function. Unfortunately the procedure requires the patient to lie flat within an enclosed space for a prolonged period of time (Hundley et al., 1999). This requirement, in addition to the limited monitoring that is available for these patients during this time, means that this investigation is not practical for the acute setting.

### **2.2.7 Nuclear Studies**

Radionuclide studies are useful for measuring both systolic and diastolic cardiac function (Klocke et al., 2003). This study involves labelling the red blood corpuscles from a patient, and then measuring the radiation emitted during the cardiac cycle. The investigation is recommended for patients with suspected heart failure who cannot be investigated by other means, for example echocardiography. Due to the time that this procedure takes, the equipment that it requires, and the expertise necessary to interpret the results, it is not practical to perform this in an acute setting (Hendel et al., 2009).

### **2.2.8 Peripheral Temperature**

Temperature has been found to be a significant marker of prognosis in terms of mortality and serious complications in patients admitted to hospital due to cardiac failure (Casscells et al., 2005; Ahmed et al., 2008). It is not clear why this should be the case but is possible that this is a marker of other underlying conditions causing the exacerbation of the cardiac failure (Payvar et al., 2006).

One paper by Clarke et al. (2005) has attempted to define cardiac from non-cardiac causes of shortness of breath by the difference in peripheral versus central temperature. This small study did suggest a significant difference but contained small numbers and did not assess variations in the acquisition of the peripheral temperature.

### **2.2.9 Response to Treatment**

Theoretically the response of a patient to a particular treatment could be used as part of a reference standard for a condition if the effect is specific for that condition. In practice, several interventions are likely to be applied to any patient at the same time and so it is difficult to elicit which, if any, the patient derived benefit from. Obviously, the response to treatment for any individual is only known after the treatment has been commenced and so is of no value in the emergency setting.

### **2.2.10 Scoring Systems for the Diagnosis of Heart Failure**

As it has been recognised that there is no single diagnostic feature that can identify patients with heart failure with any consistency, attempts have been made to combine relevant variables into scoring systems in order to facilitate the diagnosis process.

The Framingham Criteria for Congestive Heart Failure were derived from the long established, and still on-going cohort study based in the

eponymous city (McKee et al., 1971). The population of 5209 adults present in Framingham in 1948 were recruited as a prospective cohort. A biennial assessment was carried out on each person in the study that consisted of a medical history, examination and laboratory tests. The criteria to define heart failure were derived from the observations made of this population over the first twenty-three years of this study (McKee et al., 1971; Tamura et al., 1996). In the Framingham classification system, the diagnosis of CHF requires that either 2 major criteria or 1 major and 2 minor criteria be present concurrently. This is shown in Table 2-7. Minor criteria are accepted as evidence of heart failure only if they cannot be attributed to another medical condition. The authors found the criteria to be 100% sensitive and 78% specific.

Table 2-7 Framingham Criteria for Heart Failure

Framingham Criteria for Heart Failure
<b>Major Criteria</b>
Paroxysmal nocturnal dyspnoea
Neck vein distension
<b>Rales</b>
<b>Radiographic</b>
Acute pulmonary oedema
Neck vein distension
Circulation time of 25 seconds
Hepato-jugular reflux
Pulmonary oedema or visceral congestion at autopsy
Central venous pressure greater than 16 cm water
Weight loss of 4.5 kg in 5 days in response to treatment
<b>Minor Criteria</b>
Bilateral ankle oedema
Nocturnal cough
Dyspnoea on ordinary exertion
Hepatomegaly
Pleural effusion
A decrease in vital capacity by one third the maximal value recorded
Tachycardia (rate of at least 120 bpm)

Killip and Kimball (1967) published a paper promoting the use of coronary care units for patients who had suffered a myocardial infarction. Two hundred and fifty patients were treated either on a regular ward or a coronary care unit and the mortality rate of each group was compared. In order to keep the groups comparable, the severity of their condition was stratified using a clinical classification score derived by the authors: the Killip Classification as shown in Table 2-4.

Table 2-8 Killip Classification of Severity of Heart Failure

Class	Definition
I	No clinical signs of heart failure
II	Rales or crackles in the lungs an S <sub>3</sub> , and elevated jugular venous pressure.
III	Frank acute pulmonary oedema
IV	In cardiogenic shock or hypotension measured as systolic blood pressure lower than 90 mmHg), and evidence of peripheral vasoconstriction (oliguria, cyanosis or sweating).

The Boston criteria for heart failure were originally derived from 96 patients undergoing right heart catheterisation. A score greater than 4 was found to correspond with a pulmonary capillary wedge pressure greater than 12mmHg with a sensitivity of 90% and a specificity of 85%. A maximum of 4 points allowed from each category leading to a maximum possible score of 12. 8-12 points is considered definite heart failure, 5-7 possible heart failure and <5 unlikely to be heart failure (Carlson et al., 1985).

Table 2-9 Boston Criteria for Heart Failure

Criterion	Point value
<b>Category 1: history</b>	
Rest dyspnoea	4
Orthopnoea	4
PND	3
Dyspnoea walking on flat	2
Dyspnoea while climbing	1
<b>Category II: physical examination</b>	
Heart rate abnormality (1pt 91-110bpm, 2 pts >110bpm)	1 or 2
JVP (2 pts if >6cmH <sub>2</sub> O, 3 pts >6cm H <sub>2</sub> O and hepatomegaly)	2 or 3
Lung crackles (1pt basilar, 2pts more than basilar)	1 or 2
Wheezing	3
Third heart sound	3
<b>Category III: chest radiography</b>	
Alveolar pulmonary oedema	4
Interstitial pulmonary oedema	3
Bilateral pleural effusion	3
Cardiothoracic ratio > 0.50	3
Upper zone flow redistribution	2

Remes et al. (1991) validated the Boston criteria, applying it to 88 patients suspected by their GP to have heart failure. They compared it against their own system that included clinical and radiological findings, a respiratory physician's opinion and response to treatment over a six-

month period. They found that the diagnosis of 'definite' heart failure by the Boston criteria, i.e. a score of 8 or greater had a sensitivity of 80% and a specificity of 92% using their own final diagnosis as the reference standard.

Eriksson, Caidahl, et al. (1987) used the prospective population of study of cardiovascular disease, the Study of Men Born in 1913 to construct a scoring index to differentiate causes of dyspnoea. Dyspnoea graded according to Fletcher questionnaire from no dyspnoea on exertion (grade 0) to dyspnoea at rest (grade 4).

Di Bari et al. (2004) compared four methods of diagnosing heart failure in unselected adults in the community. The authors collected data on people aged 65y or over and applied the criteria from the Framingham, Boston and Gothenburg studies and also from the guidelines published by the European Society of Cardiology. The authors evaluated the patients and measured left ventricular mass index, LVEF, left atrium systolic dimension, lower extremity motor disability, summary physical performance score and a six-minute walk test. Predictive validity was assessed with follow up looking at cardiovascular mortality, incident disability and heart failure related hospitalisations. All criteria suggested a similar prevalence of heart failure in the population tested. Framingham and Boston criteria were better at correlating with the findings of the investigations. The Boston criteria were superior in indicating risk adjusted cardiovascular death.

Kim (2006) suggested that the Framingham Criteria was impractical and proposed a new modified system The Minnesota Heart Failure Criteria. This abridged form of the Framingham Criteria was derived using latent class analysis (LCA). LCA is described as a probabilistic approach to disease classification that allows for, and can potentially identify more precise categories of disease conditions. The authors calculated the criteria retrospectively using data from patients who had already been

recruited into the Minnesota heart failure study or the Minnesota post-myocardial infarction study.

Kim (2006) validated their diagnosis of heart failure on the basis of the mortality rate by making the assumption that patients with heart failure would have a higher mortality rate than patients without the condition within a two-year follow up period. The authors stated, 'in the absence of a perfect gold standard, a diagnostic test can be validated by comparing its results against a known consequence of a disease.'

The variables were taken from the Framingham Criteria with the addition of left ventricular ejection fraction (LVEF). They consisted of: dyspnoea at rest or on exertion; pulmonary rales; cardiomegaly; interstitial or pulmonary oedema; the presence of a third heart sound; tachycardia; and LVEF. A five-class model had a good fit to the observed data. Class 1 was 'non-cases'. Class 5 patients almost all had dyspnoea, pulmonary rales and cardiomegaly. They were also very likely to have a reduced LVEF and a third heart sound. These patients were considered 'advanced heart failure'. This system requires the statistics software SAS© in order to calculate which group the patient is most likely to belong to. This system does not appear to be practical for use in an emergency setting. The link to the required software was not functioning at the time of checking (February 2014).

### **2.2.11 Assessing the Severity of Heart Failure**

The New York Heart Association (NYHA) developed a classification of patients with heart disease on the basis of the relation between symptoms and the amount of effort required to provoke them (Kossman, 1964). It was originally derived in 1928 but updated and modified in 1994. This system was designed for the classification of the severity of the condition rather than performing any diagnostic function. Although the process of assigning numerical values to subjective findings has obvious limitations, this classification is nonetheless useful to grade the severity of

the condition thereby allowing comparison of groups of patients, as well as describing the condition of the same patient at different times. In addition, the NYHA class has proved to be a strong, independent predictor of survival in patients with chronic heart failure although it was intended for use in patients with known cardiac disease rather than as a diagnostic tool. The severity of the patients' heart failure may be symptomatically classified according to the amount of effort needed to produce dyspnoea, as described in

Table 2-10.

Class I	No limitation of physical activity
Ordinary physical activity does not cause undue fatigue, dyspnoea, or palpitation. Patients have symptoms only at exertion levels similar to those of relatively healthy individuals	
Class II	Slight limitation
Patients with class II disease are comfortable at rest. Ordinary physical activity results in fatigue, palpitation, dyspnoea, or angina. Patients exhibit symptoms with ordinary exertion.	
Class III	Marked limitation
Although patients are comfortable at rest, less-than-ordinary activity leads to symptoms. Patients exhibit symptoms with minimal exertion.	
Class IV	Unable to carry out any activity without discomfort
Symptoms of congestive failure are present even at rest. With any physical activity, increased discomfort is experienced.	

Table 2-10 NYHA Classification of Heart Failure

## 2.2.12 Diagnosing Diastolic Heart Failure

Diastolic failure occurs due to the inability of the heart muscle to relax fully to allow diastolic filling of the heart. This form of disease is more common in elderly patients and, although it presents with similar symptoms, it can be more difficult to diagnose, as the echocardiogram findings associated with this variation are subtler (Aurigemma et al., 2001)

Diastolic heart failure is diagnosed when the syndrome of heart failure arises in the absence of significant abnormality of left ventricular contractile function. These patients are often excluded from heart failure trials and so there is less literature available regarding this condition.

Diastole is traditionally divided into four phases: (1) Isovolumic relaxation; (2) early diastolic filling; (3) diastasis and; (4) atrial contraction. During isovolumic relaxation the myosin and actin cross-bridges within the cardiac myocytes must be dissolute. This requires the rapid uptake of calcium in order to reduce cytosole levels. This is an active process facilitated by the sarcoplasmic reticulum. Some myofilaments may maintain the cross-bridges even with a lower calcium concentration. In the failing human heart, myofilaments and calcium exchange pumps work more slowly which may be responsible for the slowing of left ventricular relaxation.

Early diastolic filling begins as the ventricular pressure drops and the mitral valve opens. Around 80-85% of the LV filling usually happens at this point. Diastasis and the contraction of the atria then occur. The effectiveness of the atrial contraction is dependent on the LV compliance but also the atrial preload, heart rate, atrial contractility and atrial geometry. The atrial contribution may be increased in impaired LV function or during exercise.

The clinical features of diastolic heart failure cannot be reliably differentiated from systolic heart failure. A definitive diagnosis requires cardiac catheterisation. Using Doppler echocardiography to measure blood velocity across the mitral valve usually gives a characteristic appearance with a large E peak representing rapid early diastolic filling followed by a smaller A peak during atrial systole as shown in Figure 2-2. If LV relaxation slows then the atrial component becomes more important and there is reversal of the E/A ratio (van Kraaij et al., 2002).

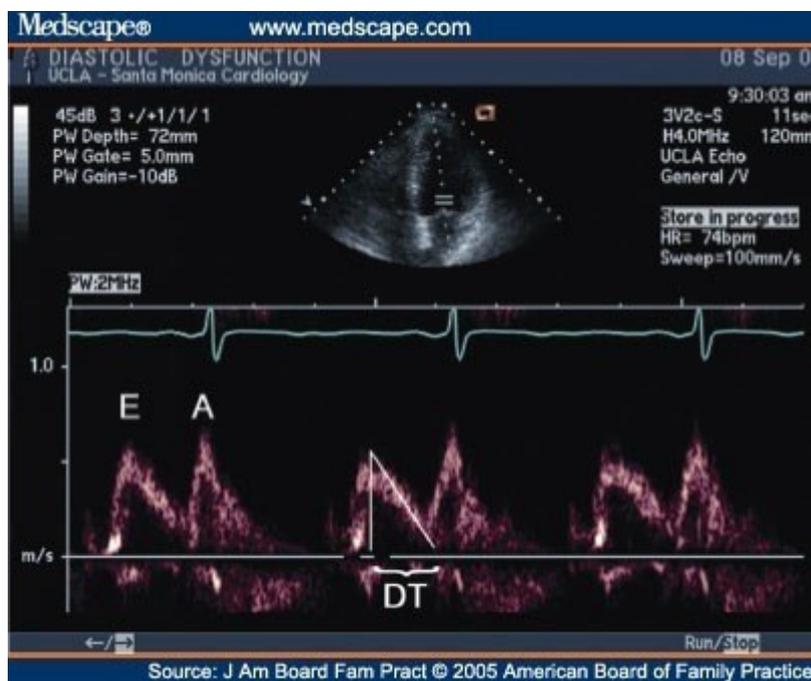


Figure 2-2 Echocardiography Demonstrating Diastolic Dysfunction

The European Study Group on Diastolic Heart Failure has suggested the following criteria for the diagnosis of primary diastolic heart failure.

1. Presence of signs or symptoms of congestive heart failure
2. Presence of normal or mildly reduced left ventricular systolic function
3. Evidence of abnormal left ventricular relaxation, filling, diastolic distensibility and stiffness (Paulus, 1998).

### **2.2.13 Emergency Department Echocardiograms**

While echocardiography is not routine practice in Emergency Departments in the UK, there have been some studies evaluating the utility of this test when seeking specific pathology such as pericardial effusion (Mandavia et al., 2001; Tayal & Kline, 2003). The value of echocardiography in the acute presentation by the Emergency Physician to assist in the diagnosis of heart failure has not been assessed.

A major limitation to the introduction of this investigation is the importance of operator skill. The individual operator will improve with time and experience but will be an issue for the introduction of a new service. Other limitations of this intervention are associated with the patient, either in terms of comorbidities such as atrial fibrillation or emphysema or as a result of body habitus that made the scans more difficult to obtain and interpret. The acute nature of the presenting illness in these patients may also mean that they are unable to assist with the investigation by changing posture or controlling their breathing. Other restraints in using this investigation in the Emergency Department environment included the service pressure that limited the time available to perform this examination.

### **2.2.14 Summary**

This chapter introduced the rationale for the diagnostic tests, findings and strategies that are currently used to make the diagnosis of heart failure. The reported diagnostic utility of these components demonstrates considerable variation in the performance of these variables that may depend on the population studied, the acuity of their symptoms and the skill and experience of the interpreting clinician. The independence of the

available variables, how they perform in isolation or combined with other diagnostic variables has not been assessed.

Although various findings are known to be associated with the syndrome of heart failure, the individual test or combination of tests that can provide a definitive answer early in the patients' journey has yet to be discovered. Scoring systems and diagnostic criteria have been derived but have an emphasis on the diagnosis of chronic heart failure, and so may be of limited value in the acute setting.

The most recent innovation in this diagnostic field has been the discovery of the value of natriuretic peptide levels to aid in the diagnosis of heart failure. As this is a novel investigation, not routinely used in the UK setting, this variable has been described in further detail in the following chapter.

## Chapter 3 The Natriuretic Peptides

### 3.1.1 Introduction

Biomarkers, measurable biological markers of a pathological process, have established a growing role in modern medical practice over the last fifty years. They are used to provide information regarding diagnosis, prognosis and response to treatment for a variety of conditions from ovarian cancer to renal failure (Jacobs & Bast, 1989; Couchoud et al., 1999). For the condition of myocardial infarction, the presence of troponin has a central role in confirming the diagnosis (O'Gara et al., 2013).

Biomarkers have the greatest utility when they can provide information required for clinical decisions. In the condition of the heart failure, where there is a clinical need for early and appropriate treatment but no objective method for a rapid diagnosis, the potential benefits are enormous for any biomarker that can reliably rule in or rule out this syndrome. Collinson et al. (2004) have set out a list of criteria for the ideal biological marker for cardiac disease.

### 3.1.2 Characteristics of an ideal marker – from Collinson (2004)

1. Well characterised
2. Completely cardiac specific
3. Easy to measure accurately and precisely
4. Stable in storage (although not necessary if 'point of care')
5. Not found in normal circulation
6. Clinical role defined

Although several potential biomarkers for heart failure and asymptomatic ventricular dysfunction have been considered, the natriuretic peptides appear the most promising in terms of potential diagnostic value (Jortani

et al., 2004). This is due to the rapid and sustained change in natriuretic peptide concentration in response to the clinical condition of heart failure. It is also relatively stable which makes it easier to measure and there is a degree of correlation with the clinical condition of the patient (Tsuji et al., 1994b).

The natriuretic peptides are well characterised and are mainly produced by the cardiac myocytes, especially in the diseased state, although a small proportion is produced in other vascular tissue. BNP can easily be measured by point-of-care systems and it is known to be relatively stable in stored blood (Wu et al., 2004). While natriuretic peptides are found at low levels in people with normal cardiac function, high levels always suggest a pathological process, although this is not necessarily due to a cardiac condition. A recognised issue with the natriuretic peptides is that they are synthesized and released in response to cardiac strain from any cause, whether this is a result of a primarily cardiac condition or not (Bhalla et al., 2004).

Natriuretic peptides are known to be present in higher concentration in patients with heart failure and therefore have a potential diagnostic role. The cardiac natriuretic peptides consist of atrial natriuretic peptide, brain natriuretic peptide and their associated metabolites. Other natriuretic peptides such as C-type natriuretic peptide and urodilatin are structurally similar and thought to have similar activity but are mainly secreted by non-cardiac tissues. The natriuretic peptides have several effects causing vasodilation, natriuresis and diuresis, and inhibition of the sympathetic nervous system and renin-angiotensin system (Dillingham & Anderson, 1986; Vesely et al., 1994; Clerico, 2002; Klokke et al., 2009; Costello-Boerrigter et al., 2010). They are also likely to have protective effects against the processes of ventricular remodelling and endothelial dysfunction in association with atherosclerosis (Ahluwalia et al., 2006).

Four distinct natriuretic peptides have been described, (A, B, C and D). All have a characteristic 17 amino acid (aa) residue ring formed by an

intra-molecular disulphide bridge between 2 cysteine residues (Suga S, 1992; Daniels, Bhalla, et al., 2006; Bayes-Genis et al., 2007). The amino- and carboxyl terminal tails vary between the different peptides, with BNP having a 32aa tail (Levin et al., 1998) All exist as pro-hormones that are cleaved prior to their release into the circulation (Ozaki et al., 1999; Mikkelsen et al., 2007). A visual representation of the structure of this molecule is provided in Figure 3-1.

The levels of circulating cardiac peptides are altered by a variety of physiological and pathological processes and the levels increase with age, female sex, hypertension, pulmonary hypertension, hyperthyroidism, renal failure and subarachnoid haemorrhage. The plasma concentration of the peptides is also affected by medications such as corticosteroids, angiotensin-converting enzyme (ACE) inhibitors and adrenergic drugs (Clerico, 2004).

### **3.1.3 Atrial Natriuretic Peptide**

Atrial natriuretic peptide (ANP) or atrial natriuretic factor (ANF) was first identified by de Bold et al. (1981) although the natriuretic effects resulting from atrial stretch were first noticed in a canine model by Henry and Pearce (1956). De Bold (1981) showed that the injection of homogenised rat atrial tissue induced diuresis and natriuresis when intravenously administered to rats. The presence of granules within the atrial tissue had been recognised with the advent of electron microscopy but the purpose of these granules had only been postulated upon up until this time. The molecular structure was identified and described by the same author following further research.

ANP is originally synthesised and released as a pro-form that is cleaved, generating a C-terminal ANP and an N-terminal ANP (NT-ANP) (Sundsford et al., 1988). As the inactive N-terminal section has a longer half-life and therefore, a higher plasma concentration, efforts have been

made to assess the diagnostic value of NT-ANP in screening for ventricular hypertrophy and dysfunction. Studies comparing the diagnostic value of this compound with that of the brain natriuretic peptides have consistently shown BNP / NT-proBNP to be superior in this purpose (Yamamoto et al., 1996; Hammerer-Lercher et al., 2001; Vasan et al., 2002; Azzazy & Christenson, 2003; Clerico, 2004).

### 3.1.4 Brain Natriuretic Peptide

Brain natriuretic peptide was first isolated from porcine brain by Sudoh et al. (1988). It was found to be a potent smooth muscle relaxant and also demonstrated hypotensive, natriuretic and diuretic effects when administered intravenously to rats.

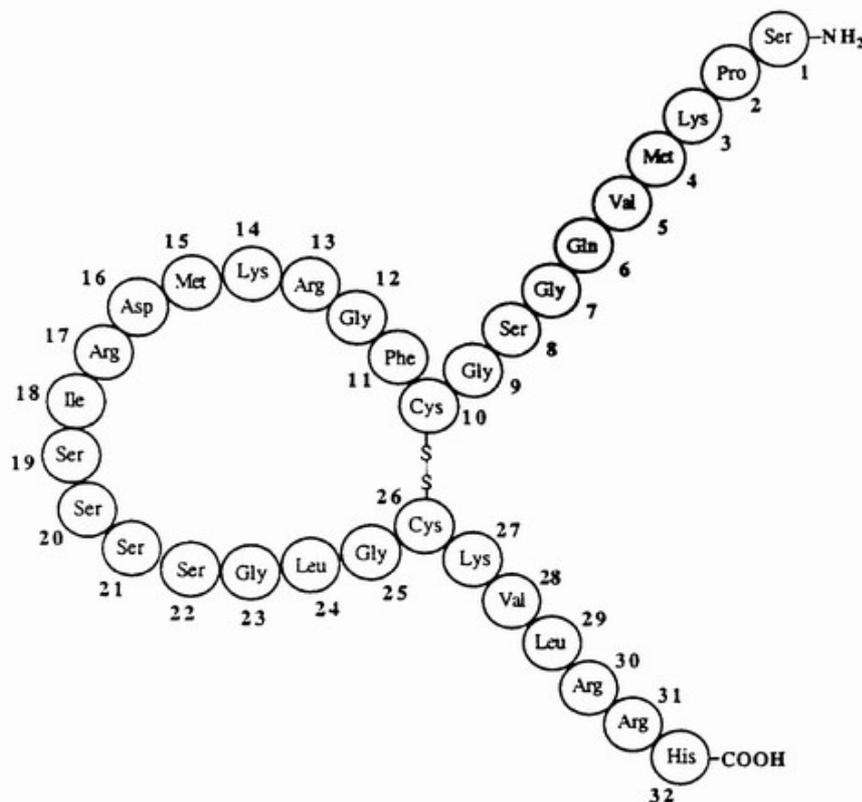


Figure 3-1 Brain Natriuretic Peptide

The human BNP gene is located on chromosome 1 (Tamura et al., 1996; Levin et al., 1998). BNP regulation appears to take place during gene expression. Activation of the BNP gene is required to boost production of BNP; it is not stored in granular form within the myocytes unlike atrial natriuretic peptide (ANP), but is produced de novo in response to cardiac strain. This means that there is not an immediate burst of BNP release in response to an exacerbation of the patient's heart failure but the levels do show a response within hours (Yasue et al., 1994; Silver et al., 2004).

BNP is initially formed as a 134 amino acid precursor known as pre-pro-BNP. Following release of this molecule, it is cleaved by the proteolytic enzyme furin to produce proBNP that consists of 108 amino acids. The compound is released in bursts and is in turn cleaved by a membrane-bound serine protease (corin) to produce the active 32aa BNP which corresponds to the C-terminal sequence and a 76aa N-terminal fragment, NT-proBNP. All 3 components circulate in the plasma and can be detected by immunoassay. The molecule is mainly released by myocytes although there is probably some released from fibroblasts (Tsunda T, 2002; Baxter, 2004). NT-proBNP has a longer half-life than BNP (118 minutes compared to 18 minutes) and is cleared by renal excretion. BNP is mainly cleared by proteolysis by peptidases (Wegner et al., 1996; Ozaki et al., 1999; McCullough et al., 2003; Wu et al., 2004) although BNP levels do correlate with glomerular filtration rates. NT-proBNP levels correlate more closely with renal function than BNP (McCullough et al., 2003; Vesely, 2003; Anwaruddin et al., 2006).

There is a positive correlation between BNP levels and left ventricular (LV) end diastolic pressure and an inverse correlation to LV function. The main stimulus for BNP secretion appears to be cardiac wall stress as ventricular stretch causes increased synthesis and release. There is some suggestion that it may be ischaemia within the ventricle walls that stimulates the BNP release (Maggia J, 1994). Yasue et al. (1994) have demonstrated that the secretion of BNP is increased according to the

level of dysfunction. The peripheral plasma levels reflect the secretion rate from the left ventricle and so allow its use as a marker of left ventricular dysfunction (Herman et al., 1990; Yasue et al., 1994; Scharhag et al., 2007).

Other neurohormones may modulate cardiac BNP production in a paracrine and possibly endocrine way (Harada M 1998). Stimuli such as tachycardia and glucocorticoids also contribute to the induction of cardiac BNP mRNA in overt heart failure (de Lemos JA, 2003; Collinson et al., 2004; Apple et al., 2005; Hammerer-Lercher et al., 2010).

The hormone is synthesised in both the cardiac atria and the ventricles, although it has been shown that the ventricles produce the majority of this compound (Yasue et al., 1994; Boomsma & van den Meiracker, 2001). This appears to be especially true for patients with symptomatic heart failure. Following a myocardial infarction, BNP production appears to be up-regulated to a greater degree than ANP production, possibly due to stretch around the infarcted area (Sumida H, 1995).

In subjects with no evidence of heart failure plasma concentrations of NT-proBNP and BNP are similar. Both peptides are continuously released from the heart and are detectable in normal individuals. As BNP has a short half-life, of around 20 minutes in blood, levels reflect pulmonary capillary wedge pressure (CPWP) changes every two hours. The plasma half-life of NT-proBNP is 120 min and so changes in this level reflect haemodynamic changes over the last 12h. In patients with left ventricular dysfunction, NT-proBNP levels rise relative to BNP levels, with plasma concentrations 2-10 times higher than BNP; the mechanism for this is unknown (Hunt et al., 1995; Morello et al., 2007).

### **3.1.5 The Function of Natriuretic Peptides**

Three forms of natriuretic peptide cell surface receptors (natriuretic peptide receptor NPR-A, NPR-B and NPR-C) mediate the effects of these

molecules. Each receptor has a single trans-membrane domain and an extracellular binding domain. NPR-A is most abundant in the large blood vessels while NPR-B is predominantly found in the brain (Porter et al., 1990; Yamamoto et al., 1996). The receptor is linked to a c-GMP (cyclic guanyl-monophosphate) dependent, signalling cascade that mediates most of biological activity (Azzazy & Christenson, 2003; Yap et al., 2004).

BNP causes direct vasodilatation and increases the permeability of the vascular endothelium allowing fluid to be shifted into the extravascular compartment. It also causes a decrease in sympathetic tone due to suppression of central sympathetic outflow; this decreases the sensitivity of baroreceptors and suppresses catecholamine release from autonomic nerve endings (Vesely, 2003; Apple et al., 2004).

The compound has a direct effect on the kidneys causing an increase in the glomerular filtration rate (GFR) by causing dilatation of the afferent and constriction of the efferent vessels and allowing relaxation of the mesangial cells thereby increasing the surface area for filtration (Weidmann et al., 1986; Hobbs et al., 2002). It also inhibits the actions of angiotensin II therefore preventing sodium and water reabsorption from the proximal tubules. BNP antagonises anti-diuretic hormone (ADH) preventing water absorption in the collecting ducts and blocking sodium reabsorption in the inner medulla (Zeidel et al., 1988; Apple et al., 2005).

The physiological effects of BNP have been studied by injection into intact organism, exposing cells or organs to increased levels, and by the genetic design of mice over-expressing BNP or with BNP gene knockout. These experiments have shown that BNP binds to natriuretic peptide receptor type A causing increased intracellular GMP production (Suga S, 1992). Biological effects of this include diuresis, vasodilatation, the inhibition of renin and aldosterone production and the promotion of cardiac and vascular growth (Ogawa et al., 1995; Vasan et al., 2002).

Natriuretic peptides are broken down after binding to natriuretic peptide receptor C that internalises the peptide before allowing enzymatic degradation. The peptides are also lysed within the circulation by peptidases, the most studied of which is neutral endopeptidase (NEP) (Levin et al., 1998; Campbell et al., 2001)

Table 3-1 Summary of Effects of BNP

Summary of effects of BNP
<ul style="list-style-type: none"> <li>• Vasodilatation</li> </ul>
<ul style="list-style-type: none"> <li>• Shifts intravascular fluid to interstitium</li> </ul>
<ul style="list-style-type: none"> <li>• Inhibits renin release from kidney cells</li> </ul>
<ul style="list-style-type: none"> <li>• Inhibits aldosterone from adrenal cells</li> </ul>
<ul style="list-style-type: none"> <li>• Inhibits sympathetic tone</li> </ul>
<ul style="list-style-type: none"> <li>• May have anti-mitogenic properties to modulate growth in vascular wall disorders</li> </ul>

### 3.1.6 Measuring BNP

Both BNP and NT-proBNP concentrations can be measured by immunoassay; various researchers have examined the potential of measuring the concentration of these molecules as markers of heart failure as described in this section.

Biosite diagnostics introduced an approved test for BNP in 2000. The Biosite BNP assay is a point-of-care assay that uses whole blood or plasma and produces a result in around 15 minutes. It uses a chemiluminescent sandwich immunoassay (Biosite, 2007). There is also a micro-particle based immunoassay for BNP available from Abbot Laboratories (Jones & Kline, 2003; Silver et al., 2004).

The Roche Diagnostics test (Roche Diagnostics GmbH 2010) measures the level of NT-proBNP using an electrochemiluminescent assay. The

Roche NT-proBNP assay uses two polyclonal antibodies aimed at epitopes on residues 1-21 and residues 39-50. One is labelled with biotin and the other with ruthenium complex that bind to NT-proBNP to form a sandwich. Detection is by addition of streptavidin labelled micro-particles. The complex is then bound via biotin-streptavidin interaction. The assay time is approximately 18 minutes. The assays can be run on the Elecsys 1010, 2010 and E170 machines. All three use the same detection technology. The performance accuracy has been calculated as 3.2-2.4% from 175-4962ng/l with an analytical range of 5-35000 ng/l.

Collinson et al. (2004) evaluated the performance characteristics according to the NCCLS protocol EP-5A. Three sample pools were prepared from patients with heart failure with known levels of NT-proBNP and measured repeatedly. The precision profile across the entire analytical range was tested by analysing eleven samples obtained from known cardiac patients and healthy controls measured on each analyser with a target number of replicates of ten.

BNP samples in whole blood or plasma are known to be stable for up to four hours at room temperature and up to twenty-four hours if refrigerated. NT-proBNP samples from plasma are stable for up to three days at room temperature or up to one year if frozen (Tsuji et al., 1994a; Collinson et al., 2004; Apple et al., 2005; Hammerer-Lercher et al., 2010).

Unlike with ANP, there is no detectable rise in BNP in mild exercise so there is no need for the patient to rest for a period before sampling although BNP levels are affected by severe exercise, e.g. marathon running, even in healthy athletes. This probably equates to moderate exercise in a patient with symptomatic heart failure (Anwaruddin et al., 2006; Herrmann et al., 2007; Scharhag et al., 2007). No evidence has been found to demonstrate that BNP levels are affected by posture (Boomsma & van den Meiracker, 2001; Harrison et al., 2002).

There are several physiological aspects and conditions that are known to affect natriuretic peptide levels. Natriuretic peptide concentrations are higher in women than men at all ages (Wieczorek et al., 2002; Wu et al., 2004; Wang et al., 2006). BNP levels are higher in patients with impaired renal function due to decreased renal clearance (McCullough et al., 2003; Anwaruddin et al., 2006; Wang et al., 2012). BNP levels are also higher in patients with atrial fibrillation (Harrison et al., 2002; Morello et al., 2007). BNP levels tend to be lower in obese patients (Struthers, 2000; Daniels, Clopton, et al., 2006; Bayes-Genis et al., 2007). Mariano-Goulart et al. (2003) have shown that the levels of BNP are higher in patients with bilateral ventricular dysfunction compared with isolated left ventricular dysfunction as measured using gated blood pool scintigraphy.

The use of diuretics, vasodilators, Angiotensin Converting Enzyme (ACE) inhibitors, angiotensin II receptor antagonists and spironolactone all decrease BNP concentration. The use of B blockers causes an initial increase in BNP as this removes the adrenergic stimulation that acts to inhibit BNP release. BNP levels are also elevated in patients with chronic respiratory failure with cor pulmonale (Bando et al., 1999; Clerico, 2002). Commencing appropriate therapy for heart failure lowers natriuretic peptide levels (Devillé et al., 2002; Mikkelsen et al., 2007).

BNP levels are also elevated in patients who have had a pulmonary embolism, presumably as a reflection of right ventricular strain. There is some evidence that a lower BNP level predicts a lower risk of adverse outcome in these patients (Kucher et al., 2003). Patients with terminal parenchymal lung disease but with normal left ventricular function do not have raised NT-proBNP levels (Schmidt et al., 1977; Goetze et al., 2004).

Other non-cardiac conditions can also be associated with a raised BNP concentration. For example, Jones and Kline (2003) described three case studies where BNP levels were significantly elevated in patients with sepsis, despite normal cardiac function on echocardiography and absence of clinical evidence of heart failure.

Table 3-2 Summary of Commercially Available Assays (Based on table from Silver et al. (2002))

Vendor	Technology	Marker	Imprecision	Dynamic Range	Cut-off (pg/ml)
Abbot Laboratories	Micro-particle enzyme immunoassay	BNP	Total %CV range 6.5-9.4	0-4000	100
Bayer HealthCare Diagnostics	Direct chemiluminescent sandwich immunoassay	BNP	Total %CV range 2.3-4.7	0-5000	100
Biosite Inc.	Single use fluorescence immunoassay device	BNP	Total %CV range 9.9-12.2	0-5000	100
Biosite Inc.	Two-site chemiluminescent immune-enzymatic assay	BNP	Total %CV range 2.1-6.7	0-5000	100
Roche Diagnostics	Electrochemiluminescent immunoassay	NT-proBNP	Total %CV range 3.6-5.8	0-35,000	<75y 125 >75y >450

### 3.1.7 Optimal Utility of BNP

As BNP levels are known to be higher in patients with ventricular strain, this presents the opportunity to use this compound to screen and diagnose patients where this finding is expected to be present, for example, in cardiac hypertrophy, dysfunction or failure.

The first researchers in this field simply sought to confirm that BNP levels could be used to differentiate those with underlying cardiac disease from those without. Yamamoto et al. (1996) measured BNP levels in 94 patients referred for cardiac catheterisation, and 15 people with no known

cardiac disease and found that a high BNP level was a good predictor of a reduced left ventricular ejection fraction. BNP levels were also demonstrated to be higher in patients with respiratory disease with an element of cor pulmonale compared to the results in patients with isolated pulmonary disease (Bando et al., 1999; Yap et al., 2004).

Groenning et al. (2002) performed a study in 48 patients with a known diagnosis of heart failure and found the cardiac size and function, as measured using magnetic resonance imaging, could be predicted with reasonable accuracy by using NT-proBNP levels. Groenning et al also recruited patients aged 50 to 90 years from four randomly selected general practitioners in Copenhagen to screen for possible heart failure. All of the patients completed a questionnaire covering relevant medical history, had an ECG recorded, had blood pressure and pulse measured and had blood tests, including NT-proBNP taken. To predict patients with a left ventricular ejection fraction of less than or equal to 50%, a cut-off value of 41.5 pmol/l of NT-proBNP had a sensitivity of 70% and a specificity of 63%. The specificity was improved if a lower ejection fraction cut-off was used.

Further research demonstrated that patients with acute decompensated heart failure have higher BNP levels than patients with stable, chronic heart failure (Sterne & Davey Smith, 2001; Logeart et al., 2002; Lainchbury et al., 2003).

Hobbs et al. (2002) investigated the use of NT-proBNP as a screening tool for heart failure. They randomly selected 591 patients over the age of 45 years from four primary care practices. The patients were from four cohorts, the general population, patients with a diagnosis of heart failure, patients on diuretics and patients at high risk of heart failure. All of the patients had a full history, examination and investigations including ECG and echocardiography. Using a cut-off value of 36 pmol/l the NT-proBNP had a sensitivity of 100% and a specificity of 70% in the general population group. In the other groups the sensitivity at this cut-off

remained above 90% but the test was less specific. The authors suggested using further research to investigate which group of patients would benefit from further investigation to confirm heart failure.

Vasan et al. (2002) looked at the value of BNP as a screening tool in the Framingham Offspring cohort study. They evaluated 3177 adult patients with no history of heart failure and measured BNP levels and performed echocardiography in each of the included patients. Patients with renal impairment (creatinine  $\geq 2.0$  mg/dl) or who had inadequate views on echocardiogram were excluded. The authors found that although there was a correlation between BNP and left ventricular systolic dysfunction (LVSD) and left ventricular mass, there was little value demonstrated as a screening test. Cut-off values that provided a reasonably high sensitivity had a very low specificity and vice versa. Adding the BNP result to a model using known clinical risk factors to predict cardiac disease did not significantly improve the model.

Lee et al. (2002) demonstrated that BNP levels correlate well with an objectively assessed NYHA status in patients with heart failure who were managed as outpatients. Wieczorek et al. (2002) performed a larger study of 1050 patients whose left ventricular function varied from normal to severely impaired. They found that the BNP assay becomes more specific with disease severity. Although a direct correlation was found between BNP levels and the degree of ventricular dysfunction it was not possible to predict NYHA class on the basis of the BNP level as the confidence intervals overlapped.

The 'Breathing Not Properly' study was a multi-national, international study looking at the diagnostic utility of BNP in patients presenting to Emergency Departments with acute dyspnoea (Maisel et al., 2002). 1586 patients were recruited and 744 of these were judged to have acute heart failure as the primary cause of their dyspnoea by two cardiologists who reviewed all of the data and acted as the reference standard. BNP had a

diagnostic accuracy of 83.4% at a cut-off of 100pg/ml and a cut-off of 50 pg/ml had a negative predictive value of 96%.

Rogers et al. (2009) performed a retrospective study on patients who had presented to a single centre with acute dyspnoea and had a BNP test performed. The authors suggested that different cut-off values should be used to diagnose heart failure in patients depending on the presence of atrial fibrillation, creatinine levels, BMI and age.

**Table 3-3 Factors Affecting Natriuretic Peptide Levels**

Factor	Effect on BNP levels
Female sex	Higher
Age	Increases with age
Renal impairment	Higher
Obesity	Lower
Atrial fibrillation	Higher
Severe exercise	Higher
Treatment of heart failure	Lower
Right ventricular strain	Higher

### **3.1.8 Comparing BNP with NT-proBNP**

Clerico et al. (2007) performed a literature review after searching for published research assessing and comparing the diagnostic roles of BNP and NT-proBNP. The authors evaluated the data from fifteen studies looking at chronic heart failure and nine studies examining acute heart failure and compared the diagnostic utility using the area under the receiver operator characteristic (ROC) curve (AUC) and the diagnostic odds ratio (DOR). The authors used the random-effects DerSimonian-Laird method to pool the data for the AUC and the DOR for each

biomarker. There were no significant differences between the assay performances for the diagnosis of acute or chronic heart failure.

Emdin et al. (2007) compared NT-proBNP and BNP levels in a prospective cohort of 820 patients, who all had a confirmed diagnosis of heart failure, and 182 healthy controls. The diagnostic utility was assessed by measuring the area under ROC curves for the assessment of the severity of the heart failure, graded from A to D compared with cardiologists blinded to these blood results. NT-proBNP performed marginally better than BNP with a significant improvement in the AUC for patients graded severity C compared to grade A patients or the controls when evaluated against BNP.

Table 3-4 - Characteristics of BNP / NT-proBNP

	BNP	NT-proBNP
Size	32 amino acids	76 amino acids
Clearance	Mainly endopeptidase	Mainly renal
Half-life	15 – 20 min	>1h
Activity	Active	Not active
Stability at room temperature	4h (EDTA tube)	72h

### 3.1.9 Prognostic value of BNP levels

Several studies have looked at the prognostic value of the BNP level, both in patients with heart failure and in patients with other conditions such as sepsis.

Selvais et al. (1998) published a small study looking at the diagnostic and prognostic properties of BNP in patients with heart failure. They found

that patients with a high level of BNP had a significantly higher mortality rate over a follow up period ranging from 18 months to 2 years.

Morrison et al. (2002) looked at a convenience sample of 325 patients presenting with acute dyspnoea. There appears to be an overlap with patients recruited for a larger, multicentre study that recruited patients with this presentation from this site over the same time period (Maisel et al. 2002). The authors followed up the patients at six months, either by clinic review, chart review or telephone interviews. End points were defined as death, hospital admissions and ED attendances with heart failure. A cardiac end point was defined as cardiac death or admission as determined by a cardiologist. The authors used ROC curves to determine the sensitivity and specificity of BNP in predicting the predetermined end points. From their data the authors determined that for the sixty-seven patients with a BNP level greater than 480 pg/ml, the probability of a cardiac event was 51% compared to a cardiac event rate of 2.5% for patients with a BNP level less than 230 pg/ml. The six-month mortality rate attributed to heart failure for patients in the high-risk group was 35%.

Groenning et al. (2004) published a study looking at the diagnostic value of NT-proBNP in screening a randomly selected population in Copenhagen. The authors also examined the prognostic value of this biomarker by following up the recruited patients for a median of 805 (range 60 – 1171) days. Various clinical findings, historical features and investigative results were examined using Cox proportional hazards analyses. NT-proBNP levels were independent prognostic markers for mortality, hospital admission for heart failure and admissions for other cardiac causes. Mortality rates were significantly higher for patients with an NT-proBNP level greater than 32.5pmol/l compared to the group of patients with a lower result.

Apple et al. (2004) enrolled 399 patients with end-stage renal disease (ESRD) who were treated by chronic intermittent haemodialysis in a single city in the US. Various blood tests, including NT-proBNP, were

performed at enrolment and the patients were followed up for two years. 99% of the patients had a raised level of NT-proBNP compared to the general population, however, if the results were split into tertiles then the levels were predictive of mortality. A higher NT-proBNP level was associated with a higher mortality rate.

The REDHOT study was a multi-centre study based in ten hospitals in the US (Maisel et al., 2004). The patients in this study were recruited if they were treated for or admitted with an acute exacerbation of heart failure and had a BNP level greater than 100 pg/ml. Higher levels of BNP (>200pg/ml) were associated with an increased risk of death or congestive heart failure admission or clinic assessment at 90 days.

As part of the Framingham study, Wang et al. (2006) measured ten different biological markers in 3209 participants to examine the predictive value of each of these tests for primary cardiovascular events or death. BNP had the strongest predictive values for death with an adjusted hazard ratio of 1.4 per SD in log values. Levels below the cut-off used to define heart failure but above the mean were still predictive of a higher risk of death during the follow up period.

Anwaruddin et al. (2006) published findings from patients followed up for 60 days after recruitment for the PRIDE study, which looked at the diagnostic value of NT-proBNP for heart failure (Januzzi et al. 2005). NT-proBNP was found to be the strongest predictor of the sixty-day mortality rate with a hazard ratio of 1.61 (95% CI 1.14 – 2.26).

Wang et al. (2012) performed a systematic review and meta-analysis of studies that provided information on BNP levels in patients with sepsis. Analysing 12 studies with a total of 1,865 patients showed a significant correlation between elevated natriuretic peptide levels and mortality with an odds ratio of 8.65 (95% CI 4.94 – 15.13).

### **3.1.10 Therapeutic use of exogenous BNP**

Due to the physiological properties of natriuretic peptides it was postulated that exogenous BNP might benefit patients in acute heart failure. Initial trials showed that intravenous exogenous BNP (Nesiritide©) decreased pulmonary capillary wedge pressure and provided self-reported improvement of dyspnoea at three hours post-treatment (Colucci et al., 2000; VMAC, 2002) The drug received approval for the treatment of acute, decompensated heart failure from the FDA in 2001 on this basis.

There were suggestions from meta-analyses of the small trials that existed that this treatment may be associated with an increased risk of mortality or worsening renal function (Sackner-Bernstein & Aaronson, 2005; Sackner-Bernstein et al., 2005). The first large randomised trial was published in 2011 by O'Connor et al (2011). In this trial 7141 patients with acute, decompensated heart failure were randomised to receive either a Nesiritide infusion or placebo in addition to standard treatment. Overall there was a small and non-significant improvement in outcome in terms of dyspnoea; there was no worsening of renal function; there was a marked increase in the number of patients having symptomatic and asymptomatic episodes of hypotension. The authors did not recommend using Nesiritide in the treatment of heart failure.

It has been estimated that \$1 billion have been spent on Nesiritide in the US healthcare system since the product was licensed (Topol, 2011).

### **3.1.11 Other Diagnostic Roles for BNP**

Campbell et al. (2001) showed that NT-proBNP did not have a useful diagnostic role in detecting acute myocardial ischaemia in patients presenting to an Emergency Department with chest pain.

As BNP levels rise in response to ventricular strain, this compound has a potential prognostic role in the assessment of the severity of pulmonary embolism (Klok et al., 2008).

### **3.1.12 Summary**

This chapter has introduced the natriuretic peptides and explained the potential role that they may have in the diagnosis of heart failure. These molecules are only produced by cardiac tissue and the concentration of these compounds has been shown to increase in proportion to ventricular strain. The natriuretic peptides are relatively stable and simple to measure with commercially available kits providing rapid results. Studies have been performed to evaluate the diagnostic performance of these biomarkers with very promising results. The next chapter examines how the optimum combination of the selected diagnostic variables could be constructed.

## Chapter 4    Diagnostics and Decision Rules

### 4.1.1 Introduction

The successful treatment of patients is reliant on an accurate diagnosis; the earlier the appropriate diagnosis is made, the earlier that the correct treatment can be commenced and the less likely the application of inappropriate treatment.

For some conditions, for example skin wounds, simple recognition is all that is required to make a diagnosis; other conditions require greater consideration. Several factors are known to influence diagnostic decision-making and include knowledge, experience, environment and affective influences (Croskerry, 2005).

Treatments are often started before diagnoses are confirmed if the risk of withholding treatment is considered to outweigh the risk of providing unnecessary treatment. In the hypothetical deductive reasoning path a selection of potential diagnoses are postulated, and then a series of diagnostic hurdles are applied to confirm or refute each of the considered conditions. Some authors have suggested that expert clinicians use a method of 'pattern recognition' (Groen & Patel, 1985; Coderre et al., 2003). This is where the diagnosis is made on the basis of experience of similar cases in the past. This is especially true of conditions such as heart failure where there is no gold standard test and the diagnosis is made by a combination of signs, symptoms and investigative findings (Rutjes et al., 2007).

Current cognitive theory suggests that there are two main processes involved in decision-making: Type 1 thinking is a rapid, heuristic method that is usually correct but is very prone to bias and does not take into account aspects such as prior probability of an event; Type 2 thinking is a

slower, more reasoned method but takes time to perform and is cognitively demanding (Kahnemnan, 2012).

Although, analytical Type 2 consideration would appear to be preferential for all diagnostic decisions, the fast-paced and time-pressured environment of an Emergency Department does not allow for this to be practical.

This is reflected in the fact that most diagnostic errors that occur are not through ignorance, where the physician is not aware that a condition exists, but due to cognitive errors where the diagnosis is not considered (Croskerry, 2013).

The rate of diagnostic errors is difficult to quantify accurately as it is difficult to follow up patients who may improve or attend another healthcare facility. Croskerry (2005) states that data from autopsy results suggests an overall diagnostic error rate of around 40%, and that in a third of these cases an autopsy would not have been necessary had the correct diagnosis been made and treated at the time of presentation.

#### **4.1.2 Definition of Diagnosis**

The Oxford English Dictionary defines diagnosis as ‘the identification of the nature of an illness or other problem by examination of the symptoms’ (Pearsall, 1998;508). Such a definition of diagnosis carries weight in terms of relevant treatment, prognosis and outcome for the patients. Misdiagnosis also carries potential maleficence due to both the withholding of treatment and potential harm from inappropriate treatments intended for other conditions.

#### **4.1.3 The Reference Standard**

For conditions such as heart failure, where there is no simple definitive finding to rule in, or rule out the condition, an accepted reference

standard is a clinical diagnosis made after review of all the available information. The 'Three Wise Men' method, where three experts in a field come to a level of agreement after independently reviewing all of the relevant facts, is accepted practice for creating a fair and pragmatic reference standard for such decisions where no gold standard for the condition exists (Rutjes et al., 2007).

Ideally a reference standard should be applied blind to whichever diagnostic test is being assessed in order to reduce the risk of inclusion bias. The problem with inclusion bias is that assessing a diagnostic test for a condition and using the same test as at least part of the reference standard for the condition leads to a circular argument with a test that has 100% sensitivity and specificity with no capacity for assessing for false positive or false negative results. In this research it has been possible to blind the clinicians to novel tests that would not usually form part of their diagnostic repertoire, but considered unethical to blind them to the results of other tests that they would normally rely upon to make the diagnosis. In this research each diagnostic test will only form part of the overall picture but it is likely that the inclusion bias will remain.

#### **4.1.4 The Measurement of Outcomes**

Outcomes or findings are usually divided into a Boolean choice of true or false. In practice, while this may be true of some conditions, for example fracture to bones, there are many other conditions where there is a continuous spectrum between normal physiological findings and findings considered consistent with a diseased state, e.g. blood pressure. Even conditions as definitive and uncontroversial as death can arouse some debate as to the precise definition (Brenner, 1996). Many medical conditions are part of a spectrum where there is a loosely defined central zone of normality with the extremes of the condition being considered pathological. For example, a pulse rate may be considered abnormal if it is of a rate less than sixty or greater than one hundred beats per minute.

### 4.1.5 Normality

If a certain diagnostic test or finding is to signify the diseased state in an individual, then there must be a corresponding result or range defined as a 'normal' finding. The cut-off of what is normal and what is abnormal can be defined in many ways. For characteristics with Normal distribution, normal values are often defined as within a certain number of standard deviations of the population mean. Results may also be considered abnormal if they directly result in pathological consequences. The cut-off values are usually chosen for practical considerations rather than having distinct biological consequences, e.g. a person may be considered to be hypertensive if their systolic blood pressure is 151mmHg while a patient with a reading of 149mmHg would not require treatment.

This is important for three main reasons:

1. Defining a normal range implicitly means that there is at least one abnormal range; people with results in this range are then considered to be positive for a disease diagnosis.
2. Dichotomising a test into simply positive or negative results may mean excluding useful information. Often there is a scale of probability where a patient with a very high result may have a higher probability of a condition than a patient with a result that is just above the normal range. Choosing a low cut-off value will mean that both patients will test positive and it may be assumed that both patients have the same probability of the disease. Choosing a higher cut-off will ignore the fact that the patient with the lower result still has a higher probability of the disease than a patient with a result in the normal range.
3. For biological variables there are often two cut-off values, and so the relationship with disease should not be assumed to be linear. For example, a very high or a very, low body temperature can signify significant illness in patients.

A positive test for a patient may mean that they are labelled with a disease condition. This may have implications for other areas of their life;

there may be stigmata associated with a condition or it may affect their prospects of receiving a mortgage or insurance. It can also subject them to a further barrage of investigations or instigate life-long treatment. Deciding that a patient is disease-negative may also have consequences by providing them with false reassurance of their risk of a negative outcome and may deny them risk-modifying treatment.

#### **4.1.6 The Diagnosis or 'Not the Diagnosis'**

In this study, the main intention is to rule-in or rule-out heart failure for the presenting patients. By the nature of the inclusion criteria, i.e. adult patients presenting due to acute dyspnoea, the study group will be distinct from the general population. All of the included patients will have some form of underlying pathology. The 'not heart failure' group are likely to include a heterogeneous collection of conditions. A diagnostic test that is successful in differentiating heart failure from pneumonia may be less able to differentiate this condition from pulmonary embolism. There may also be one or more underlying comorbidities. Thus, the diagnostic tests applied to these patients may have variable success in differentiating the individual underlying condition from heart failure due to variety in underlying conditions (Thygesen et al., 2007).

Although it is not the focus of this study, all of the patients will need to have an accurate diagnosis made of the cause of their symptoms so that appropriate treatment can be commenced and the patient can be advised appropriately. It is not sufficient just to inform the patient that they do not have heart failure as an underlying aetiology.

In medical literature, the majority of papers looking at diagnostic testing will concentrate on whether or not a patient has a condition and use binary diagnostic test outcomes such as sensitivity, specificity, likelihood ratios, and positive and negative predictive values. There is limited work looking at the ability of diagnostic procedures to differentiate between two or more alternate diagnoses. As patients present due to symptoms that

are not normally present within the general population, the question is not do they have disease X or no disease, but do they have disease X or disease Y. For example, does a patient have ischaemic vs. haemorrhagic stroke, ulcerative colitis vs. Crohn's disease, heart failure vs. COPD (Brenner, 1996).

There is usually a spectrum of the severity of a disease or condition. It would be reasonable to assume that a patient with severe heart failure, with lots of clinical signs due to long-standing disease, will be easier to diagnose than the patient with asymptomatic left ventricular dysfunction. It is also possible that there may be a synergistic effect when multiple pathologies are present in the same patient; for example, the patient with mild heart failure may become symptomatic due to a sepsis or anaemia.

There is also limited information about the value of combining tests in the diagnosis of disease although this is precisely what we do in practice (Irwig et al., 2006). While it may seem intuitive that two positive tests for a given patient would mean that the diagnosis is more likely, in fact the interpretation of the second test is likely to be less accurate than as the clinician will be biased by the findings of the first test. This end diagnosis decided by a clinician can depend on the basis of the order that the test results become available (Moons & Grobbee, 2002).

Although, in theory, the sensitivity and specificity of a diagnostic test should be independent of the prevalence of the target condition there is some evidence that this not the case (Willis, 2012). This is another potential source of heterogeneity when comparing diagnostic studies performed with different populations.

#### **4.1.7 Diagnosis Within Emergency Medicine**

The setting of the Emergency Department places considerable constraints on the diagnostic process. In the UK setting there is a requirement for 95% of all patients to be discharged or admitted within

four hours of presentation. This limits the diagnostic investigations to those that will return a result within this time. In addition, there is the factor that these patients have presented to the emergency setting because they are acutely unwell and treatment has to be commenced as quickly as possible to improve their condition and outcome, and prevent associated complications.

Additional limitations in this setting arise from the equipment available and the level of seniority of the medical staff that are available at the time of presentation. Investigations that require considerable expertise to perform or interpret, are expensive, take a substantial amount of time or need specialised equipment are unlikely to be available. Another consideration is that most departments aim to provide a consistent standard of care twenty-four hours a day but may rely on investigative services that may subscribe to more traditional office hours.

#### **4.1.8 Measuring Diagnostic Utility**

Diagnostic tests are used to define whether a particular patient does or does not have a particular condition. Assuming that no test is perfect, the diagnostic tests are described by two methods to reflect the test accuracy. The sensitivity records the proportion of patients who have the condition and test positive divided by the total number of patients who have the condition, and the specificity measures the proportion of patients who do not have the condition and test negative over the total number of patients who do not have the condition. Sensitivity is therefore calculated only from the patients who have the condition and specificity is calculated only from the patients without the condition.

Meta-analysis involves a combination of statistical methods in order to summarise the results from more than one study. This can provide information about the accuracy of the diagnostic test and also how it is

affected by other variables. The advantage of meta-analysis is that by combining the data from more than one study can give it greater power and hence make estimates that are closer to true values than may be gleaned from individual studies. This means that the results of meta-analysis should be closer to the true performance in the entire population of the world compared with individual studies looking at different pockets of populations.

Meta-analysis can be beneficial by providing information that can appear greater than the sum of its parts but the risks of providing false information must be guarded against. Where studies include different populations with different prevalence of disease the accuracy of the test may be substantially different and combining these studies may compound evidence in an artificial way.

Meta-analysis of diagnostic studies is more complex than that of intervention studies as there are two summary statistics involved. There is also a threshold effect that links the sensitivity of the test with the specificity. The statistical methods involved in diagnostic test meta-analysis have been designed to account for these factors.

There may also be variation in the reference standard for the disease or condition that the diagnostic test is looking for. If a perfect reference standard existed for a condition then there would be no need to develop new ones, so any new diagnostic test must be judged against the older, established tests with any inaccuracies or limitations that they may encompass. Any variations in the applied reference standard between studies may also be reflected in apparent differences in how the diagnostic test performs.

Receiver Operator Characteristic (ROC) curves are useful for displaying the sensitivity and specificity values of diagnostic tests (Lehr & Pong, 2010). By comparing various studies for the same diagnostic test in the same ROC curve space it is possible to demonstrate if a threshold effect

is present. The threshold effect occurs as reducing the threshold for a positive test result means that more patients will test positive for that condition so sensitivity will increase but specificity, the proportion of patients who don't have the disease who test negative, will decrease. In contrast, raising the positivity threshold will decrease sensitivity and increase specificity. Plotting the corresponding sensitivity and specificity values in ROC space for a test with a variable threshold produces a characteristic shoulder-shaped curve as shown in the figure below.

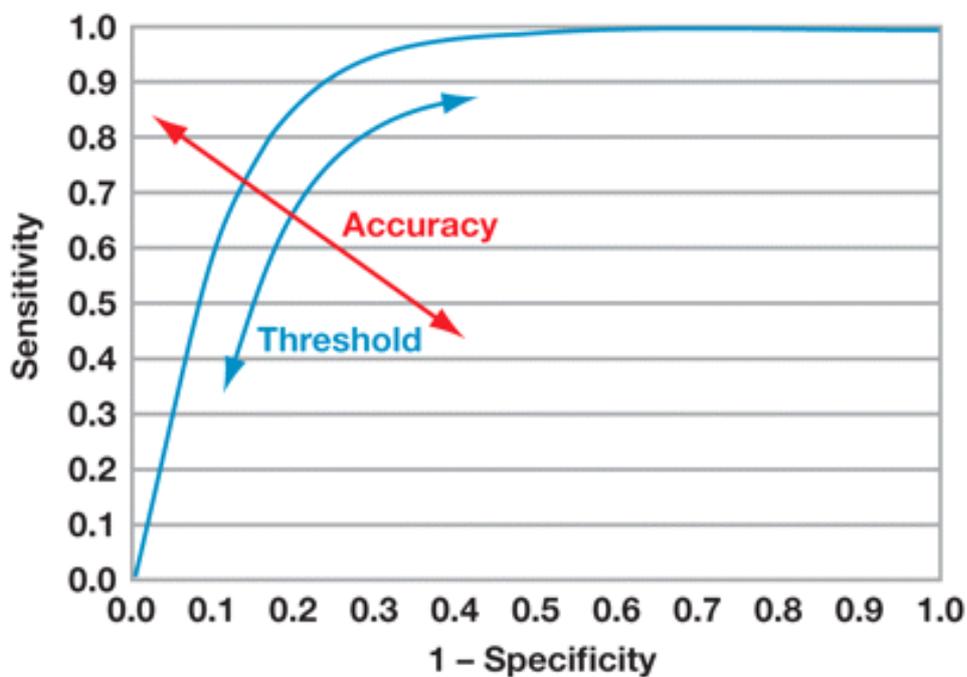


Figure 4- Threshold Effect on Summary ROC Curve (Feinstein, 1990)

The accuracy of different tests can also be compared using ROC curves. This can demonstrate the area under the curve, the optimal diagnostic cut-off levels and the Q\* point where sensitivity equals specificity. The cut-off level can be chosen depending on whether the user wishes to use the diagnostic test to rule-in or rule-out a condition, or choose the optimal mid-point between these values.

Sensitivity analyses are done to explore heterogeneity in the test accuracy. The meta-analysis is performed with all the available studies and then performed again with certain studies missed out on the basis of a defined variable. It may be that the quality of certain studies are dubious or are considered to deal with substantially different populations. This is not the same as sub-group analysis; this is performed to investigate the discovered heterogeneity.

The Cochrane Group advises that meta-analysis involving regression modelling may be useful to investigate how poor methodological quality can lead to bias in results (Macaskill, 2010).

#### **4.1.9 Reporting of Significance of Findings**

In reporting findings in diagnostic studies, the significance of the presented results is often represented by using a p value or probability value. P values are calculated on the understanding that the null hypothesis is correct. It represents the probability of getting a result equal to, or more extreme than was actually observed if the null hypothesis is true and no effect exists. It does not represent a probability that the null hypothesis is incorrect; this can only be calculated by Bayesian methods, using likelihood ratios. A major criticism of the use of p values is that a large effect in a small study or a small effect in a large study can result in the same p value.

Neyman and Pearson (1933) used two hypotheses: the null hypothesis and an effect hypothesis. This brings in type I error, finding an effect that does not exist and type II error, finding no difference although there is one. By concentrating on the p value for a demonstrated effect, this only takes account of type I error and does not address any type II error that may occur. This is in keeping with the ethical belief that it would be worse to give an ineffective treatment with potential side effects and risks than

to fail to find a treatment that exists as the effect is too small to be picked up or it has not been found by chance.

The p value represents the probability of this particular result happening if the null hypothesis is true but does not provide information about the likely mean effect or the distribution of the effect. So if a treatment halves the one-year mortality rate with a p value of .05 then this tells us that the probability of seeing this effect is 1 in 20 if the null hypothesis is true and there is no treatment effect on mortality. This may represent an extreme outcome of a small but true beneficial effect (Schmidt et al., 1977).

Current theories suggest that the optimal method of reporting results is by confidence intervals and likelihood ratios (Goodman, 1999; Sterne & Davey Smith, 2001). The use of confidence intervals suggests significance if the interval does not cross zero and also conveys some idea of the possible distribution of the result.

#### **4.1.10 Models of Decision Making**

Thomas Bayes made the first steps to a scientific understanding of probability with his theorem which linked the prior or pre-test probability of an event or condition with the post-test probability depending on the result of another variable such as a diagnostic test (Price, 1763b). This has been further refined by the derivation of sensitivity and specificity to define the accuracy of diagnostic variables. The understanding of diagnostic tests in this way depends on a binary definition of a disease, i.e. present or absent, and the binary classification of the diagnostic test into positive or negative.

Feinstein (1990) suggests that many clinical decisions are actually trichotomous. The author suggests that the results of a test may be used as a gold standard of diagnostic evidence, as a contributory factor towards a diagnosis or as a surrogate for the diagnosis. In order to calculate Bayesian values such as sensitivity and specificity etc.,

diagnostic variables must have a point defined as true or false. The point where the cut-off is made will hugely influence the sensitivity and specificity of the test, hence the use of ROC curves to demonstrate this and make it easier to choose the optimal value.

Likelihood ratios can also be used to provide information of the utility of diagnostic tests but still rely on cut-offs. Feinstein suggests that in clinical practice people apply diagnostic tests into almost certainly present, almost certainly absent or maybe. The 'maybes' require further investigation. This diagnostic uncertainty can be managed within a clinical decision rule better than by applying a single diagnostic test as the rule can steer the user towards further investigation for the patients who do not have a clear cut diagnosis as a result of their initial investigations.

For most diagnoses, all of the available information is combined, with the positive and negative findings weighed up by the attending physician to formulate a final decision. There is significant scope for variation in the tests performed, the interpretation of the results and the weight attributed to the findings which all account for the differences in opinion that arise when the process is studied (Lok et al., 1998; Joshua et al., 2005).

Medical decision making is known to be a significant source of error and this is particularly true of specialties such as Emergency Medicine where the clinician is confronted with undifferentiated patients in a time-restricted environment (Croskerry, 2013). Solutions to avoid cognitive errors involve changing systems to decrease reliance on memory, relieving time pressures and making tasks easier.

One solution to this is to select the population of interest, the diagnostic procedures to be considered and the order in which they should be performed by synthesising a clinical decision rule (CDR).

CDRs allow a consistent and logical approach to the application of diagnostic testing for conditions where the end-diagnosis is generally

made by combining the results of several variables rather than by one definitive test result (Phillips, 2010). CDRs are structured around the probability of a condition being present in a patient, if specific findings are true or false. Evidence suggests that clinicians vary significantly in their predictions of post-test probability of a condition even if they are provided with information regarding a tests performance (Sox et al., 2009). The introduction of CDRs is one way of encouraging a change in practice to utilise the available evidence in the optimal fashion (Haynes & Haines, 1998).

#### **4.1.11 Methodology**

CDRs are intended as tools to guide physicians in the clinical decision making process. This can take the form of decision trees with each decision node leading to a 'branch' in the tree. They provide either a diagnostic probability or a management plan when applied to a particular clinical situation. They are generally derived through multivariable analysis. As clinical decision rules are designed to directly influence patient care it is important that they are safe and robust. They should be derived from one group of patients and validated on another group of patients, ideally from a different centre, before being applied in mainstream practice in order to reduce the risk of overfitting of the data.

The condition or situation that the decision rule is designed to meet must be common enough for the creation of the rule to be worthwhile. There must be enough confusion, indecision or inconsistency in the management of the condition to provide value for the application of the clinical decision rule. The rule must be applied in a situation where the condition can be confidently diagnosed by the application of simple tests or clinical findings. A cumbersome or impractical rule, or one that does not address a problem that people come across in daily practice is unlikely to be incorporated into clinical practice. It should be perceived as

solving a recognised clinical problem rather than creating further bureaucracy or paperwork or interfering with clinical independence.

Laupacis et al. (1997) define a clinical prediction rule as a decision-making tool for clinicians that includes three or more variables obtained from the history, physical examination or simple diagnostic tests, and that either provides the probability of an outcome or suggests a diagnostic or therapeutic course of action.

Stiell and Wells (1999b) described a CDR as a decision making tool that is derived from original research and incorporates three or more variables from the history, physical examination or simple tests. These authors state that there must be a need for the clinical rule in that it must be a common condition that is not addressed in a consistent manner.

McGinn et al. (2000) describe a CDR as a clinical tool that quantifies the individual contributions that various components of the history, physical examination and basic laboratory results make toward the diagnosis, prognosis or likely response to treatment in an individual patient.

It is generally agreed that the rule itself should be derived according to methodological standards (Stiell & Wells, 1999a). The clinical data should be collected according to well-standardized assessment techniques, carried out prospectively and recorded on specifically designed forms. There should also be a definition of outcome and of reliable predictor variables. The rule should make sense in that it should comply in a consistent manner with what is already known about the condition and there should be a logical order to it. The rule should be sensible.

(Feinstein, 1987) first described the 'sensitivity' in the context of CDR meaning that it should be clinically reasonable, easy to use and provide a course of action. The overall applicability depends on judgement rather than any statistical analysis. CDRs that end in prescribed course of action rather than estimate of risk of condition are probably better followed by treating clinicians.

When developing a CDR, investigators identify a number of variables that they believe will be useful. Many of these are dropped either because it has no predictive value, it has some predictive value on its own but does not add predictive value to the rule that is not provided by the other variables or because assessment of the variable is too unreliable to justify inclusion.

Prospective validation should be completed to ensure robustness of the rule. It should be practical to incorporate it into clinical practice. In order to be of practical use, the application of the CDR by the physician must be able to provide an accurate diagnosis using simple and available tests. The clinical findings or tests need to be consistent. The presence of inter-observer and intra-observer variation should be sought. The outcome measure should be clearly defined and assessed blindly to avoid observation bias.

The study group must be clearly defined so that physicians know exactly who has been involved in the derivation of the rule and which population they should apply the CDR to. The setting of the derivation studies also needs to be clearly described. A derived rule should also be cost effective to apply. If all these factors are present and correct then the message needs to be disseminated in order to promote uptake into clinical practice. Stiell & Wells (1999) have derived the following six-stage guide to developing a clinical decision rule.

*Six Stages in the Development of a Clinical Decision Rule (Stiell & Wells, 1999a)*

- Stage 1. Is there a need for the clinical decision rule?
- Stage 2. Was the rule derived according to methodological standards?
- Stage 3. Has the rule been prospectively validated and refined?
- Stage 4. Has the rule been successfully implemented in clinical practice?
- Stage 5. Would the use of the rule be cost effective?
- Stage 6. How will the rule be disseminated and implemented?

#### **4.1.12 The Risks of Error**

Application of a diagnostic test or decision rule will inevitably apportion the wrong diagnosis to some patients. The dangers of mislabelling these patients will carry immediate risks for them in that they may receive the wrong treatment and fail to receive the correct treatment for their true condition. There is also the risk of 'labelling' a patient as having a condition, an action can have many effects in terms of their treatments and lifestyles and home and relationships. This risk occurs whenever a diagnosis is made and is in no way limited to clinical decision rules. The application of the clinical decision rule should reduce the risks associated with the less experienced clinician by providing clear diagnostic and treatment guidance. The guidance can also reduce the risks of the more experienced clinician who may make omission errors due to overfamiliarity and automation of the management of this condition.

The clinical decision rule can be assessed in terms of its diagnostic accuracy against a reference standard like any other diagnostic test. Minimum acceptable levels of sensitivity or specificity can be selected to ensure sufficient diagnostic utility of a particular model before considering incorporating it into clinical practice.

#### **4.1.13 The 'Art' of Medicine**

The application of an accepted pathway can empower the junior clinician to be confident in the management of the patient with heart failure in terms of treatment and referrals. A theoretical concern with the application of clinical decision rules is that it may remove the 'art' from clinical practice. Research into the utilisation of diagnostic testing in the UK National Health Service shows very high levels of variation on a regional basis; it is unlikely that such variation is consistent with universally applied optimal practice (RightCare, 2013).

By standardising the approach to patients with certain presentations using an agreed pathway there is no need for staff to negotiate with 'gate-keeping' staff for access to the diagnostic test in each individual case. Setting up the diagnostic pathway and demonstrating that it is based on the best available evidence allows the necessary negotiations to be carried out at a senior level. This can reduce the number of tests performed by removing unnecessary tests that do not provide independent new diagnostic information.

#### **4.1.14 Established Clinical Decision Rules**

A number of clinical decision rules have been embraced by the Emergency Medicine specialty following validation demonstrating a consistent, evidence-based approach to different diagnostic quandaries. Examples include the Ottawa Ankle Rule which has been used to standardise radiological investigation of ankle injuries, Well's score which is used to decide which patients need further investigation for possible deep vein thrombosis and the PERC rule which is used to discern when it is appropriate to investigate patients for possible pulmonary embolus (Stiell & Wells, 1999a; Wells et al., 2006; Kline et al., 2010).

#### 4.1.15 Mathematical Techniques to Derive the CDR

Univariate analyses are relatively simple to perform but do not allow exploration of the relationship of predictor values with each other and with the outcome. In order to account for these relationships multivariate statistical approaches such as logistic regression and recursive partitioning are used. Potential variables for the multivariate analysis are usually screened by assessment of the univariate association with the outcome as well as an assessment of the reproducibility of the variable. Variables that demonstrate a significant association with the outcome are selected for further assessment within a multivariable model.

Recursive partitioning analysis progressively divides patients into subpopulations that only include patients with particular outcome. This may be appropriate where wish to develop a decision rule with high sensitivity.

Logistic regression also predicts likelihood of binary outcome (positive or negative) and this is given as log odds that can be used to calculate a simple odds ratio. Logistic regression tends to give decision rules with higher overall accuracy i.e. better overall classification of all patients but possibly with less sensitivity.

Several computer programs have been written to generate decision trees. The programs assess all the provided data and variables to decide the optimum combination of diagnostic variables and cut-off values using three main methods:

- CHAID (Chi-squared automatic Interaction Detection) - uses standard Chi-square tests on cross-tabulation.
- CRT (Classification and Regression Trees) which use a measure of 'impurity' based on the 'gini' coefficient
- C5.0 which uses a measure of information gain based on entropy

The aim is to create decision nodes that are as different as possible and produce sub-populations that are as homogenous as possible. The CHAID method handles missing values as a separate category, CRT replaces missing values with a value from a surrogate variable closely associated with the original variable with the missing value.

Decision trees are not true multivariate models, i.e. they don't simultaneously assess the effect of one variable while controlling for all the others but they use different predictors in different branches and control for the effect of the preceding variables but only for that particular branch of the tree.

End-points can be specified for the decision trees so that the process is stopped before over fitting. The stopping decision can be made on the basis of the number of available predictors, the number of cases left to split, the number of cases in the terminal child nodes or it may be decided to synthesise the tree to a predetermined depth.

#### **4.1.16 Validation of the CDR**

The initial study should be prospectively validated, as many CDRs do not perform as well as predicted when applied in other settings. The reason for this may be statistical, due to over-fitting or instability in the original model, or may be due to differences in prevalence of disease, the threshold of diagnostic criteria or differences in how the CDR is applied.

Bootstrapping involves the removal of the data from a single patient from the study, deriving the rule from the rest of the patients and then applying this to the first patient and seeing if the rule holds true. This process is then repeated in sequence for each patient in the study. This can be used as a form of internal validation but is inferior to an independent, prospective validation study.

Prospective validation should be in new patient group, ideally by a different set of clinicians in order to demonstrate the robustness of the CDR. Clinicians using the CDR should understand the procedures and be confident in the application of the rule. The overall accuracy of the rule in the new population should be mapped out in a two-by-two matrix with relevant calculations and compared with the original study data.

There should also be measurement of the inter-observer agreement regarding the outcome of the CDR providing a value for reliability. Centres applying the CDR should be observed to see how effective the CDR is in practice, whether it is indeed adopted and how it changes practice over a period of time.

#### **4.1.17 Application to this Study**

The diagnosis of heart failure is complex despite this being a common condition there is considerable variation in clinical practice. The end-diagnosis is derived from a selection of clinical findings and diagnostic test results. Patients with this condition would benefit from the derivation of a clinical decision rule that allowed a rapid and accurate diagnosis.

Although many symptoms, signs and investigative findings are considered to have diagnostic value in defining this condition, the true diagnostic weight of each of these elements and the combination of these findings remains poorly understood. Early and appropriate treatment improves symptomatic relief and is likely to improve outcomes and reduce associated health care utilisation and cost.

In order for such a rule to be accepted it would have to be seen as making clinical sense, and to perform well in the environment of Emergency Medicine. It must be reproducible in application by a mutable workforce of variable seniority and experience, and harness available diagnostic features that can be interpreted reliably.

To demonstrate the reliability and reproducibility of any decision rule derived from this study, further validation studies would have to be performed with different populations of patients, ideally in independent centres. Following this, dissemination of a validated clinical decision rule would best be facilitated by publication in peer-reviewed journals.

Having established the potential application for an effective CDR for the diagnosis of heart failure, the next stage in the synthesis of the decision rule is the selection of predictor variables. The following chapters present a systematic review and meta-analysis of diagnostic studies in patients presenting to emergency department with symptoms suggestive of heart failure. The intention was to assess the utility of all potential predictor variables and then use meta-analytic techniques to select those with the optimal diagnostic value for consideration in the CDR.

# Chapter 5    Meta-analysis of Diagnostic Studies

## 5.1.1 Introduction

In order to select which variables have the greatest diagnostic utility, and to be able to define summary statistics of their diagnostic performance, meta-analysis of the potential predictor variables from the studies included in the systematic review was necessary.

Meta-analysis is essentially the analysis of analyses. Glass (1976) coined the term in 1976, although the concept and process existed before this. The intentions behind meta-analysis are two-fold: the first intention is to enable a summary or review of various studies of the same subject, with a method other than that of a simple narrative; the second intention is to produce an answer to a clinical dilemma which may be greater than the sum of its parts. Meta-analysis may also facilitate explanation of observed heterogeneity between different studies (Trikalinos et al., 2012).

Meta-analysis is the aggregation of results from multiple studies. Each study has 'weight' corresponding to its precision, which is estimated by the variance. The weight of a study is equivalent to one under the power of the variance. The general formula for meta-analysis the global outcome (D) expressed as a weighted mean.

For interventional studies the size of effect of the selected intervention can be measured by a chosen outcome. The outcomes values of multiple studies can then be combined while weighting the results according to relevant factors, for example, the size of the study. Meta-analysis methods for looking at diagnostic tests, rather than treatments, are less well established in the literature. The rationale, however, is similar: to summarise the information in a valid way and to provide information on the factors affecting the estimates. The validity of the estimates is

dependent on the absence of publication bias, the absence of bias due to poor quality of the selected studies and that belief that the meta-analysis estimates are generalisable in a useful way.

In diagnostic studies or screening studies the outcome is more complex than interventional studies as there are two components to consider, namely sensitivity and specificity. Moreover, sensitivity and specificity are not independent of each other but usually have a negative correlation, whereby increasing the threshold of a diagnostic test increases the sensitivity but decreases the specificity of the test. (Jones & Athansiou, 2009; Sousa & Ribeiro, 2009; Willis & Quigley, 2011; Begum et al., 2012). This is described as a threshold effect. The threshold may be explicit, where different cut-off values are chosen deliberately, or implicit, where different cut-offs are applied due to some degree of subjectivity or there is an unrecognised variability in the way that the test is performed.

Meta-analysis of diagnostic studies is a developing field and there is no unanimous agreement as to the best method to use (Riley et al., 2007). Although it is possible to perform meta-analysis on the sensitivity and specificity of the test individually, there are theoretical reasons and modelled examples that demonstrate better results when a combined method is performed (Leeflang, Moons, et al., 2008).

### **5.1.2 Threshold Effect**

Most diagnostic tests will have a positivity threshold to dichotomise patients into condition positive or condition negative. The threshold may be explicit or latent. For example one study may classify patients over 185cm in height as tall, while another study may ask the attending physician to choose tall patients; the explicit threshold should mean greater consistency but this is not always possible when assessing patients for more subjective findings, for example, pulmonary oedema. There may also be an implicit threshold where the difference in applied thresholds is not apparent between different studies (Harbord et al.,

2008). Reducing a positivity threshold will tend to increase the number of patients who test positive, increasing the true positive and false positive proportions. This, in turn, increases the sensitivity but decreases the specificity of the test.

It is possible to use a Spearman correlation coefficient to examine the relationship between sensitivity and specificity for all of the included studies. In studies where the threshold exists, the diagnostic odds ratio will remain approximately constant for all the given sensitivity and specificity values and there will be a strong negative correlation between them. When plotted on a ROC plane, a threshold effect will provide a symmetrical, curvilinear line (Irwig et al., 2002). In examining the values of the selected studies for the presence of a threshold effect, it is important to observe how far the plotted study points are from the summary ROC curve rather just than how far they are from each other.

### **5.1.3 Heterogeneity**

Statistical heterogeneity, usually referred to as simply heterogeneity, is defined as the variation between studies that is considered to be beyond random effect. It results from the differences in study design, conduct, patients, procedures, outcome measurements and other difference that may be difficult to predict or measure. Heterogeneity is a recognised feature of systematic review of diagnostic studies as between-study heterogeneity is widespread for all measures of diagnostic accuracy (Higgins & Thompson, 2002). It is important to have an idea of the level of heterogeneity present in order to understand the influence of this on the results of the meta-analysis (Deville et al., 2002). The presence of heterogeneity can be tested for statistically using a chi-square test, or Fisher's exact test for smaller numbers of studies. The power of this test tends to be low (Deeks, 2001).

An impression of the level of heterogeneity present can be obtained by examining the appearance of the summary ROC curve. The presence of

a threshold effect affects the diagnostic odds ratios to the extent that a smooth curve will be seen, often described as resembling a shoulder (Cochran, 1954). Studies with marked differences in the study populations or results will be plotted far from the ROC curves. The basic method of examining a forest plot of sensitivity and specificity with 95% confidence intervals is also a reasonable method of assessing heterogeneity.

Estimating the between-study variance is done as part of a random effects meta-analysis. The variance can be used to describe the extent of variability but can be difficult to interpret as it is quantified on the scale of the log-odds ratio.

Cochran's  $Q$  test, also known as Cochran's chi-square test, is a way of indicating the extent of heterogeneity with a  $p$ -value indicating the extent of variability between studies (Hardy & Thompson, 1998). This test does have some limitations: it has a low statistical power when meta-analyses include few studies and it may also detect clinically unimportant heterogeneity in the presence of many studies (Higgins & Thompson, 2002).

$H$  is the square root of the chi-square for heterogeneity statistics divided by its degrees of freedom. It describes the relative excess in  $Q$  over its degrees of freedom.  $R$  is the ratio of the standard error of the mean from the random effect meta-analysis to the standard error of a fixed effect meta-analytic estimate. It describes the inflation in the confidence interval for a summary estimate under a random effect model compared with a fixed effect model.

An inconsistency ( $I^2$ ) score is a measure of the variation across studies due to their heterogeneity. Higgins and Thompson (2002) described  $I^2$  as the percentage of total variability in the estimates that is due to heterogeneity between studies rather than sampling error (Ioannidis et al., 2008).  $I^2$  is a transformation of the  $H$  that describes the proportion of

total variation in study estimate that is due to heterogeneity, that is, the variation that is not due to chance.  $I^2$  is given as a value between 0 and 100%. This number also reflects the number of studies within the analysis. Confidence intervals can be calculated for this variable.

An advantage of  $I^2$  over  $Q$  is that it can allow comparisons between meta-analyses of different sizes. It has been suggested that  $I^2$  values of 25%, 50% and 75% are evidence of low, moderate and high heterogeneity, respectively. However, while the  $I^2$  index and its 95% confidence intervals are helpful in expressing quantitative values, decisions should not be based entirely on these results (DerSimonian & Laird, 1986). Ioannidis et al. (2008) recommend that the interpretation of heterogeneity in a systematic review still needs consideration of the clinical and methodological diversity between studies.

#### **5.1.4 Random Effects Model**

The procedures in the calculation can either use fixed effect models or random effect models. Fixed effect models assume that the available studies all give an estimate of the same effect so that the estimated effects can all be considered as part of the same distribution. This method is only reliable for models with low heterogeneity.

The random effects model does not assume that all of the patients in the selected studies come from the same population. Each study is considered to have its own population, and each population, its own means. Therefore, variability of the estimate may have two sources, variation within the study and variation between different studies. The DerSimonian-Laird method is used to calculate this (Moses et al., 1993).

If the test for statistical heterogeneity has a significant result then the random effects model should be used. Some statisticians argue that the random effects model should always be used in meta-analysis, as in the absence of significant heterogeneity, the results are almost identical to

fixed effect model, while there is general agreement that the fixed effect model should not be used in the presence of significant heterogeneity. Heterogeneity is to be expected in the results of test accuracy studies thus random effects models are required to optimally describe the variability in the test accuracy.

### 5.1.5 Methods of Meta-analysis of Diagnostic Studies

Moses et al. (1993) proposed a method of meta-analysis that has received some acceptance in the field of medical statistics (Irwig et al., 2002). Their approach utilises diagnostic odds ratios (DOR) as a method of unifying the sensitivity and specificity that can be used even in the presence of a threshold effect. The DOR represents the odds of a positive test result in patients with a target condition compared to the odds of a positive test result in patients without the condition. The DOR expresses how much greater the odds of having the disease are for the people with a positive test result than for the people with a negative test result. It is a single measure of diagnostic test performance that combines both likelihood ratios. Using the log of the odds or the logit values of the DOR it is possible to construct a line describing the relationship between these values and then transform the values back to create a summary Receiver Operating Characteristic (ROC) curve (Glas et al., 2003). This curve describes the diagnostic odds ratio (DOR) over a range of corresponding sensitivity and specificity values. The DOR is calculated by the following equation:

$$DOR = \frac{TP}{FN} + \frac{FP}{TN}$$

The DOR is a useful indicator of a diagnostic tests utility as it is a single value, making it easier to visualise and perform statistical calculations. The disadvantage is that it is not possible to weight for true positive and false positive results separately{Moses 1993}. The first step is to convert the true positive results (TPR) and false positive results (FPR) to their

logistic transformations. The logit of the TPR is the logarithm of the TPR divided by one minus the TPR. The logit of the FPR is calculated in the same manner. To account for any cells that contain zero or one, 0.5 is added to all the cells in the two by two tables for each study. Two equations are involved:

$$S = \text{logit (TPR)} + \text{logit (FPR)}$$

$$D = \text{logit (TPR)} - \text{logit (FPR)}.$$

S is the sum of the two transforms and is related to how often the test is positive, which is related to the test threshold. D is the difference between the two transforms and is a measure of how well the test discriminates between the two populations of disease positive and disease negative patients.

Generally, TPR and FPR are inversely related. The main factor affecting these variables is the threshold. The relationship between D and S is estimated using a linear model (least squares regression):

$$D = bS + i$$

Once the slope and intercept of the transformed line is known it can be back-transformed. Since  $D_k$  is the log-odds ratio of the  $k$ th study, a summary of D (by equally weighted mean, weighted mean or median) can provide reasonable estimate of the power of the test.

As this method involves regression analysis it is susceptible to influence from extreme outliers. If a test is being used with the intention of confirming that a patient really has a condition, then the area of interest will be the part of the curve with high specificity. If some included studies involve results with low specificity and high sensitivity this may influence the data even though that cut-off point has little to do with how the test would be used in clinical practice. Moses et al. (1993) therefore recommend that only studies that lie within a clinically determined

“relevant range” should be used (Deeks, 2001). The authors also recommend that the curve not be extended beyond the plotted study points. They concede that reducing the size of the curve in this way will inflate the apparent accuracy of the index test but state that the bias is not large.

The area under the curve (AUC) is equal to the probability that if two individuals are chosen at random and one has the condition and the other doesn't have the condition, the diseased individual will have a higher test result than the non-diseased. It is also an average sensitivity over the entire range of specificity and vice versa. In symmetric ROC curves all points have a common DOR. In asymmetric curves the DOR varies with the threshold.

In a summary ROC curve, individual studies are plotted in ROC space. In this study the plotted studies have been scaled by sample size with the height representing the number of patients with disease (precision sensitivity) and the width representing the number of patients without the disease (precision specificity).

Unfortunately there are some limitations of this technique. This method does not supply summary values for sensitivity or specificity. It is possible to calculate the sensitivity and specificity for any point on the ROC curve that is produced, but these values may not be represented in the original data (Gatsonis & Paliwal, 2006; Leeflang, Deeks, et al., 2008). There are also assumptions made about the relationship between the sensitivity and the specificity that may not hold true. The threshold effect may be negated to some degree by selecting the point where the ROC curve intersects with a line drawn to illustrate where the sensitivity is equal to specificity but this approach is only appropriate for homogenous data (Walter, 2002). This technique also fails to allow for variation in the prevalence of the disease in the included studies (Leeflang, Moons, et al., 2008).

Other criticisms include the fact that sampling variability in the individual studies is not taken into account and that there is an underestimation of test accuracy due to zero cell corrections (Moses et al., 1993). Zero cell corrections, where 0.5 is added to every cell, are necessary so that logarithmic calculations can be performed even if there are no patients in one of the categories without causing an error.

### **5.1.6 Bivariate Analysis**

An alternative method of meta-analysis has been described by Kardaun and Kardaun (1990). Bivariate analysis overcomes some of the shortcomings of the method described by Moses et al. (1993) (Reitsma et al., 2005). In the method described by Kardaun and Kardaun (1990) the assumption is made that there is Normal distribution of both the sensitivity and specificity around a mean, following logit transformation. A random effects model is used to incorporate unexplained variability in the analysis. It is assumed that there is a negative correlation between the two values and this is incorporated into the analysis as a single model. The combination of two normally distributed outcomes, the logit transformed sensitivities and specificities and the acknowledgement of a possible correlation between them leads to bivariate normal distribution. The studies are weighted so that those with a more precise estimate of sensitivity or specificity are given a higher weight in the analysis. The bivariate model is then analysed using linear mixed model techniques.

One advantage of this method is that it provides summary values of sensitivity and specificity together with confidence intervals (Lau et al., 1998). Covariates can be added to the bivariate model to allow further analysis. There are issues with using a random effects model, it does not explain any heterogeneity that is present and it may provide a more conservative level of statistical significance than fixed effect models (Trikalinos et al., 2012).

The distributions of the logarithm of the likelihood ratios are approximately normal and so it is possible to calculate confidence intervals for these values. The distribution for the diagnostic odds ratio is also approximately normal and the standard error and confidence interval of the DOR can be calculated.

### **5.1.7 Hierarchical Summary ROC Curve Method**

A third model for meta-analysis of diagnostic tests is the Hierarchical Summary ROC curve method (HSROC). This method is similar to the bivariate method but instead of allowing for the variability of the sensitivity and specificity of the test includes parameters for the scale and accuracy of the studies. This method produces identical results to the bivariate method unless covariates are added to explore between-study heterogeneity. While both methods allow the introduction of covariates to examine the effects of other factors on the diagnostic performance of the test under examination the bivariate method allows the direct evaluation of the effect on the sensitivity or specificity while the HSROC method allows evaluation of the effect on accuracy or threshold parameters.

### **5.1.8 Likelihood Ratios**

Likelihood ratios are a useful tool to describe the application of diagnostic tests to a given population. A positive likelihood ratio is calculated by dividing the proportion of patients with a positive result who really have a condition (true positive) by the proportion of patients with positive results who do not have the condition (false positive) (Sackett et al., 2002). This can also be represented by dividing the sensitivity of a test by one minus the specificity of the test.

A negative likelihood ratio is the proportion of patients with a false negative result over the proportion of patients with a true negative result. This is the equivalent of one minus the sensitivity divided by the specificity.

Likelihood ratios have the advantage of being independent of the prevalence of the disease condition. They can be used to calculate the post-test odds for a given population from pre-test odds when a diagnostic test is applied. The only issue with their use is the need to convert the probability of a condition into the odds of a condition before applying the likelihood ratio. Odds represent the ratio of a positive result against a negative result, while probability represents the proportion of positive results over the total population. Probability is a more intuitive method of description though odds have more useful mathematical properties.

Use of a Fagan nomogram makes working with likelihood ratios easier to use as approximate values can be obtained simply by drawing a straight line through the pre-test probability and the likelihood ratio to read off an approximate post-test probability (Fagan, 1975).

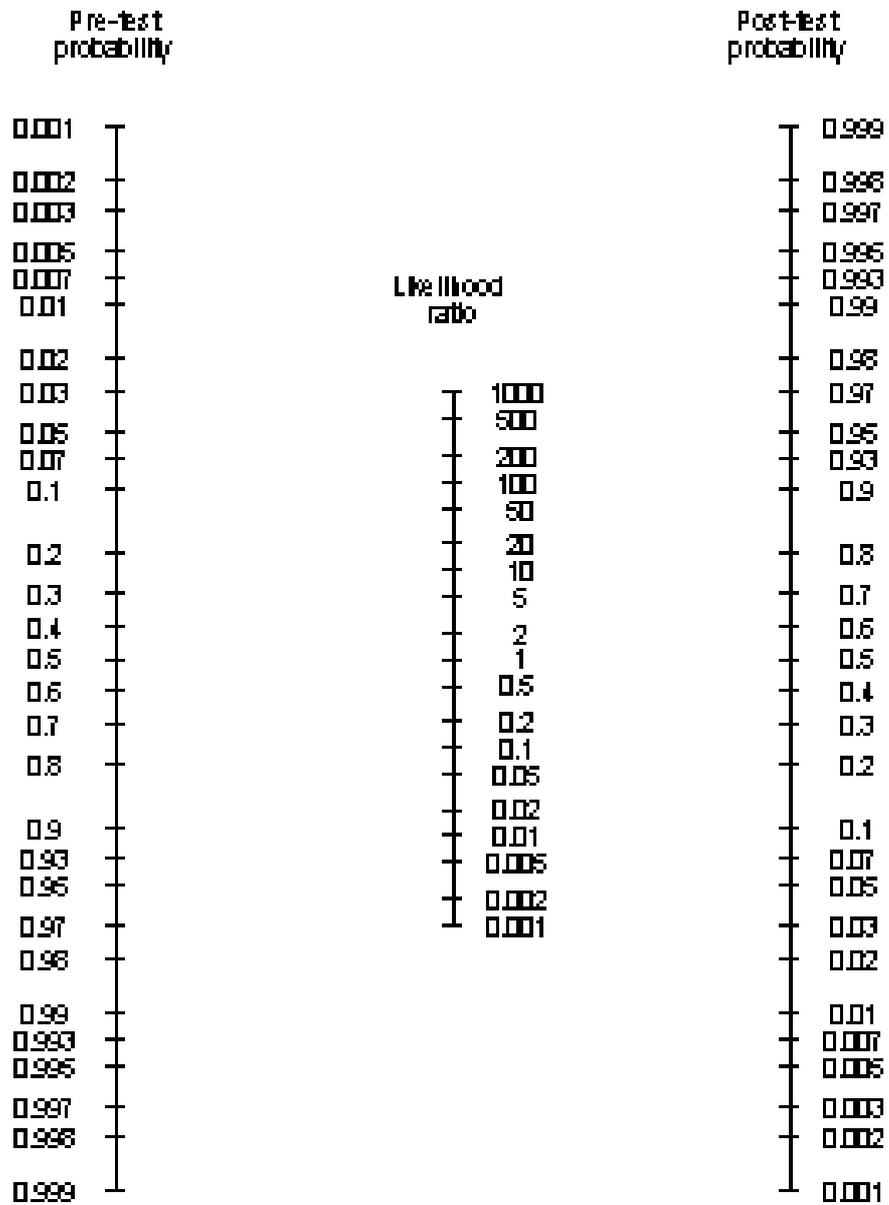


Figure 5- Fagan Nomogram (By Mikael Häggström [Public domain], via Wikimedia Commons)

### 5.1.9 Selected Methods of Meta-analysis

The first stage of the meta-analysis was the calculation of the summary statistics for the selected predictor variable from each study that provided

information about it. The sensitivity and specificity were calculated for that diagnostic variable. Forest plots were compiled providing graphical representation of the tests performance as the first stage in the examination of the data. This allowed visual assessment of the degree of variability between studies. The confidence intervals were also calculated and displayed for the sensitivity and specificity.

Positive and negative likelihood ratios (LR) were derived from the sensitivity and specificity.

$$LR+ = \text{Sen} / (1 - \text{Spec}).$$

$$LR- = (1 - \text{Sen}) / \text{Spec}.$$

Likelihood ratios express how much more frequent the respective result is among subjects with disease than among patients without disease. These were used to calculate the diagnostic odds ratio (DOR).

$$DOR = LR+ / LR-$$

The DOR expresses how much greater the odds of having the disease are for the people with a positive test result than for the people with a negative test result. It is a single measure of diagnostic test performance that combines both LRs and so incorporates both sensitivity and specificity.

There is an important extra source of variation in diagnostic studies, as compared with treatment studies, and that is the use of different thresholds to define positive and negative results. To explore this source of variation the sensitivity and specificity were plotted on a ROC plane. In the presence of a threshold effect it is expected that the points will show a curvilinear pattern. Calculating the Spearman correlation coefficient between sensitivity and specificity can also test for a threshold effect. If the threshold effect exists an inverse correlation appears (Devillé et al., 2002; Bewick et al., 2004).

For the next stage, overall test accuracy indexes were calculated as the weighted average of these summary statistics. These were plotted in ROC space using the MetaAnalyst software. In the presence of considerable heterogeneity between the studies the differences were reported rather than providing a pooled estimate. Where there were studies that were obvious outliers, reasons for the heterogeneity were sought. In the presence of marked heterogeneity it is reasonable to assume that they are substantial differences in the population or the diagnostic test between the included studies. These studies were excluded from the meta-analysis and the effects of their exclusion on the results examined.

For variables where the meta-analysis was considered appropriate, in that there were sufficient numbers of studies available that displayed a reasonable level of homogeneity or heterogeneity attributable to a threshold effect then meta-analysis was carried out. MetaDisc© software was used to summarise the data from each study in a Summary ROC Curve. This allowed the exploration the homogeneity of the studies graphically and statistically, and computation of the summary statistics and their confidence intervals.

In the presence of significant diagnostic threshold variation among studies, the best summary of study results may be a ROC curve rather than a single point. The shape of the ROC curve depends on the underlying distribution of test results in patients with and without the disease.

The confidence intervals of sensitivity and specificity were calculated using the F distribution method to compute the exact confidence limits for the binomial proportion (Leemis & Trivedi, 1996; Friedman & Popescu, 2003).

As meta-analyses often include small numbers of studies, the power of both tests  $G^2$  and  $Q$ , is low so they are poor at detecting true

heterogeneity among studies as significant. An alternative approach is to quantify the effect of heterogeneity is the  $I^2$  index that describes the percentage of total variation across studies that is due to heterogeneity rather than chance (Higgins, 2003; Bewick et al., 2004).

#### **5.1.10 Methods Used**

MetaDiSc 1.4 software was used to examine the heterogeneity of the selected studies. The number of patients included, the prevalence of the target condition and the sensitivity and specificity for each variable in each study were entered allowing the calculation of the Spearman coefficient and  $I^2$  values (Zamora et al., 2006). The bivariate analysis was performed using Meta-Analyst software (Wallace et al., 2009). This provided the overall summary point for each selected variable and the data regarding logit values for sensitivity, specificity and the covariate were entered back into the RevMan software to produce the summary ROC curve and summary point figure for each selected variable.

#### **5.1.11 Summary**

This chapter explained the statistical methods used for the meta-analysis of the diagnostic variables taken from the systematic review. There was also an exploration of alternative methods and justification for the methods chosen. The following chapter presents the results of the systematic review and the assessment of all of the possible diagnostic variables gleaned from the literature.

## Chapter 6 Systematic Review and Meta-analysis

### 6.1.1 Scope

In order to select the optimum variables for the synthesis of the clinical decision rule an exhaustive search of the relevant literature was performed. The intention was to find all available data pertaining to the question: 'For adult patients presenting to an emergency setting, with acute dyspnoea, which diagnostic tests can be used to confirm or refute this diagnosis?'

The methods for the systematic review were in accordance with current guidance. An open literature search was performed to minimise the risk of missing relevant studies. Pertinent studies were included and critically appraised to assess their quality. Any variables that were included in the selected studies that may have some utility in the diagnosis of heart failure were examined, and the relevant data extracted. Data from each study was used to compile two-by-two tables in order to calculate the sensitivity and specificity and this was combined for each variable to assess its diagnostic utility. The data was then examined further by creating forest plots and summary ROC curves for each group of variables. Variables that appeared to have reasonable diagnostic value were further scrutinised, and where appropriate, meta-analysis was performed in order to summarise the overall diagnostic utility for that particular variable.

This chapter summarises the literature search, the selection process for relevant diagnostic papers, the quality of the selected papers and the summary data for the selected variables from the studies. The selected variables are grouped by their diagnostic utility and a summary of the findings is presented at the end of the chapter.

### 6.1.2 Literature Search

A search was originally performed in August 2009 of the Medline and EMBASE databases using Ovid interface. The literature search was optimised in order to minimise the risk of missing relevant studies. The literature search is detailed in appendix 1. Filters were not applied to minimise the risk of excluding relevant studies (Wilczynski & Haynes, 2005; Leeflang, Deeks, et al., 2008). Two readers (KMJ and CJF) assessed all the titles and abstracts and all relevant papers were accessed and read in detail by CJF. The literature search was repeated in January 2014 by CJF to look for further data.

Additional searches were carried out using GoogleScholar © using the terms “heart fail\$” and “diagnos\$” looking at the first one hundred hits.

The references of selected papers were also examined to look for any further relevant studies. Lead authors were contacted by email to ask if there was any further unpublished data available. Grey literature was also searched for using the DevonAgent © software with the search strategy: (“heart failure” OR adhf) AND (acute OR new OR decompensated) AND diagnos\*  
258 unique results were obtained from 7064 results.

Two further grey literature websites were also checked: [www.mediondatabase.nl](http://www.mediondatabase.nl) and the North West Grey Literature website ([www.fade.nhs.uk](http://www.fade.nhs.uk)).

The papers were critically appraised to assess quality and the data extracted and entered into the Review Manager (RevMan 5.1) software downloaded from the Cochrane collaboration (Cochrane, 2012). This

software provides a structure for quality assessment, data reporting and basic data analysis.

6909 papers were identified using the EMBASE search and 5633 papers were identified using the OVID search in 2009. In the repeat search in 2014, EMBASE and OVID were searched in combination. This identified 21,852 papers that were searched back until 2008. From this selection forty-nine papers were included and forty-one papers were excluded. The included studies are detailed in Section 6.1.12 and summarised in Appendix IV; the excluded papers are detailed, with the reasons for exclusion in Section 6.1.14.

A diagram of the format recommended by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) group shows the results of the literature search in Figure 6-1.

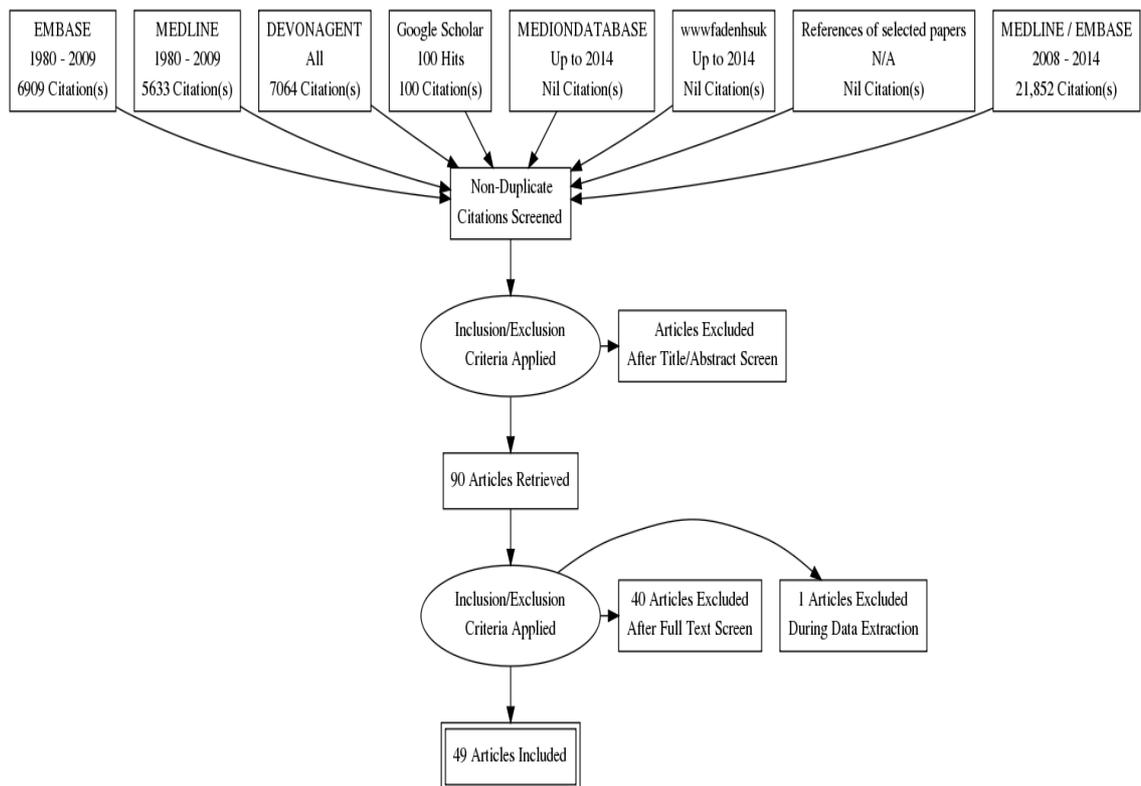


Figure 6-1 PRISMA Diagram of Literature Search

### 6.1.3 Risk of Bias

The New Oxford Dictionary of English defines 'bias' in a statistical context as 'a systematic distortion of a statistical result due to a factor not allowed for in its derivation' (Pearsall, 1998;169). Given the number of studies, the number of patients, the number of settings and the number of variables involved in this analysis it is inevitable that some degree of bias will be present. The forms of bias considered relevant for this study have been detailed in Table 6-1. While some degree of bias is accepted as inevitable, steps have been taken to minimise the effects whenever possible, and to at least openly acknowledge its presence in other areas.

While it is very likely that publication bias, where studies with positive findings are more likely to be published than those with negative findings, there are no universally accepted or validated methods to assess for publication bias in diagnostic studies (Irwig et al., 2002). Research has suggested that smaller studies are more likely to have more positive results and that the best method of reducing publication bias in diagnostic studies is by including as many studies as possible (Song et al., 2002).

An attempt has been made to avoid selection bias by involving two people in the selection process and defining clear, transparent inclusion criteria prior to beginning the search. The possibility of extractor bias has been reduced by having the process of data recording repeated independently by a second reader for each included study. Arrangements were made for a further independent reader to assess the literature in the event of any disagreement between these two readers.

The inclusion dates and settings for each of the studies were examined and cross-referenced to try to avoid multiple publication bias and multiple use of the same subject bias.

Reporting bias, where the authors look at many variables but only report the optimal results, may also be present (Macaskill, 2010).

Recording error bias, where data in the original study has been accidentally recorded incorrectly is unavoidable given the number of studies and patients involved. The overall effect is unlikely to be significant as any accidental errors are as likely to increase as decrease any effect.

Other biases that may exist include geographic biases. There may be different disease aetiology and prevalence and different diagnostic thresholds in different countries, or even different regions within the same country.

Incorporation bias will be present in the findings of the analysis for several of the variables. As there is no single diagnostic test for heart failure the reference standard is usually created by a combination of various findings. For novel diagnostic tests such as BNP or NT-proBNP where this test does not form part of standard practice, it is ethical to withhold the result of this test from the treating physician. For standard diagnostic tests, for example, the chest x-ray, which has an established, essential role in forming a diagnosis for the breathless patient, it is not ethical to withhold this information from the treating physician. As a result of this bias, a circular process will be present where the same diagnostic feature is used to make and confirm the diagnosis and then found on subsequent analysis to have a significant diagnostic role. This form of bias is recognised but considered unavoidable in this field (Sackett et al., 2002).

Table 6-1 Different Forms of Bias

Type	When?	Under/over estimate?
Patients		
Spectrum bias	Included patients do not represent intended spectrum of severity of the condition	Depends on difference between targeted and included spectrum.
Selection bias	When eligible patients are not enrolled consecutively or randomly.	Usually leads to overestimation.
Index test		
Information bias	Index test results interpreted with knowledge of the reference standard or with different information than would be known in practice	Usually leads to overestimation
Reference Standard		
Misclassification	Reference standard does not correctly classify patients with the target condition	Depends on whether both tests make the same mistakes
Partial verification bias	Set of patients are identified with a second or third reference standard, especially if this depends on the index test	Usually leads to overestimation
Incorporation bias	Index test is incorporated in reference standard	Overestimation

Disease progression	Patients condition changes between applying index test and the reference standard	Over or underestimation depending on change in patient's condition
Information bias	When the reference standard is interpreted knowing the index test results	Usually overestimation
Data analysis		
Excluded data	Uninterpretable or intermediate test results and withdrawals not included	Usually overestimation

### 6.1.4 Inclusion and Exclusion Criteria

The two readers (CJF & KMJ) used the following criteria to guide the selection of papers to consider for inclusion.

Table 6-2 Inclusion Criteria

Prospective, diagnostic study
Set in an emergency department or acute care environment
Breathlessness the main presenting complaint.
Adult patients.
Reasonable reference standard.
Heart failure the target condition.
Prospective studies

Table 6-3 Exclusion Criteria

Exclusion Criteria
Inappropriate inclusion criteria e.g. >75 years only
Inappropriate reference standard.
Inappropriate exclusion criteria.
Not possible to calculate sensitivity / specificity from data provided.
Poor quality.

### 6.1.5 The Reference Standard

Choosing a reference standard for the diagnosis of heart failure is difficult. There is no universally agreed definition of heart failure and no accepted gold standard tests. Some guidance exists regarding reviews of

diagnostic tests in this situation (Deeks, 2001; Irwig et al., 2002; Gatsonis & Paliwal, 2006). A combination of different diagnostic tests is a legitimate end-point in the absence of a gold standard. Clinical opinion appeared the reasonable choice to make as a reference standard. It is chosen as the reference standard for most of the included studies and it is also a practical method that reflects real life. Physicians have to make diagnostic and therapeutic decisions about patients who may have heart failure based on the available information in daily practice.

A decision was made that the optimal reference standard should be the opinion of two or more physicians, after reviewing of all the diagnostic evidence available including response to treatment and post-mortem results if applicable. Studies with a less rigorous reference standard such as the review of one physician or the review of echocardiogram results have also been included. Echocardiogram results have a high correlation with the clinical diagnosis of heart failure although this is not perfect and the interpretation of this investigation still involves some degree of subjectivity.

Where the reference standard is sub-optimal, the studies have been included in the initial assessment of the data for the particular diagnostic test. For diagnostic tests that appear to have demonstrated useful diagnostic utility, the contribution of the poorer quality studies to the overall results has been examined. A recommended technique is sensitivity analysis, whereby instead of simply excluding poorer quality studies, they are initially included and then their contribution to the analysis result is assessed (Hui & Zhou, 1998).

The threshold of the positivity for the reference standard will vary between studies in addition to the threshold that is explicitly applied to the diagnostic test. This will also affect the diagnostic accuracy of the test and needs to be examined as a source of heterogeneity during the analysis of the data.

Hui and Zhou (1998) make the point that most studies treat the gold standard as though there was conditional independence while this is often not the case. In a condition where there is a spectrum of disease, severe cases are unlikely to be missed by any test, while in the least severe cases the patients may test negative on more than one test. The diagnostic value of a test may also vary with the pre-test probability of a condition in some situations.

### **6.1.6 Methods**

An assessment of each included study and all of the diagnostic data provided by it was entered into the RevMan 5.1 software for diagnostic systematic reviews downloaded from the Cochrane Collaboration website (Cochrane, 2012). This software allows the storage and basic analysis of the available data and also provides a framework for the summary and quality assessment of each study.

The data from each included study was entered by the lead researcher and verified independently by another researcher (RB, DH, SC, KMJ, BF, JB or TB). Data consisted of any information pertaining to the diagnosis of heart failure where it was possible to create a two-by-two table and calculate sensitivity and specificity for the variable. Most papers provided data about more than one diagnostic variable although often the primary intention of the paper was to compare one or more diagnostic tests against the reference standard. On the occasions when a novel diagnostic test was being investigated the index test was often blinded; for established variables the results were usually known and often incorporated as part of the basis of the reference standard.

### **6.1.7 Extraction of Information**

Data was taken directly from the study papers and entered into the RevMan Software. This allows the diagnostic data to be entered either as raw patient numbers or as sensitivity, specificity, prevalence and total

number of patients. This software was then used to compile forest plots for each diagnostic variable from all the studies that provided information about it. Where explicit positivity thresholds were present, the same threshold was used for each of the combined test results. For variables where the threshold was less objective the issue has been addressed in the accompanying text. To reduce the risk of extractor bias or error a second reader was asked to read a selection of the chosen articles and perform the data extraction and quality assessment. A total of seven readers were involved in addition to the main researcher, each reading a selection papers to independently assess the quality and extract the necessary data (RB, DH, SC, KMJ, BF, JB and TB). The results available from the independent readers were compared with the lead researcher results and any differences settled by discussion.

### **6.1.8 Assessment of Quality**

The quality of each study has been assessed using the QUADAS (Quality Assessment of Diagnostic Accuracy Studies) system (Whiting et al., 2003). This system has been designed specifically to assess quality in systematic reviews of diagnostic studies and has been validated by the authors (Whiting et al., 2006). A summary of this system is presented in Appendix III. However, any scoring system requires some level of subjectivity so there remains a possibility of a quality score bias. In an effort to counter this effect, two researchers assessed the quality of each of the included studies independently and summarised the quality assessment on a table derived from the QUADAS criteria.

Scoring systems such as QUADAS are useful as a method of assessing and displaying the quality of the included studies but the use of these systems still requires some degree of subjectivity. While individual aspects of quality can be demonstrated to affect diagnostic accuracy, the overall quality scores produced by scoring systems such as QUADAS, or even the stratification of studies into high or low quality, has not been

shown to be of value (Hui & Zhou, 1998). A recommended strategy is to examine the role of individual aspects of quality and how this affects the analysis of the diagnostic utility of the variable.

### **6.1.9 Data Presentation**

The data for each variable has been summarised using a forest plot and a summary receiver operator characteristic (SROC) curve. Coupled forest plots allow the simultaneous display of sensitivity and specificity from each study along with the 95% confidence intervals. The study identifier and numbers of true positive, false positive, false negative and true negative values are provided for each included study. This method allows concise display of the diagnostic information provided by each study for the relevant variable. The graphic display of the forest plot allows the presence of heterogeneity to be observed directly although it can be more difficult to discern a threshold effect.

A summary ROC curve has been presented for each of the diagnostic variables to display the results of the included studies. This allows visualisation of any clustering of results, threshold effects or outlying studies in the analysis. Each analysis was examined to see if a threshold effect was apparent by visual examination of the calculated ROC curve. The information provided by the summary ROC curve is discussed for each of the included variables.

### **6.1.10 Selection of Variables for Meta-analysis**

Sensitivity is a measure of the proportion of patients who truly have the condition who are found to be positive on application of the diagnostic test of interest. A test of that has high sensitivity can have a useful role in ruling out the condition for patients who have a negative test result.

Specificity is a measure of the proportion of patients who do not have the condition and who test negative for the condition. This measure is very useful in that tests with a high specificity can be used to rule in the condition in patients for whom the diagnostic test is positive.

The intention of this research is to create a decision rule to rule-in and rule-out heart failure therefore it needs to involve variables with reasonable sensitivity and specificity. Tests such as JVP are very specific and can rule-in the condition but have low sensitivity. Other variables such as a history of hypertension are sensitive but not specific.

Youden's Index is a method of examining a variables diagnostic merit (Youden, 1950). Adding the sensitivity result to the specificity result, and subtracting one from this figure obtains Youden's Index value. The perfect test would score one; tests that score less than zero have little diagnostic value. For example, a diagnostic test with a sensitivity of 0.8 and a specificity of 0.6 would have a Youden's Index score of 0.4 ( $0.8 + 0.6 - 1.0$ ). Although Youden's Index provides a rather crude measure of diagnostic accuracy it was decided that variables that failed to score greater than 0.4 in at least one included study should be excluded from further analysis. The intention was to concentrate on variables that had a reasonable sensitivity and specificity. A score of 0.4 corresponds with the equivalence of a sensitivity and specificity of 70%.

Variables that scored less than 0.4 on Youden's Index but had very high specificity for heart failure were also investigated further. These tests were considered to be potentially useful, as a positive result would mean that a patient had a high likelihood of having heart failure. The disadvantage is that these findings are present in only a few patients so are likely to have a limited role in a diagnostic pathway. Information regarding the positive likelihood ratio associated with the presence of these clinical signs was also provided.

### **6.1.11 Included Studies**

Forty-four studies were included, derived from forty-eight papers with a total of 15,496 original patients. Table 6-4 provides basic information about the included studies. A summary of the each study is also provided in Appendix IV. Table 6-5 displays the quality assessment of each of the included studies. Figure 6-2 displays the number of patients in each study against the prevalence of heart failure in that study.

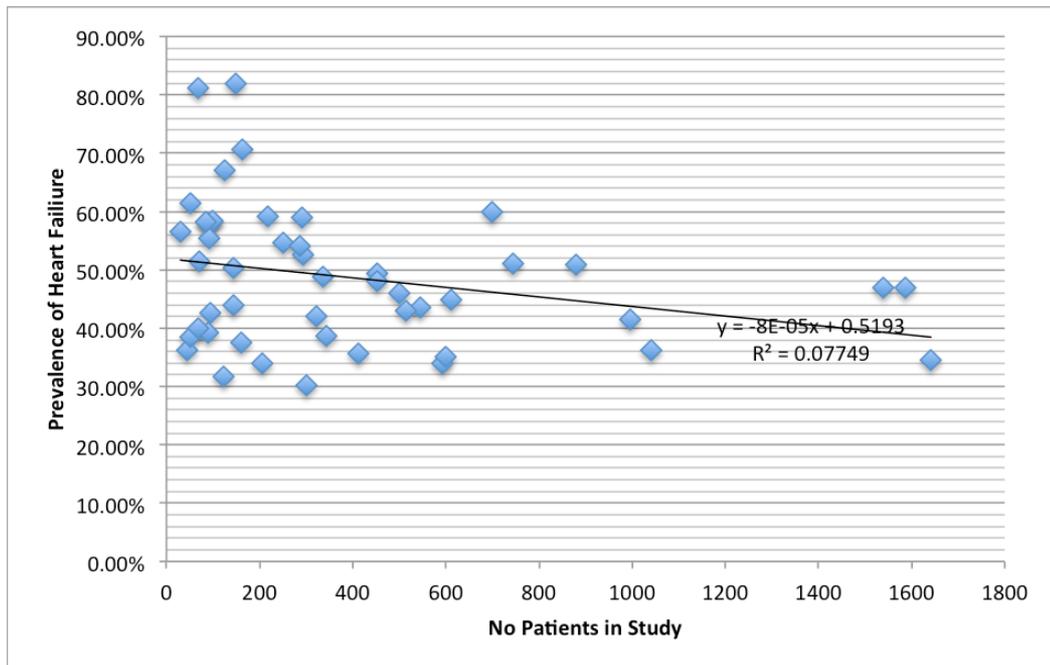


Figure 6-2 Study Size Versus Prevalence For Included Studies

Table 6-4 – Included Papers

Author / Year	Country	Total Patients	Prevalence of HF	Recruitment dates	Notes
Ababsa et al. (2005)	France	86	58.1%	?	Dates unclear
Alibay et al. (2005)	France	160	37.5%	6/2003 - 1/2004	
Barcarse et al. (2004)	US	98	58.2%	6/2001- 4/2002	
Bayes-Genis et al. (2004)	Spain	100 (11excl.)	58.4%	4/2002 - 2/2003	
Boldanova et al. (2010)	Austria	452	49.3%	5/2001-4/2002.	Same patient group as Mueller C (2005)

Choi et al. (2007)	Korea	1040	36.3%	10/2004 -10/2005	
Chung and Hermann (2006)	Australia	143	50.3%	?	Dates unclear
Collins, Lindsell, Peacock, Eckert, et al. (2006)	US	343 (96 excl.)	38.7%	9/2003 - 6/2004	
Collins et al. (2009)	International	995	41.5%	3/2006 - 10/2006	
Coste et al. (2006)	France	699	60.0%	9/2003 - 2/2005	
Davis et al. (1994)	New Zealand	52	61.5%	?	Dates unclear

Fleischer et al. (1997)	New Zealand	123	31.7%	11/1994 - 1/1995	
Garcia (2013)	US	44	36.3%	4/2010 – 8/2010	
Gargani et al. (2008)	Italy	149	81.9%	11/2004 – 3/2006	
Gegenhuber et al. (2006)	Austria	251	54.6%		Patients from Mueller (2005)
Green et al. (2008)	US	592	34.0%		Patients from Januzzi (2005)
Huang et al. (2006)	Taiwan	92(12 excl.)	55.4%	1/2005-10/2005	
Januzzi et al. (2005)	US	599	35.0%	4 months?	Duration unclear

Jourdain et al. (2002)	France	125	67.0%	7/2001 -10/2001	
Kang et al. (2006)	Korea	69	81.2%	11/2003-5/2004	
Klemen et al. (2009)	Slovenia	546	43.6%	1/2005 – 6/2007	
Knudsen, Omland, Clopton, et al. (2004)	International	880(706 excl.)	50.8%		As per Maisel (2002)
Lainchbury et al. (2003)	New Zealand	205	34.0%	?	Dates unclear
Lakhdhar et al. (2013)	Tunisia	30	56.6%	1/3/2010 – 20/6/2010	

Liteplo et al. (2009)	US	94(6 excl.)	42.6%	12/2006-6/2007	
Lo et al. (2007)	Taiwan	52 (8 excl.)	38.5%	?	Dates unclear
Logeart et al. (2002)	France	163 (72 excl.)	70.6%	6/1999-6/2001	
Lokuge et al. (2010)	Australia	612 (87 excl.)	44.8%	8/2005 – 4/2007	
Maisel et al. (2002)	International	1586	47.0%	4/1999-12/2000	
Maisel et al. (2010)	International	1641	34.6%	3/2007-2/2008	
McCullough et al. (2002)	International	1538 (48 excl.)	47.0%		As per Maisel (2002)
Miller et al. (2012)	US	89	39.3%	1/9/2008 – 30/4/2009	3 patients excluded due to poor views.

Moe et al. (2007)	Canada (multi-centre)	500 (34 excl.)	46.0%	12/2004-12/2005	
Morrison et al. (2002)	US	321	42.0%	6/1999-6/2000	
Mueller, Frana, et al. (2005)	Austria	293	52.5%	10/2003-2/2004	
Mueller et al. (2004)	Switzerland	452	48.0%	5/2001 – 4/2002	
Nazerian et al. (2010)	Italy	145	44%	1/2007-3/2007	
Parrinello et al. (2008)	Italy	292	58.9%	12/2004 - 12/2006	
Potocki et al. (2010)	Switzerland	287	54%	4/2006 – 3/2007	

Prosen et al. (2011)	Slovenia	218	59.2%	7/2007 – 4/2010	
Ray and Lefort (2006)	France	514	43.0%	2/2001- 9/2002	
Reichlin et al. (2010)	Switzerland	743	51%		Patients taken from Mueller, C (2005) and Potocki (2010) studies.
Robaei et al. (2011)	Australia	68	40%	?	Dates unclear
Seronde et al. (2013)	France	336	48.8%	June 2009?	Dates not given in studies but on clinicaltrials.gov. Acute dyspnoeic subpopulation data provided separately in paper.
Shah et al. (2009)	US (Multi-centre)	412	35.7%	5/2003 - 12/2006	

Singer et al. (2009)	US (Multi-centre)	301	30.2%	2005	
Villacorta et al. (2002)	Brazil	70	51.4%	4/2001-7/2001	
Wang et al. (2010)	Taiwan	84	58.3%	6/2005 – 9/2006	

## 6.1.12 Assessment of Quality

The following table provides the summary assessment of quality for all of the included papers. The summary of the assessment is based on the QUADAS scoring system for diagnostic studies. Blank rows occur where the authors have reported on another aspect of a study that has already been assessed under the primary authors.

Table 6-5 Quality Assessment of Included Papers

Author / Year	Representative	Acceptable Reference Standard?	Acceptable delay between tests?	Partial verification	Differential Verification avoided?	Incorporation	Reference standard blinded?	Index test results	Relevant clinical	Uninterpretable results reported?	Withdrawals
Ababsa (2005)	?	?	+	+	+			+	+	+	
Alibay et al. (2005)	?	?	+	+	+	+	-	-	+	?	?
Barcarse et al. (2004)	-	+	+	+	+	+	+	+	+	+	+
Bayes-Genis et al. (2004)	+	+	+	+	+	+	-	+	+	-	+
Choi (2007)	+	+	+	+	+	+	+	+	+	?	
Chung (2006)	+	?	?	+	?	-	+	-	+	?	?
Collins (2006)	+	+	+	+	+	+	+	+	+	+	+
Collins (2009)	+	+	+	+	+	+	+	+	-	+	+
Coste (2006)	+	+	+	+	+	+	+	+	+	+	+
Davis (1994)	?	+	+	+	+	?	-	-	+		+
Fleischer (1997)	+	-	+	+	+	+	+	-	+	+	+
Garcia (2013)	?		+	+		+	+		+	+	+
Gargani (2008)	?	+	+	+	+	+		+	+	+	+

Gegenhuber (2006)	-	+	+	+	+	+	+	+	+	+	-
Green (2008)	+	+	+	+		-		+	+	+	+
Huang (2006)	+	+	?	+	+	+	+	+	+	-	+
Januzzi (2005)	+	+	+	+	+	+	+	+	+	+	+
Jourdain (2002)	?	+	+	+	+	+	+		+	?	
Kang (2006)	+	?	+	+	+	+	+			+	+
Klemen (2009)	+	?	+	?	+	+	?	+	+		?
Knudsen (2004)	+	+	+	+	+	+					+
Lainchbury (2003)	+	+	+	+	+	+	-	+			+
Liteplo (2009)	+	+	+	+	+	+	+	?	+	+	+
Lo (2007)	?	+	?	+	+	+	+	+	?	+	?
Logeart (2002)	+	+	+	+	+	+	-	+	+		+
Maisel (2002)	+	+	+	+	+	+	+		+	+	+
Maisel (2010)	+	+	+	+	+	+	+	+	+	+	+
McCullough (2002)											
Miller (2011)	+	+	+	+	+	+	+	-	+	+	+
Moe (2007)	+	+	+	+		-	-	+	+	+	+
Morrison (2002)	+	+	+	+	+	+	+		+	+	+
Mueller, Gegenhuber, et al. (2005)	-	+	+	+	+	+	+	-	+	+	+
Mueller C (2005)	+	+	+	+	+	-	+	-	+	+	+
Nazerian (2010)	+		+	+		+	+	+	+	+	

Parrinello (2008)	+	+	+	+	+	+	-	+	+	+	+
Potocki (2010)	+	+	+	+	+	+	+		+	+	
Prosen (2011)	+	?	+	+	+		+	-	+		+
Ray (2006)	+	?	+	+			+	+	+	+	+
Reichlin (2009)											
Robaei (2011)	+	+	+	+		-	-	+	+	+	+
Seronde (2013)	+	+	+	-	-	-			+		
Shah (2009)	+	+	+	+	+	+	+	+	+	+	+
Singer (2009)	?	?	+	?		?	+	?	?		+
Villacorta (2002)	+	+	+	+	+	+	+	+	+	+	+
Wang (2010)	+	+	+	+	+	+	+	+	+	?	+

### 6.1.13 Excluded Studies

Table 6-6 describes the characteristics of excluded studies and details the reasons for their exclusion.

Table 6-6 Excluded Papers

Study	Reason for Exclusion
Behnes et al. (2009)	Recruited patients with acute dyspnoea and/or peripheral oedema. Unable to separate results to look at diagnostic function in dyspnoeic patients only.
Berdagué et al. (2006)	Only selected patients over the age of 70y and excluded patients in whom they could not decide on a clear diagnosis.

Chenevier-Gobeaux et al. (2005)	Reference standard of two 'urgentists' in the Emergency Department not acceptable.
Cinar et al. (2012)	Only recruited dyspnoeic patients where the Emergency Physician had diagnostic uncertainty and informed them of the result of the novel test within 30 – 60 minutes to assess if that altered their diagnosis.
Collins, Lindsell, Peacock, Eckert, et al. (2006)	Patients already included in Collins, Lindsell, Peacock, Hedger, et al. (2006).
Dao et al. (2001)	Subset of patients from Morrison et al. (2002) paper.
deFilippi et al. (2007)	Breaks patient data into those with renal impairment and those without and then provides cut-off values of BNP and NT-proBNP for these groups but doesn't provide overall values for all patients and not enough data provided to calculate this.
Dieplinger et al. (2009)	Data for tests with independent diagnostic utility for heart failure already presented in Maisel et al. (2005) and Gegenhuber et al. (2006) papers.
el Mahmoud et al. (2006)	Only looked at patient over 75y age so not representative sample.
Eckstein et al. (2012)	Includes patients from study by Potocki (2010) & Reichlin (2010). Divides patients into those with AF and without and looks at diagnostic performance of NT-proBNP and MR-ANP in these groups.

Eurlings et al. (2012)	Prognostic but not diagnostic study.
Gillespie et al. (1997)	Patients had already been admitted to hospital.
Gruson et al. (2008)	Patients presenting with chest symptoms, so included chest pain etc. not just dyspnoea.
Havelka et al. (2011)	Reference standard was discharge diagnosis as entered by a primary care physician.
Jang et al. (2010)	Reference standard was a radiologist interpretation of chest x-ray.
Jang et al. (2012a)	Same study as Jang et al. (2010)
Jang et al. (2012b)	Same study as Jang et al. (2010)
Januzzi et al. (2006)	Patients already included in individual studies.
Javadzadeh et al. (2008)	Included patients with dyspnoea and others presenting with chest pain.
Jose et al. (2003)	Used combination of patients presenting acutely and those with chronic shortness of breath either presenting to the emergency department or referred to outpatients.
Leuppi et al. (2005)	Looked at patients presenting with chest symptoms, not just dyspnoea.
Malas et al. (2003)	Excluded patients with both cardiac and respiratory disease from the study retrospectively.

Marantz et al. (1990)	Poor reference standard. Applied modified scoring system in ED to patients as reference standard. ECG and CXR only performed at the discretion of the attending physician so full reference standard not consistently applied to all patients.
McCord et al. (2005)	Subgroup analysis of patient group reported in Knudsen, Omland, Clapton, et al. (2004)
McNamara and Cionni (1992)	Excluded patients from the study retrospectively so that only patients with either heart failure or COPD were included.
Mendez Bailon et al. (2007)	Recruitment methods unclear. NT-proBNP measured. No mention of blinding. Reference standard is application of Framingham criteria by GPs plus echo data if available, not clear at what point this was carried out. Little in the way of inclusion or exclusion criteria mentioned, not clear if patient group part of another study.
Mueller and Laule-Kilian (2004)	Randomised patients to have BNP test or not and then looked to see if this altered outcomes. Not diagnostic test. Data from clinical and investigative findings from same study included as Mueller, Frana, et al. (2005)paper.
Mueller-Lenke et al. (2006)	Sub-group analysis of (Mueller, Frana, et al., 2005) study in patients who had chest x-ray in emergency department which was reported by a radiologist.

Pascual Figal et al. (2005)	Selected 70 out of 1267 patients who presented with acute dyspnoea whom the investigators felt may have heart failure but may have had another diagnosis.
Peacock et al. (2006)	No reference standard applied. Asked ED physicians if they thought the patient had heart failure then provided information from a bio-impedance reading and recorded how many clinicians altered their diagnosis.
Ray et al. (2004)	Data included in Ray et al. (2006) paper. The diagnostic result of BNP is taken from this paper and included in the analysis but referenced as the 2006 paper as this included most of the information about this study.
Ray et al. (2005)	Patients already included in the Ray and Lefort (2006) paper. The diagnostic result of NT-proBNP is taken from this paper and included in the analysis but referenced as the 2006 paper as this included most of the information about this study.
Rogers et al. (2009)	Retrospective analysis of patients who had a BNP checked, not patients presenting with heart failure.
Sanz et al. (2006)	No reference standard provided. Not clear how or when diagnosis of heart failure made. Seems to assume presence of AF confirms cardiac cause of dyspnoea. No patient demographics. No time scale provided for when tests were taken etc.

Schneider et al. (2009)	Not diagnostic study. Patients randomised to have BNP tested or not tested and looked at outcomes in these two groups.
Shah et al. (2012)	Data derived from PRIDE study by Januzzi et al. (2005).
Strunk et al. (2006)	Sub-group analysis of Maisel et al. (2002) study.
Studler et al. (2008)	Sub-group analysis of Mueller and Laule-Kilian (2004) study.
Styron et al. (2009)	Not diagnostic study, retrospective cohort. Looked at patients who had had an echo performed and then compared their LV ejection fraction with their systolic BP to look to see if there was any correlation.
van der Burg-de Graauw et al. (2009)	Not clear who made the diagnosis of heart failure or at what stage in the patient journey. A definitive diagnosis was made using all available data including an NT-proBNP level and then looked to see if the NT-proBNP result made a difference. Cannot calculate final diagnosis from available results. Other factors such as COPD, hypertension, ECG, age, glucose and clinical judgement were of value on stepwise logistic regression but only odds ratios provided.

van Kimmenade et al. (2006)	Further analysis of PRIDE study by Januzzi et al. (2005). Looking at other biomarkers but nothing added diagnostically, suggest may have a role in predicting adverse outcomes.
--------------------------------	---

## Systematic Review

The literature search is detailed in the preceding sections. Duplicates and papers that were not diagnostic studies or had no relevance to the diagnosis of acute heart failure on the basis of their title were excluded leaving 500 papers that were reviewed at greater depth. Ninety papers were selected from these on the basis of the inclusion and exclusion criteria detailed in the following section. From this selection forty-nine papers were included and forty-one papers were excluded. The included studies are detailed in Section 6.1.12 and summarised in Appendix IV; the excluded papers are detailed, with the reasons for exclusion in Section 6.1.14.

Forty-four articles were selected that provided information about the diagnosis of heart failure in the breathless patients in an acute setting. The articles were derived from thirty-eight studies. Data regarding forty diagnostic variables was present in the selected articles. Some of the papers presented different information from the same study. If relevant, the articles have been listed separately although it is made clear in the text that the patients were from the same study. Articles dealing with studies that did not provide original diagnostic information of diagnostic value were excluded. One author, Ray, provided relevant information pertaining to the same study over three articles (Ray et al., 2004; Ray et al., 2005; Ray & Lefort, 2006). For the sake of simplicity data derived from these three articles has been referenced to as Ray et al. (2006).

BNP was measured in a number of studies and information was provided about twenty-nine different cut-off levels. Twenty-nine different cut-off values were also available for NT-proBNP. Several studies provided several cut-off values with a sensitivity and specificity for each, or data such that these values could be calculated. In the absence of any statistical method of combining the data for the different cut-off values from all the different studies, this data was discarded. Cut-offs corresponding to the manufacturers recommendations were selected for further analysis and are detailed at the relevant sections in the results section.

A coupled forest plot has been produced for each variable, and where appropriate, a summary Receiver Operating Characteristic (ROC) curve has been calculated. Further analysis has been performed for variables that demonstrate significant diagnostic value including meta-analysis where appropriate. Each included study is represented by a square plotted in the ROC space, the size of the square is proportionate to the size of the study and hence the weight it carries in the analysis. Where summary values have been plotted they are shown as a solid black circle.

## 6.2 Variables with Little Diagnostic Value

### 6.2.1 Absence of History of Asthma

As patients with asthma tend to present with acute dyspnoea, a previous diagnosis with this condition could suggest that this is more likely to be the cause of the patients' breathlessness than a coincidental occurrence of acute heart failure. Three studies provided information about the history of asthma in a total of 1,596 patients presenting to emergency departments with acute dyspnoea. They demonstrate the absence of a history of asthma is sensitive for the correct diagnosis of heart failure but has very poor specificity.

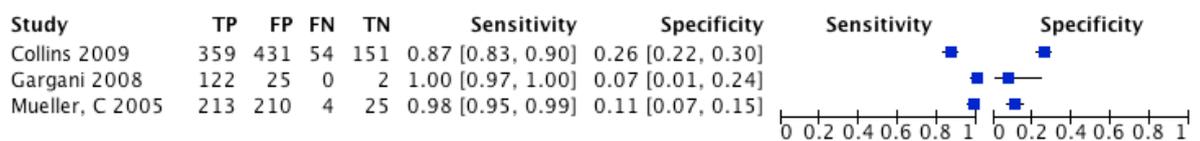


Figure 6-3 Forest Plot of History of Asthma

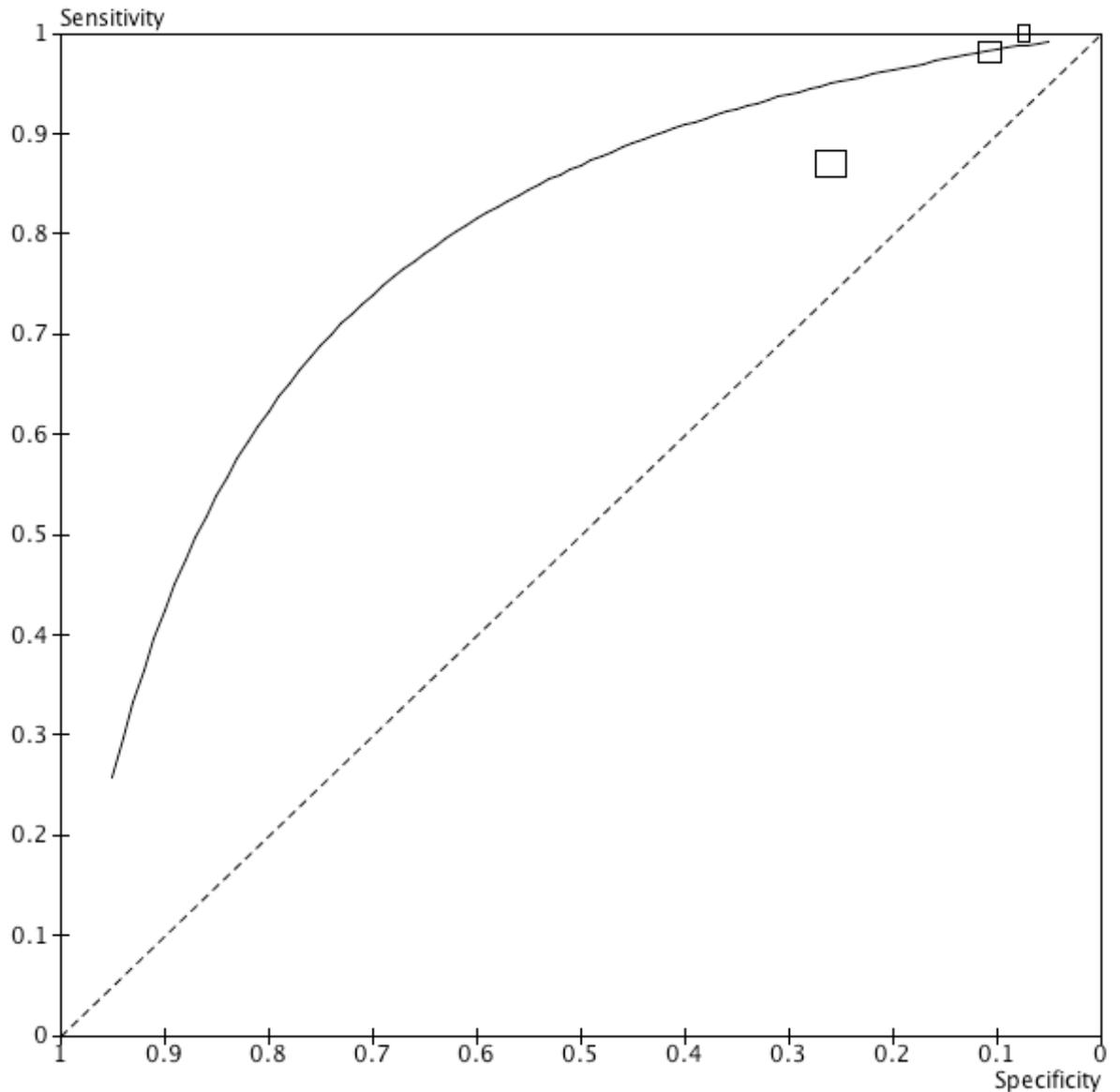


Figure 6-4 Summary ROC Curve for Absence of History of Asthma

### 6.2.2 Presence of an Abnormal Electrocardiogram

There are no specific changes on an electrocardiogram (ECG) that are diagnostic of heart failure but cardiac disease is associated with both ECG changes and heart failure and it has been suggested that heart failure is an unlikely diagnosis in the presence of a completely normal ECG (Mant et al., 2010). Two studies looked at this factor, determining whether the ECG appeared within normal limits or was abnormal (without defining what abnormalities were present) in 1,321 patients. Although an abnormal ECG

was suggestive of heart failure it did not obtain a useful level of diagnostic utility as demonstrated in the following figures.

Study	TP	FP	FN	TN	Sensitivity	Specificity	Sensitivity	Specificity
Klemen 2009	137	39	101	164	0.58 [0.51, 0.64]	0.81 [0.75, 0.86]		
Knudsen 2004	259	95	188	338	0.58 [0.53, 0.63]	0.78 [0.74, 0.82]		

Figure 6-5 Forest Plot for Abnormal Electrocardiogram

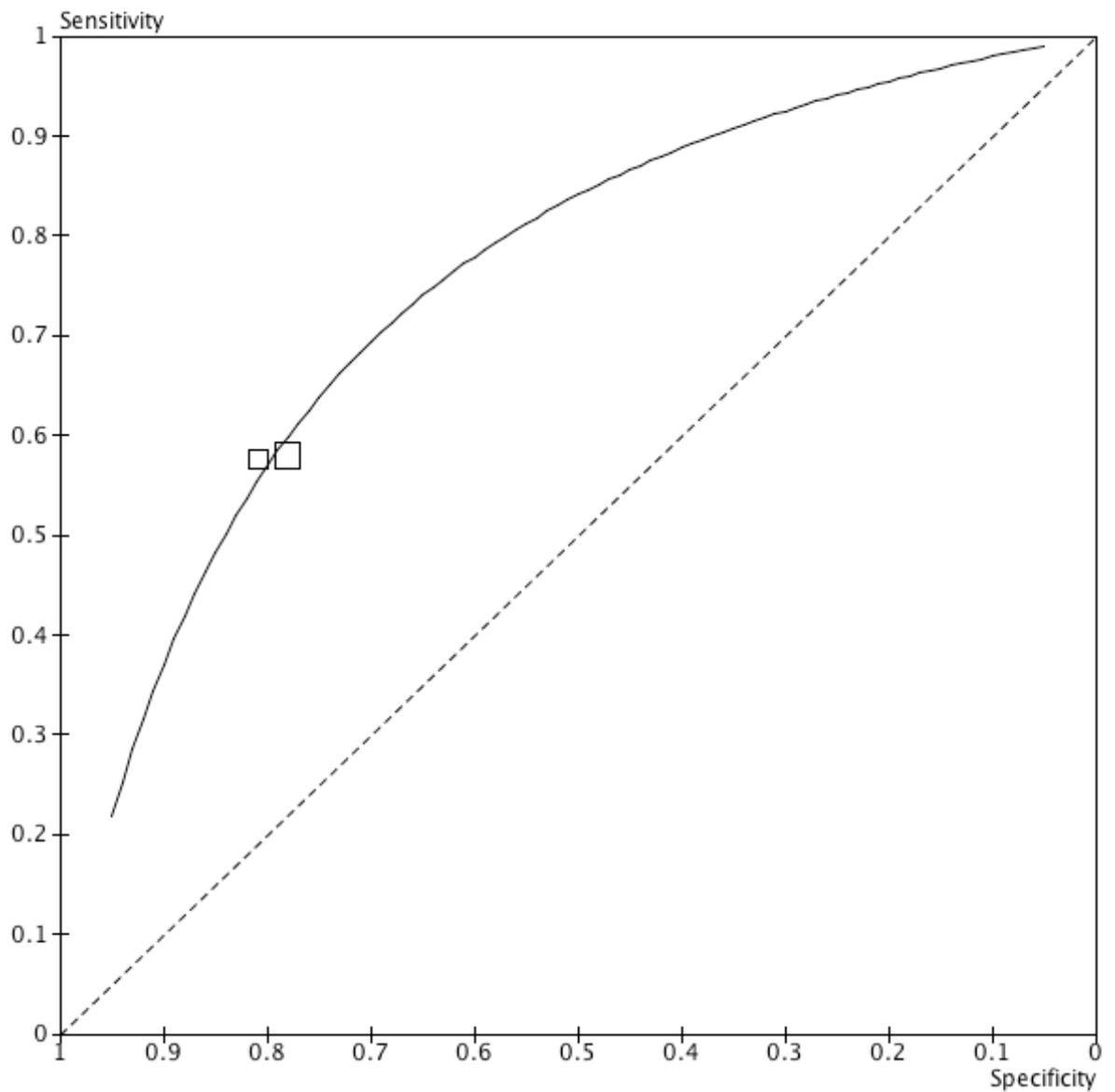


Figure 6-6 Summary ROC Curve for Abnormal Electrocardiogram

### 6.2.3 History of Smoking Tobacco

Smoking is a major risk factor for the development of ischaemic heart disease but also for COPD. Four studies examined this risk factor in a total of 2249 patients. The four studies providing data about this risk factor have discordant results and do not suggest any diagnostic role for this variable.

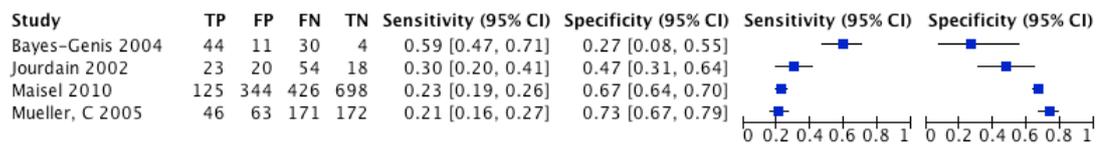


Figure 6-7 - History of Tobacco Smoking

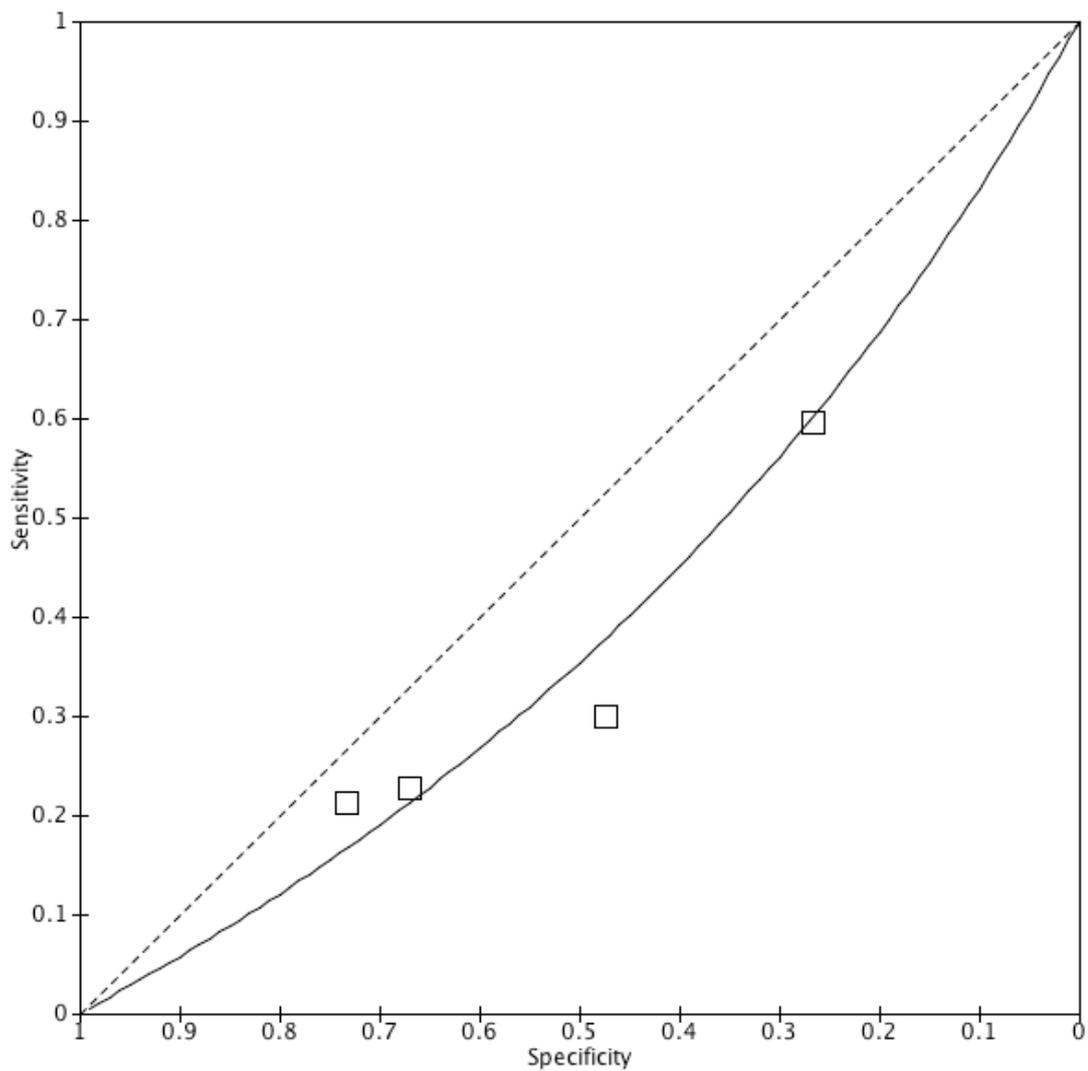


Figure 6-8 Summary ROC Curve - History of Tobacco Smoking

### 6.2.4 Hypertension

Twenty-two studies provided information about the diagnostic value of a history of hypertension in predicting patients with heart failure. A total of 9,360 patients were included. The sensitivity and specificity of this condition as a predictor of heart failure was very variable in the different studies; this may reflect different definitions of hypertension, different levels of screening within the population, different thresholds for commencing treatment or genuine population differences. While there was a tendency for patients with a history of hypertension to have heart failure this did not translate into any useful

diagnostic criteria in any individual study. The overall diagnostic utility for this finding, as presented by the summary ROC curve, was poor. The calculated values, forest plot and summary ROC curve are provided below.

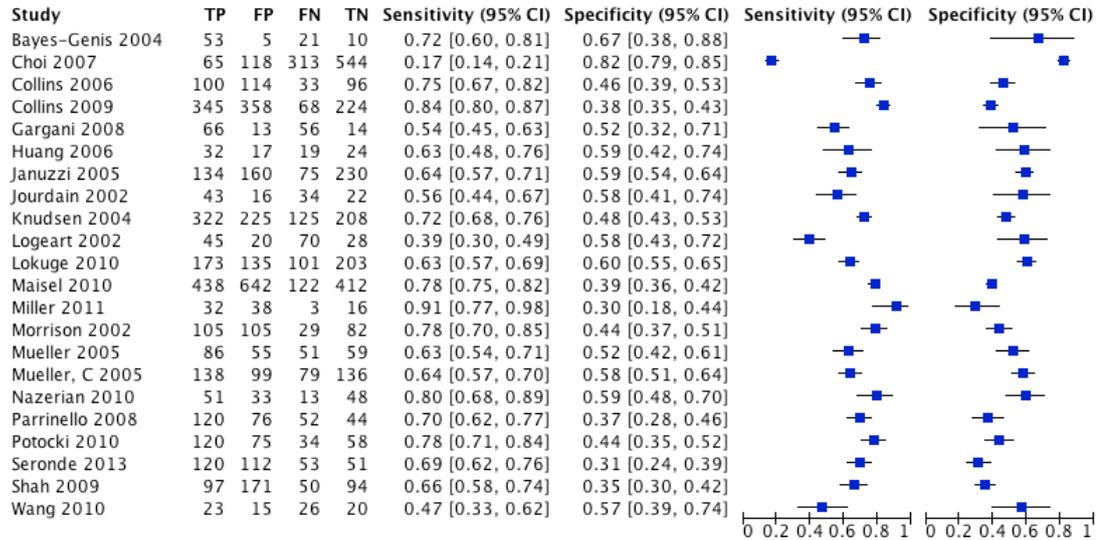


Figure 6-9 Sensitivity and Specificity of Hypertension as a Predictor of Heart Failure

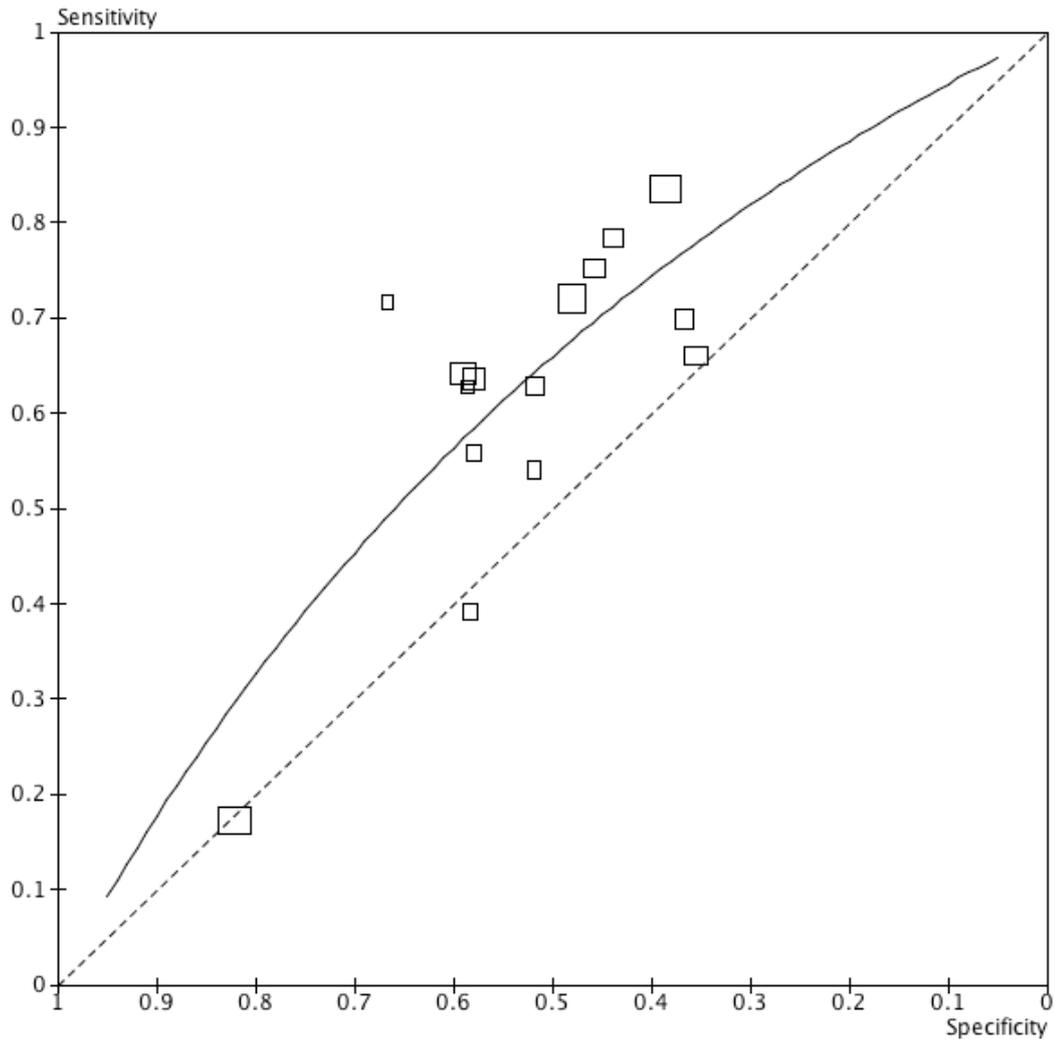


Figure 6-10 Summary ROC Curve - History of Hypertension

### 6.2.5 History of a cough

Patients with heart failure may complain of the presence of a cough although this is usually due to respiratory conditions. Six papers looked at this symptom in a total of 2,514 patients. As can be seen from the forest plot in Figure 6-11 and the summary ROC curve in Figure 6-12, the presence of a cough suggests a non-cardiac cause for the patient's dyspnoea but the sensitivity and specificity for absence of a cough are not high enough to have a useful diagnostic role.

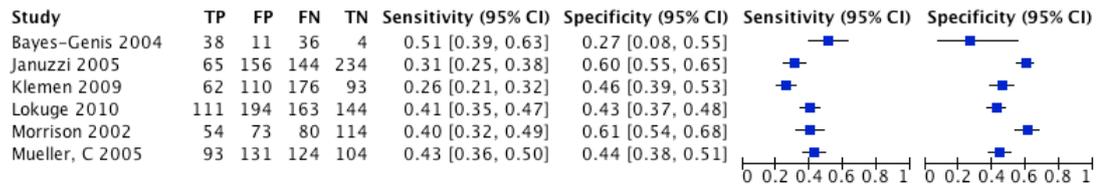


Figure 6-11 Sensitivity and Specificity of History of Cough

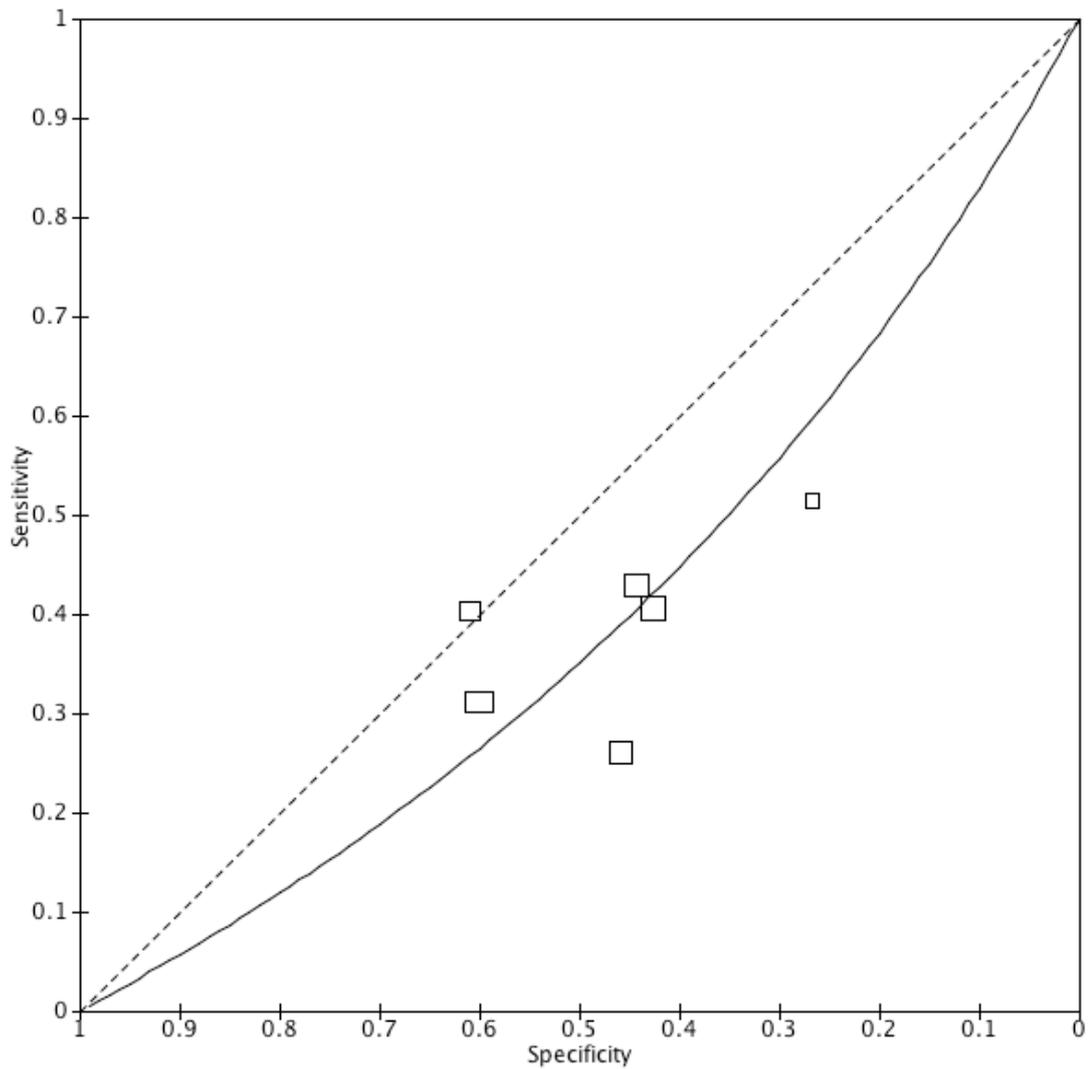


Figure 6-12 Summary ROC Curve – History of Cough

## 6.2.6 History of Orthopnoea

Thirteen studies provided information about the presence of orthopnoea derived from 5,757 patients. This symptom, a sensation of increased breathlessness on lying flat, is thought to occur because of increased venous return in this position (O'Connor, 2005). There is considerable heterogeneity in the results as can be seen from the table and forest plot provided below this text. A summary ROC curve demonstrates the distribution of the data in Figure 6-14.

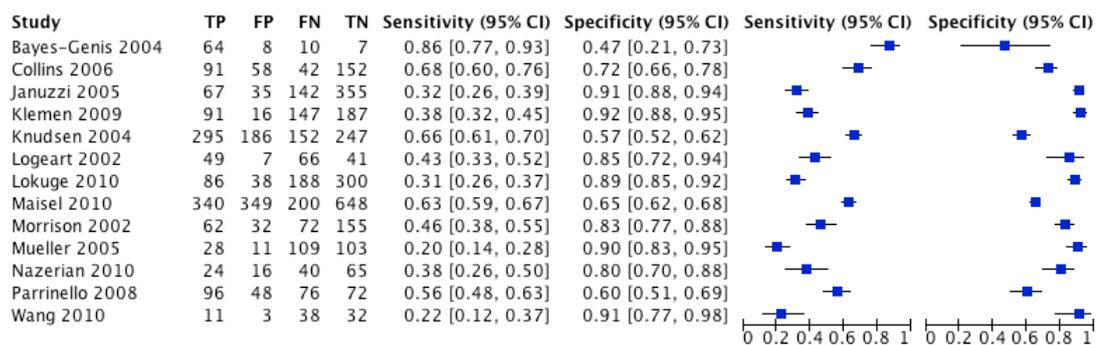


Figure 6-13 Forest Plot – Presence of Orthopnoea

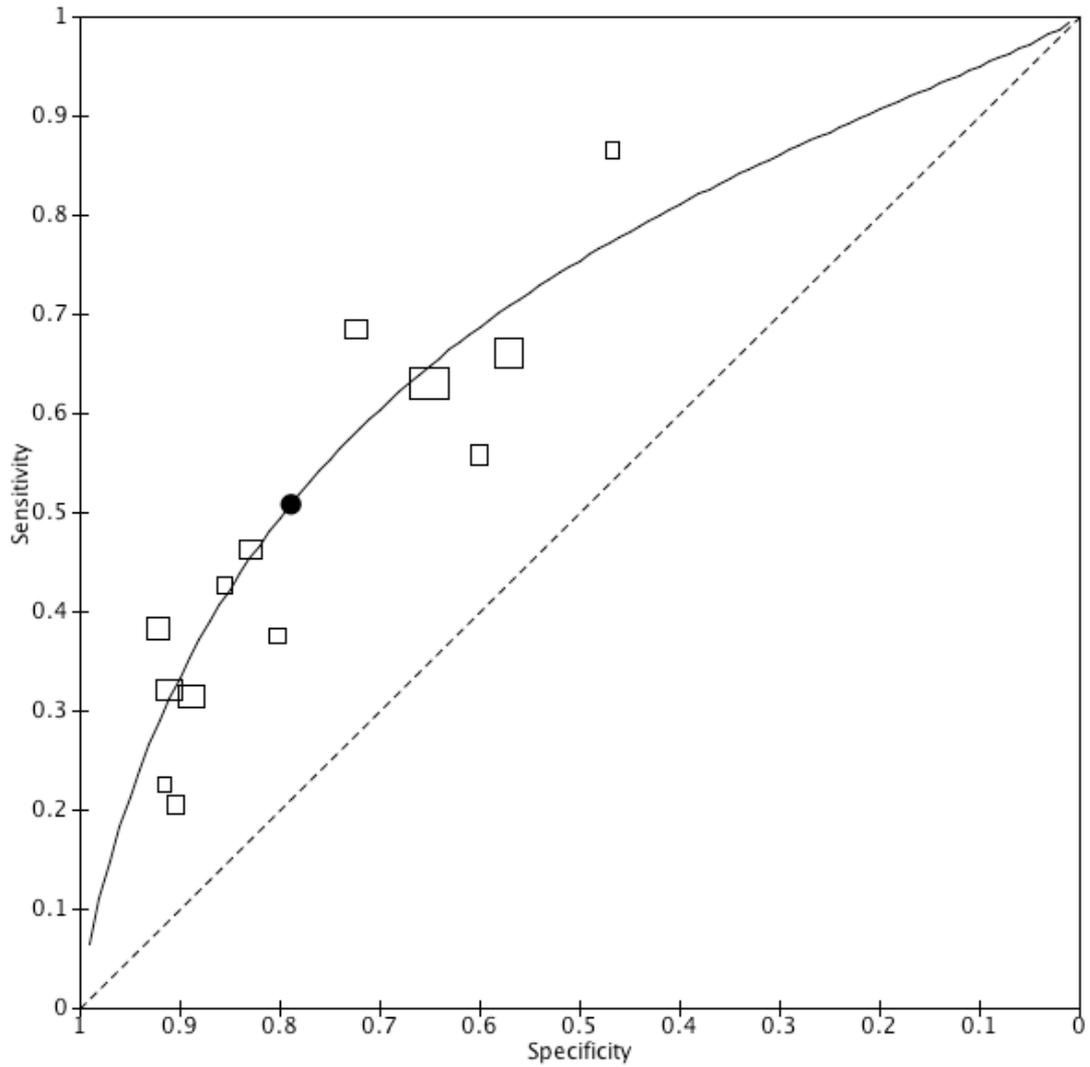


Figure 6-14 Summary ROC Curve – Presence of Orthopnoea

The data was entered into Meta-DiSc that provided the following information about the heterogeneity of the diagnostic odds ratios.

Heterogeneity chi-squared = 27.54 (d.f. = 12) p=0.006

Inconsistency ( $I^2$ ) = 56.4

Estimate of between-study variance (Tau-squared) = 0.0727

Analysis of Diagnostic Threshold

Spearman Correlation Coefficient: 0.883 p value = 0.002

## Meta Analyst analysis

Specificity = 0.801 (95% CI 0.705 – 0.875)

Sensitivity = 0.404 (95% CI 0.365 – 0.590)

Positive Likelihood Ratio = 2.03

Negative LR = 0.74

The pattern of distribution of points suggests a possible threshold effect. The summary point and wide dispersion of the included study outcomes suggest that the test does not have sufficient diagnostic accuracy to consider as part of an optimal diagnostic strategy.

### **6.2.7 Absence of Fever**

Fever is usually associated with an infective process so it may be assumed that the presence of pyrexia would suggest an infective cause of the patient's dyspnoea rather than heart failure. Four studies provided information about this finding in 1723 patients. This finding demonstrated good sensitivity but poor specificity as shown in Figure 6-15 and Figure 6-16.

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Klemen 2009	224	155	14	48	0.94 [0.90, 0.97]	0.24 [0.18, 0.30]	■	■
Lokuge 2010	250	282	24	56	0.91 [0.87, 0.94]	0.17 [0.13, 0.21]	■	■
Mueller, C 2005	185	158	32	77	0.85 [0.80, 0.90]	0.33 [0.27, 0.39]	■	■
Prosen 2011	122	68	7	21	0.95 [0.89, 0.98]	0.24 [0.15, 0.34]	■	■

Figure 6-15 Forest Plot for Absence of Fever

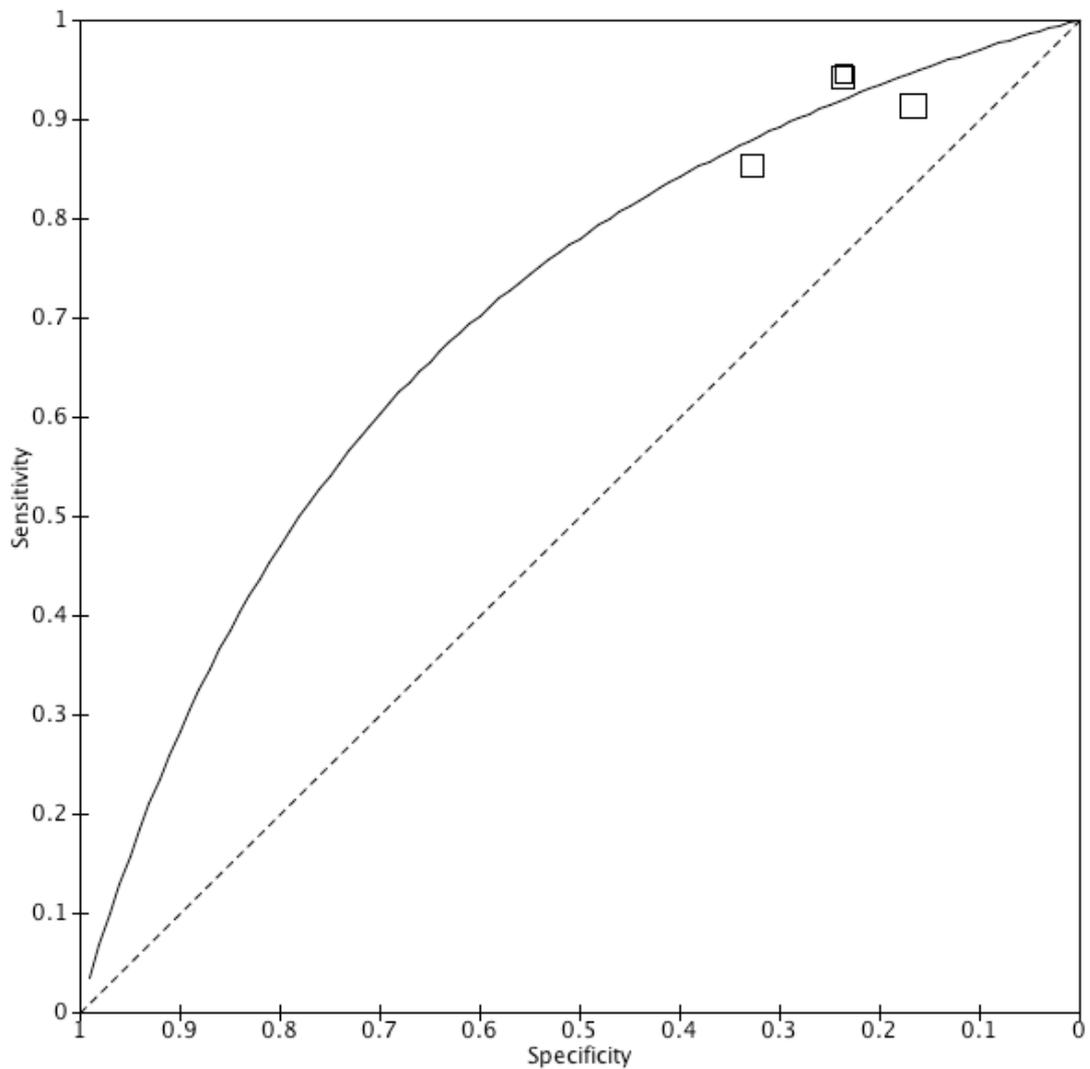


Figure 6-16 Summary ROC Curve for Absence of Fever

## 6.2.8 History of Nocturnal Cough

The presence of a nocturnal cough is thought to be associated with heart failure due to the worsening of respiratory compromise on becoming recumbent (Irwin & Madison, 2000). Three studies looked at this symptom in a total of 686 patients. From the variable and poor sensitivity and specificity

values and the summary ROC curve it can be seen that this aspect of the history has no diagnostic utility.

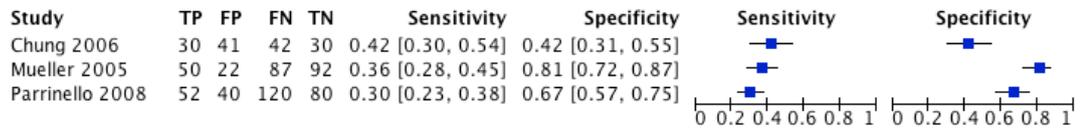


Figure 6-17 Forest Plot of History of Nocturnal Cough

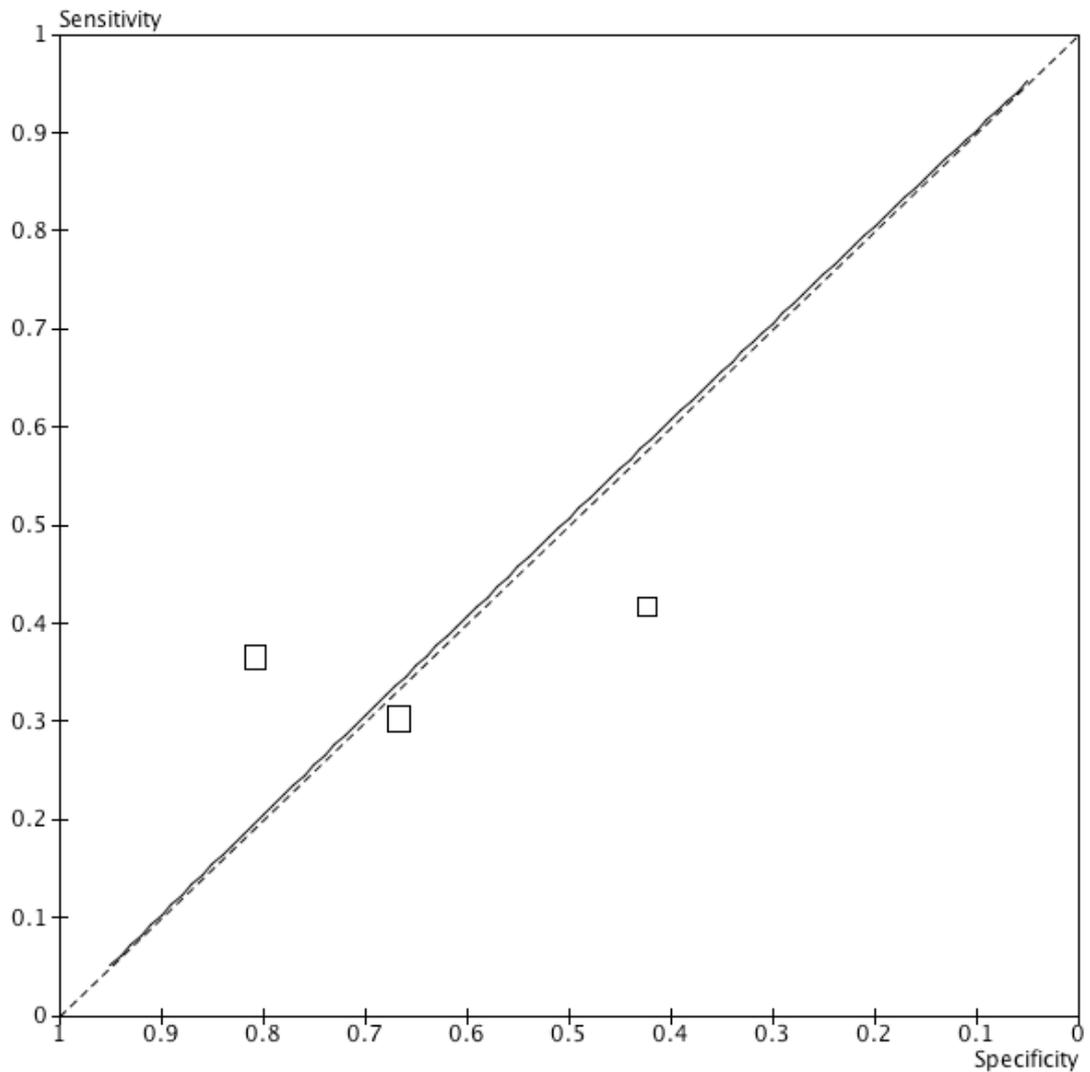


Figure 6-18 Summary ROC Curve for History of Nocturnal Cough

## 6.2.9 The Use of Diuretics

Loop diuretics have long been used as a treatment for heart failure to attempt to reduce fluid overload. Thiazide diuretics are added as a second-line treatment for heart failure and are also used in the treatment of hypertension, a known risk factor for developing heart failure. It is reasonable to assume that patients displaying symptoms consistent with heart failure may be commenced on diuretics, and so to examine the data for a correlation. There were five studies, involving a total of 3,670 patients, which provided this information.

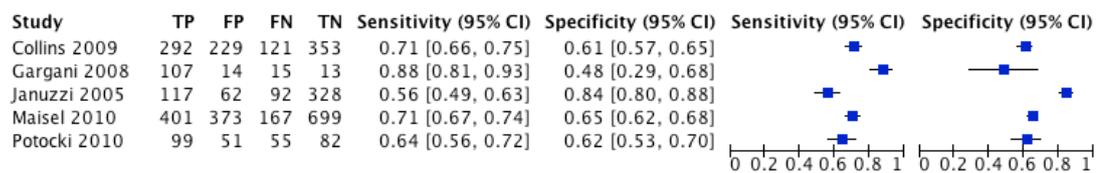


Figure 6-19 Forest Plot - Patients Using Diuretics

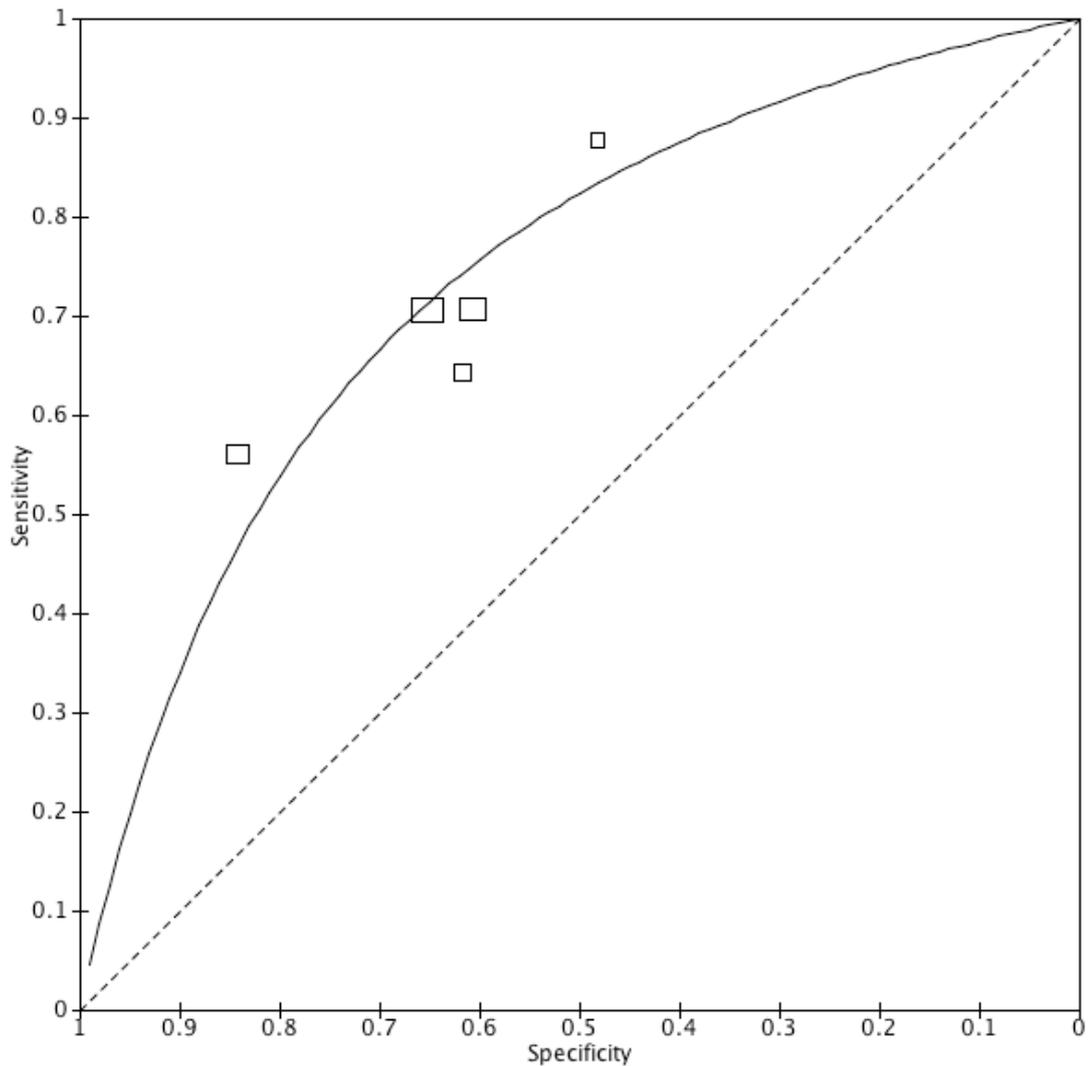


Figure 6-20 Summary ROC Curve - Patients Using Diuretics

Only one study, Januzzi et al. (2005), had a sensitivity of 56% and a specificity of 84%. This provided the best diagnostic predictive value of the five studies. This variable is likely to be associated with a medical history of heart failure, but could be considered a proxy for the diagnosis, as this is the main indication for commencing this treatment.

### 6.2.10 Use of ACE Inhibitors

Patients are prescribed angiotensin converter enzyme inhibitors for treatment for hypertension and as treatment for heart failure. It is reasonable to assume

that there may be a correlation between the patients receiving this treatment and the presence of heart disease leading to acute failure. Two studies provided this information as derived from 748 patients. This data shows reasonable specificity but poor sensitivity for acute heart failure. This is unlikely to be independent of a previous diagnosis of heart failure but may be considered a proxy marker. The positive likelihood ratio derived from the summary sensitivity and specificity from the available studies is 2.2, the negative likelihood ratio was 0.79.

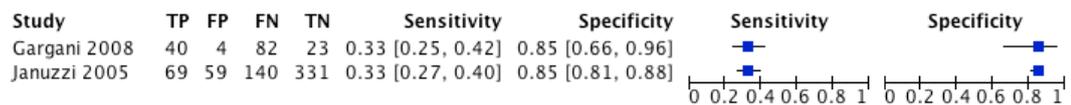


Figure 6-21 Forest Plot for the Use of ACE Inhibitors

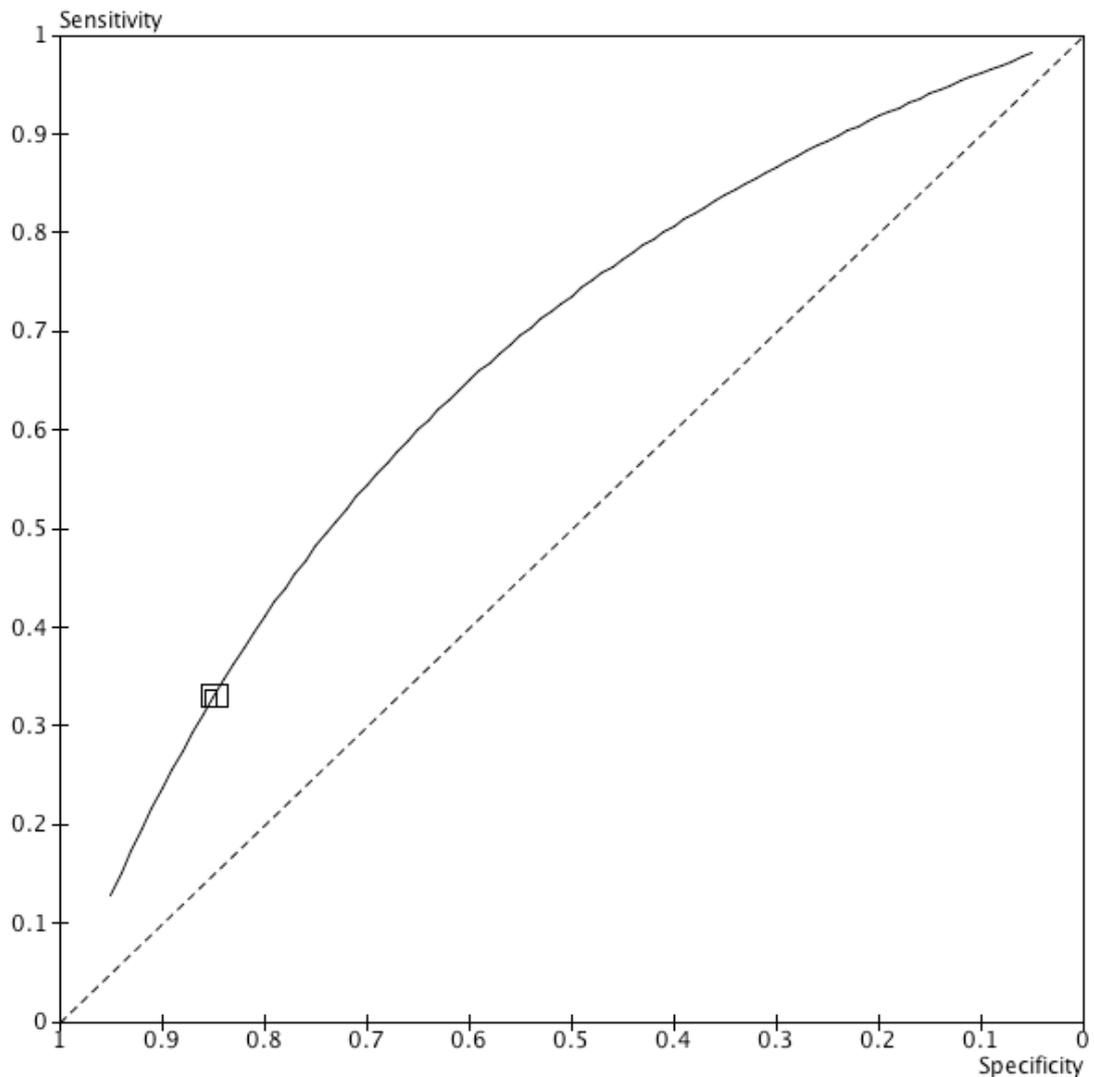


Figure 6-22 Summary ROC Curve for the Use of ACE Inhibitors

### 6.2.11 Use of ARB Medication

Angiotensin-receptor blocker (ARB) medication is primarily used in the treatment of heart failure and hypertension, particularly for patients who cannot tolerate the side effects of ACE inhibitors. As with ACE inhibitors, patients who are receiving this treatment may be expected to be more likely to have heart failure than patients who are not receiving them. One study

provided information about the use of the treatment in presenting patients and showed reasonable specificity but poor sensitivity. This corresponded with a PLR of 1.38 and a NLR of .86.



Figure 6-23 Forest Plot for Use of ARB Medication

### 6.2.12 Presence of Peripheral Oedema

Oedema due to venous congestion, secondary to a failing heart, may be present in the lower limbs, the abdomen or the sacral area. Various other conditions can also cause this condition. Twenty studies (8,171 patients) assessed this sign as a predictor of heart failure. While there was some association between the presence of peripheral oedema and the end-diagnosis of heart failure, in patients presenting with acute dyspnoea, the results demonstrated significant variability and failed to show real diagnostic value as shown in Figure 6-24 and Figure 6-25.

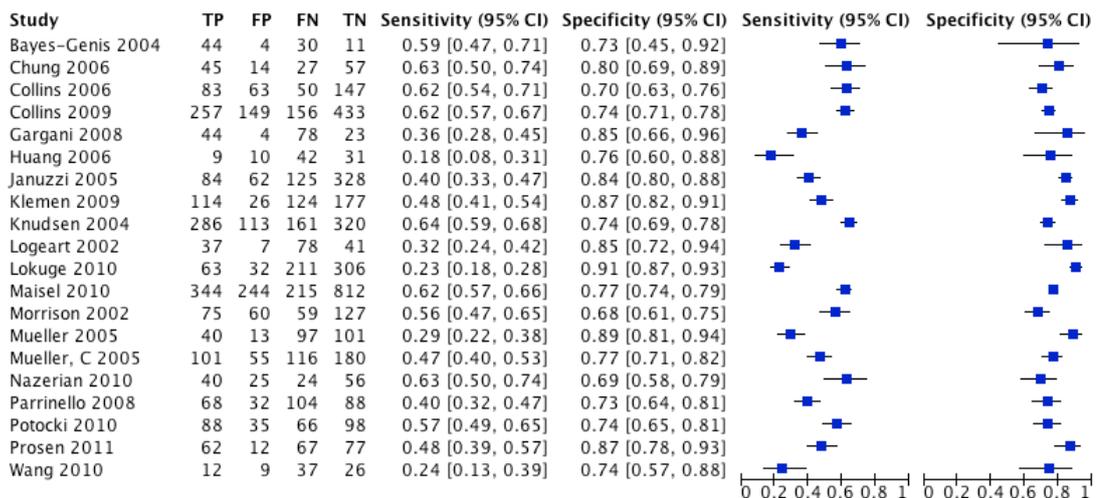


Figure 6-24 Sensitivity and Specificity - Presence of Peripheral Oedema

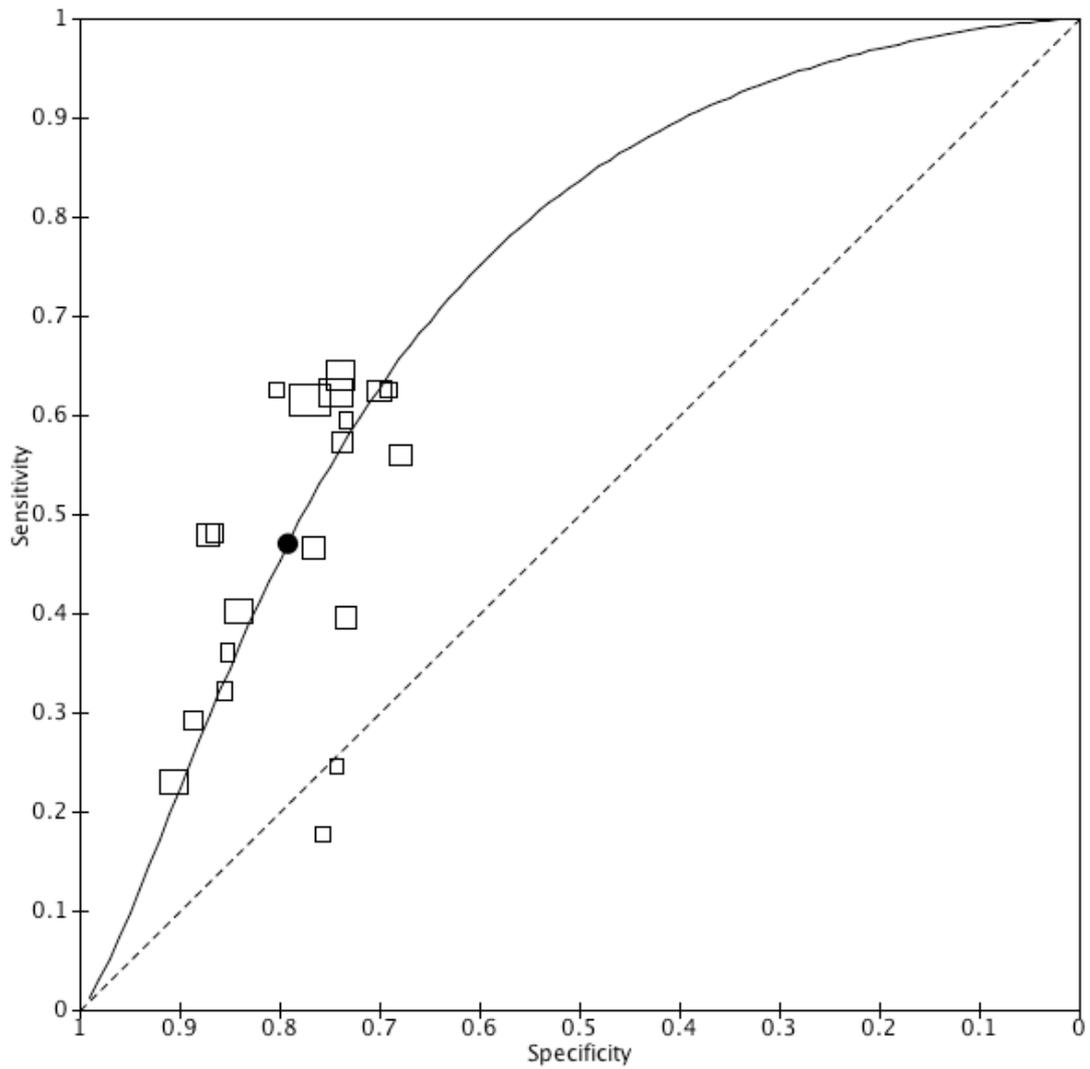


Figure 6-25 Summary ROC Curve - Peripheral Oedema

Examining the data regarding the presence of peripheral oedema for the diagnosis of heart failure only one study, Chung et al. (2009), provided a Youden's index greater than 0.4.

Analysis using MetaAnalyst provided the following data:

Specificity = 0.797 (95% CI 0.760 – 0.832)

Sensitivity = 0.467 (95% CI 0.399 – 0.540)

PLR = 2.30

NLR = 0.67

This test does not have sufficient sensitivity or specificity to provide useful diagnostic information for the diagnosis of heart failure in isolation but may have value if combined with other, more sensitive variables.

### 6.2.13 Presence of Dyslipidaemia

The presence of a raised cholesterol level is associated with an increased risk of ischaemic heart disease and so could be associated with the presence of heart failure. Three studies provided information about this condition in 1,753 patients. These small studies demonstrate a reasonable specificity, although with wide confidence intervals, but a low sensitivity.

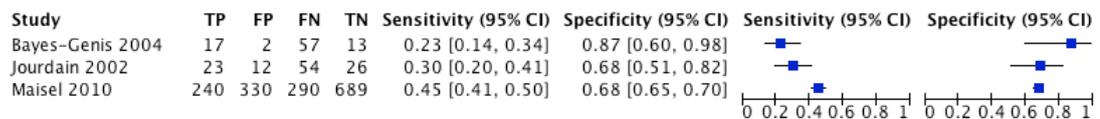


Figure 6-26 Forest Plot - History of Dyslipidaemia

### 6.2.14 The Presence of Rales

The sudden opening of small fluid-filled airways is thought to cause the presence of 'rales', crepitations or crackles during inspiration that are audible on auscultation of the lung fields (Forgacs, 1978). The presence of pulmonary oedema with oedematous small bronchi is thought to cause the late opening. Similar findings can be found in patients with pulmonary fibrosis. Nineteen studies examined this clinical finding and presented data derived from 7,731 patients. There is considerable variation in the study findings with a tendency towards reasonable specificity, but limited sensitivity.

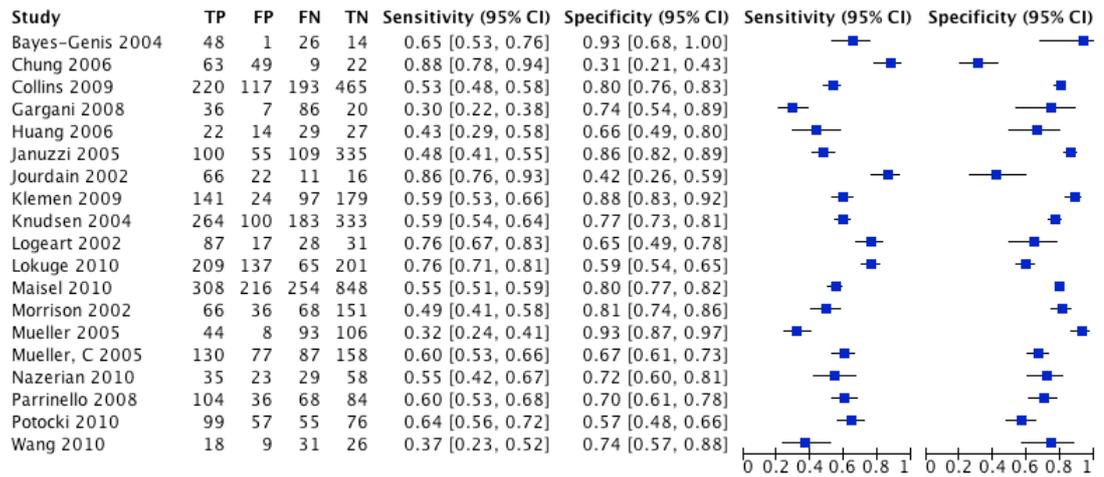


Figure 6-27 Sensitivity and Specificity of Presence of Rales

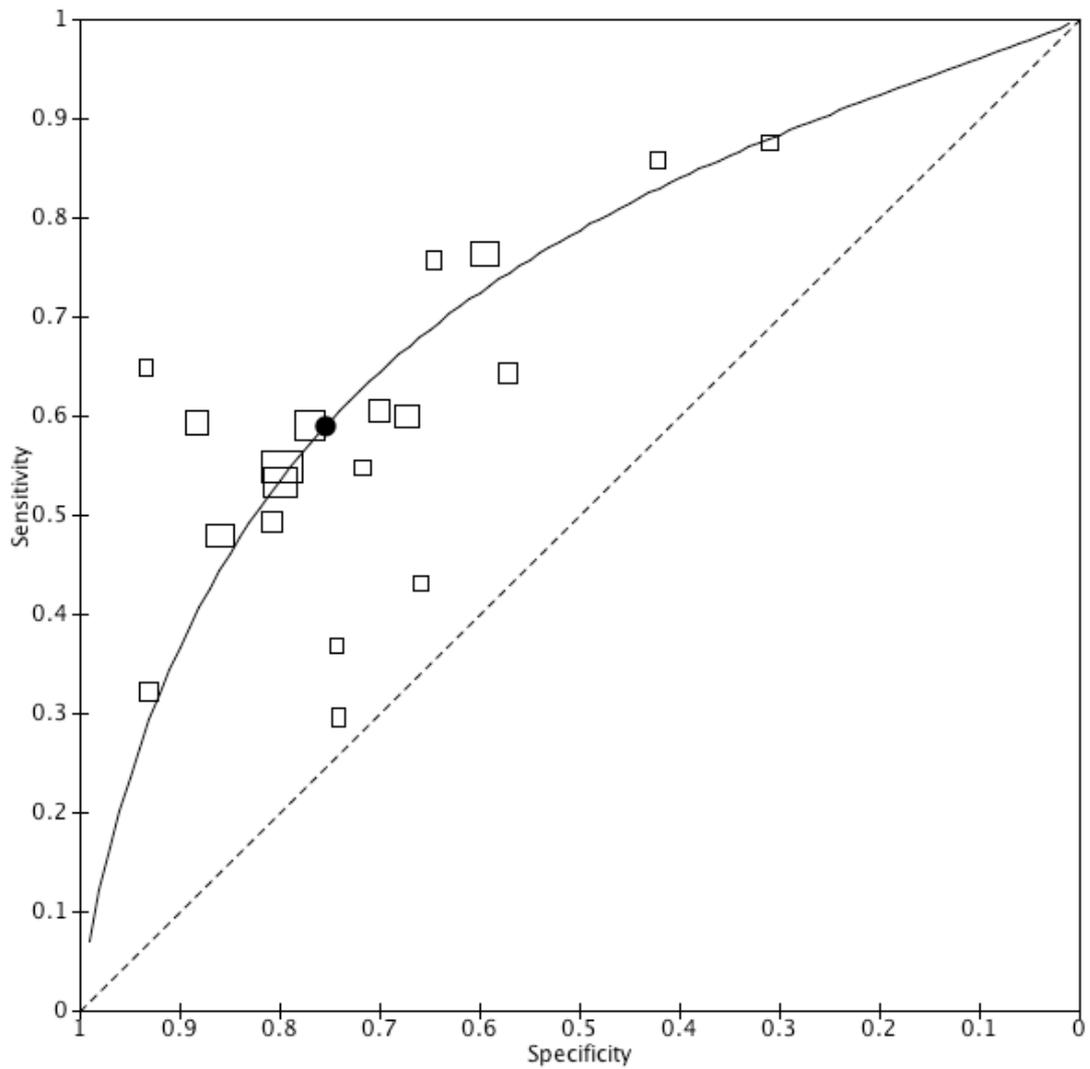


Figure 6-28 Summary ROC Curve - Presence of Rales Presence of Rales

This data was entered into Meta-DiSc and the following results were obtained.

Heterogeneity chi-squared = 44.82 (d.f. = 17) p = 0.000

Inconsistency ( $I^2$ ) = 62.1%

Estimate of between-study variance (Tau-squared) = 0.0872

Data from MetaAnalyst

Specificity = 0.736 (95% CI 0.656 – 0.802)

Sensitivity = 0.591 (95% CI 0.509 – 0.670)

PLR = 2.34

NLR = 0.56

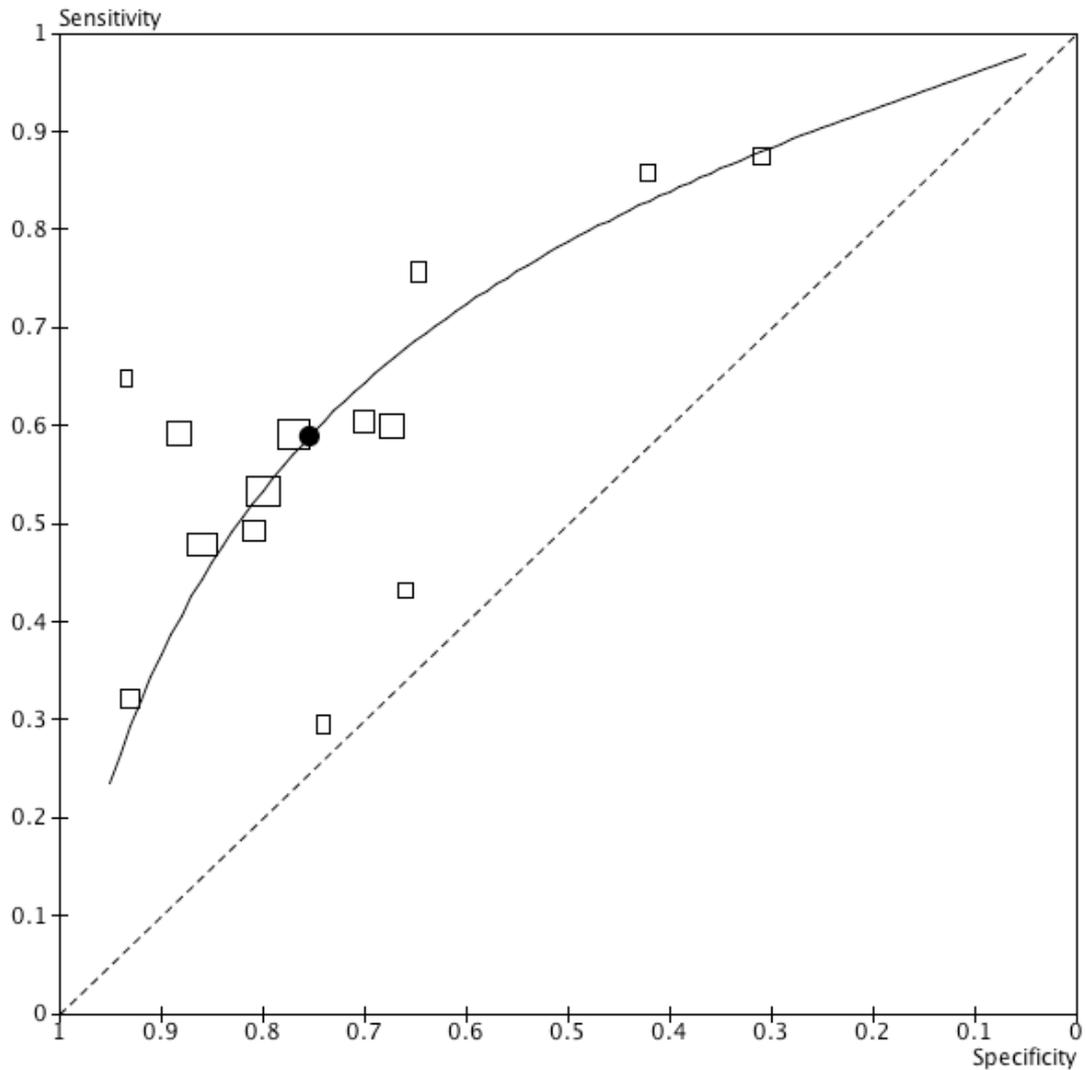


Figure 6-29 Summary ROC Curve - Presence of Rales

Although the data appears to be skewed by three studies that appear to be out-lying on the SROC chart, eliminating these studies from the analysis did not improve the diagnostic utility of this test. The summary point suggests that this finding has limited diagnostic utility.

### 6.2.15 Presence of Wheeze

The presence of a polyphonic expiratory wheeze is usually considered to be due to narrowing of small airways (Forgacs, 1978). This clinical finding is suggestive of asthma or chronic obstructive pulmonary disease but can also be present in patients with heart failure. In the results of the search nine papers provided information about the presence of wheeze from 5,274

patients. It can be seen that from the forest plot in Figure 6-30 and the summary ROC curve in that the absence of wheeze is suggestive of a cardiac cause for the patient's symptoms.

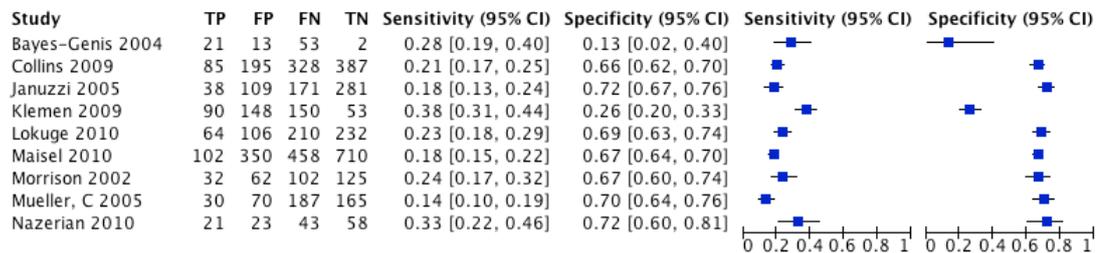


Figure 6-30 Sensitivity and Specificity of Absence of Wheeze

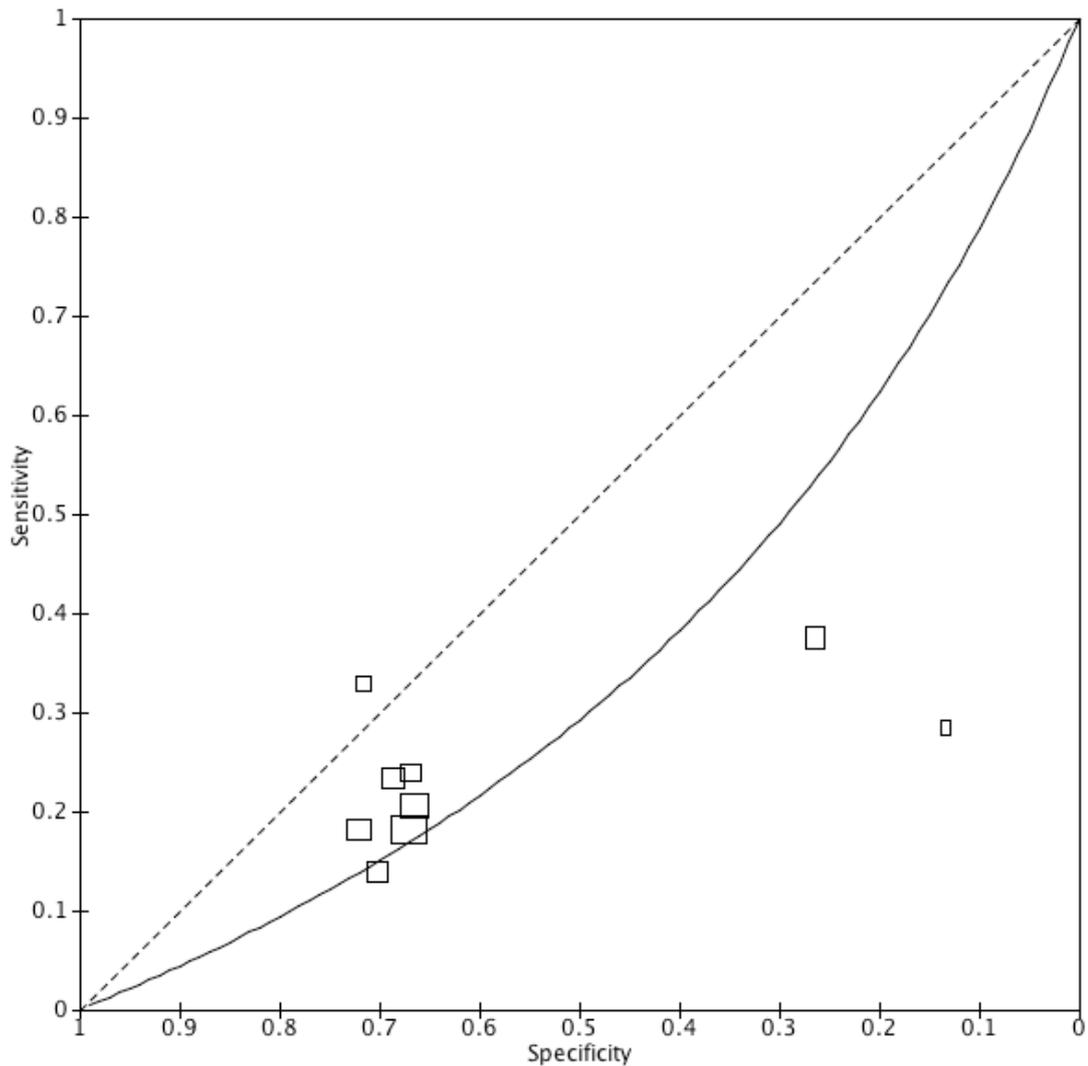


Figure 6-31 Summary ROC Curve - Absence of Wheeze

Bayes-Genis et al. (2004), demonstrated good specificity and sensitivity for this clinical finding but in the other studies it had unremarkable diagnostic value and there was marked heterogeneity in the results. This single study was an outlier on the summary ROC curve, and following its exclusion, absence of wheeze showed poor diagnostic utility overall.

### 6.2.16 Patient Taking Beta Antagonist Medication

Beta antagonists or beta-blocker drugs have been shown to improve long-term survival in patients with heart failure and reduce the damaging remodelling that can occur in the failing heart (Krum et al., 1995). The beneficial effects are thought to be due to down-regulation of the sympathetic

nervous system response(Packer, 1995b). Patients are also commenced on beta antagonists as treatment for hypertension, itself a significant risk factor for the development of heart failure. It may be reasonable to assume that patients who are taking this class of medication are more likely to present with heart failure. Six studies provide information about whether 4,007 patients were receiving this treatment prior to their attendance, the data is summarised in Figure 6-32.

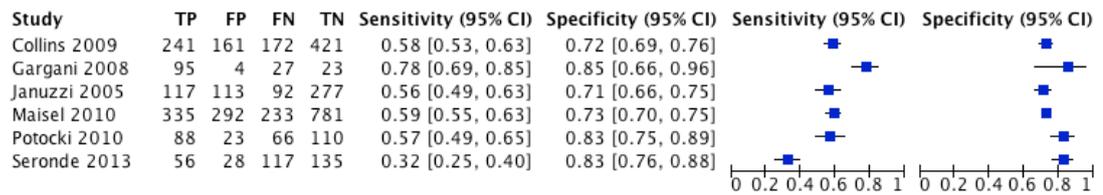


Figure 6-32 Forest Plot for Patients Taking Beta-blockers

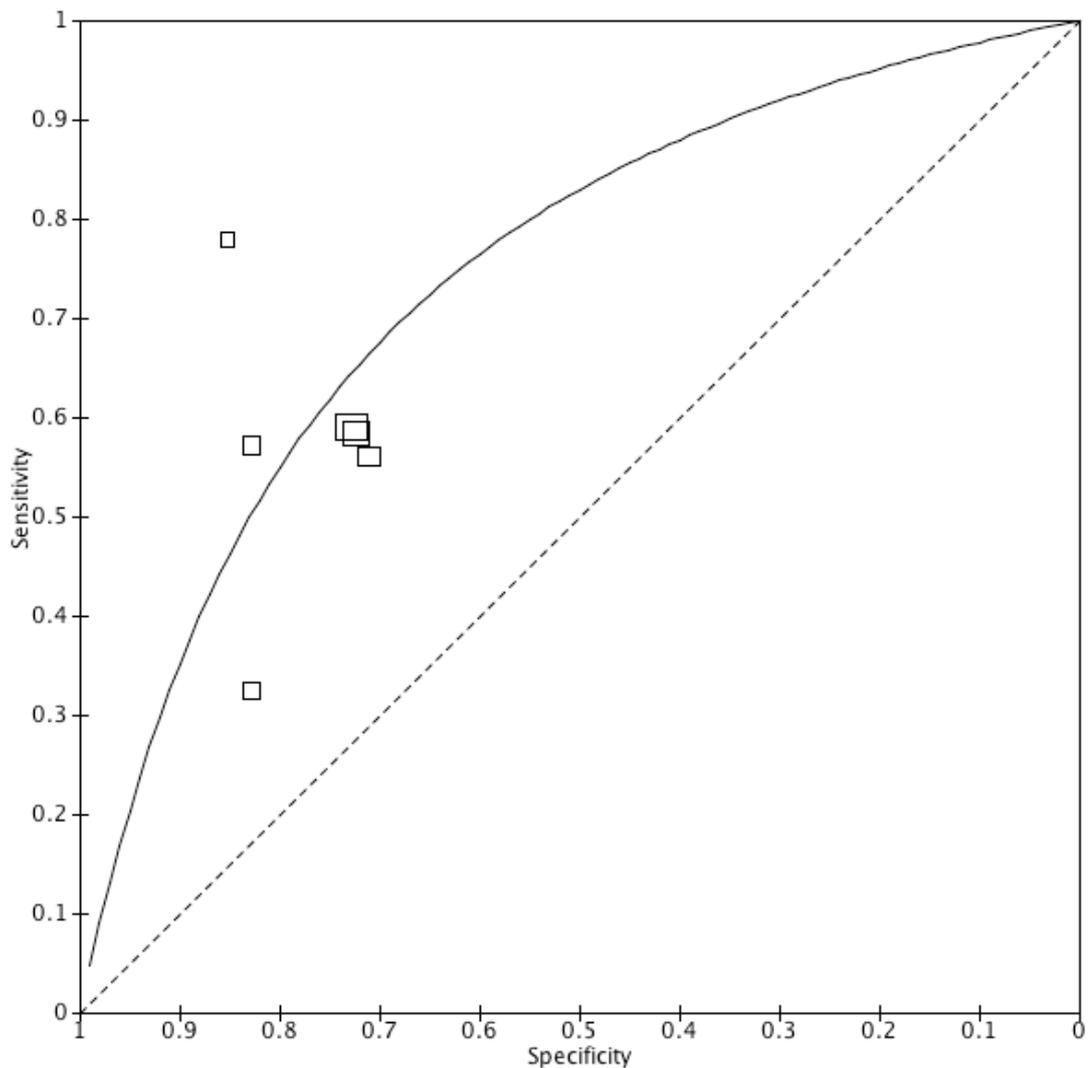


Figure 6-33 Summary ROC Curve for Patients Taking Beta-blockers

The summary ROC curve in **Error! Reference source not found.** suggests some diagnostic value from knowing if a patient has been prescribed this form of medicine.

The data was entered into Meta-DiSc that provided the following information about the heterogeneity of the diagnostic odds ratios.

Heterogeneity chi-squared = 16.45 (d.f. = 5) p=0.006

Inconsistency ( $I^2$ ) = 69.6%

Estimate of between-study variance (Tau-squared) = 0.0805

## Meta Analyst analysis

Specificity = 0.765 (95% CI 0.692 – 0.835)

Sensitivity = 0.571 (95% CI 0.441 – 0.692)

Positive Likelihood Ratio = 2.43

Negative LR = 0.56

This result suggests that this variable has some diagnostic value although the performance is skewed by a single study, Gargani et al. (2008). This finding is unlikely to have diagnostic value independently of a positive medical history of heart failure, as this would be a common reason to commence a patient on this form of medication.

### 6.2.17 History of Chronic Obstructive Pulmonary Disease

Patients with a history of Chronic Obstructive Pulmonary Disease (COPD), also known as Chronic Obstructive Airway Disease (COAD), may present in a similar fashion to patients with heart failure, as they tend to have periods of exacerbation and remission of their symptoms that includes breathlessness. The clinical picture is further confused as they share some common risk factors and indeed, often co-exist. The distinction between the two conditions is likely to affect clinical out-come as the treatments for each condition are very different. Starting inappropriate treatment on the basis of an incorrect diagnosis is likely to be at best ineffectual, and at worst may cause harm. As the condition of COPD tends to have intermittent episodes of exacerbation, a previous diagnosis of this condition could be assumed to make heart failure less likely to be the cause of the patient symptoms. Fifteen studies including a total of 6,989 patients provided data about a previous diagnosis of COPD. The summary ROC curve of this data is shown in **Error! Reference source not found**. This demonstrates that there is likely to be an association with a history of COPD and a cause other than heart failure for patients presenting

with acute dyspnoea. The results are variable and may reflect differences in prevalence of COPD within the recruited population.

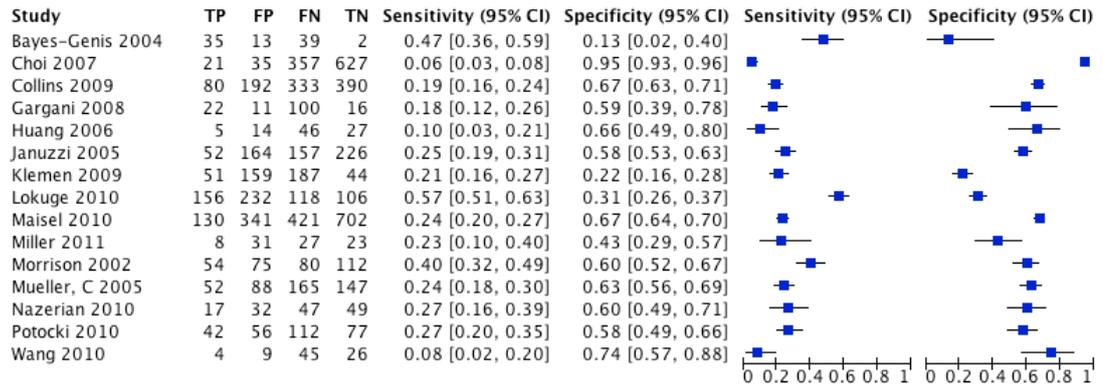


Figure 6-34 Forest Plot - History of COPD

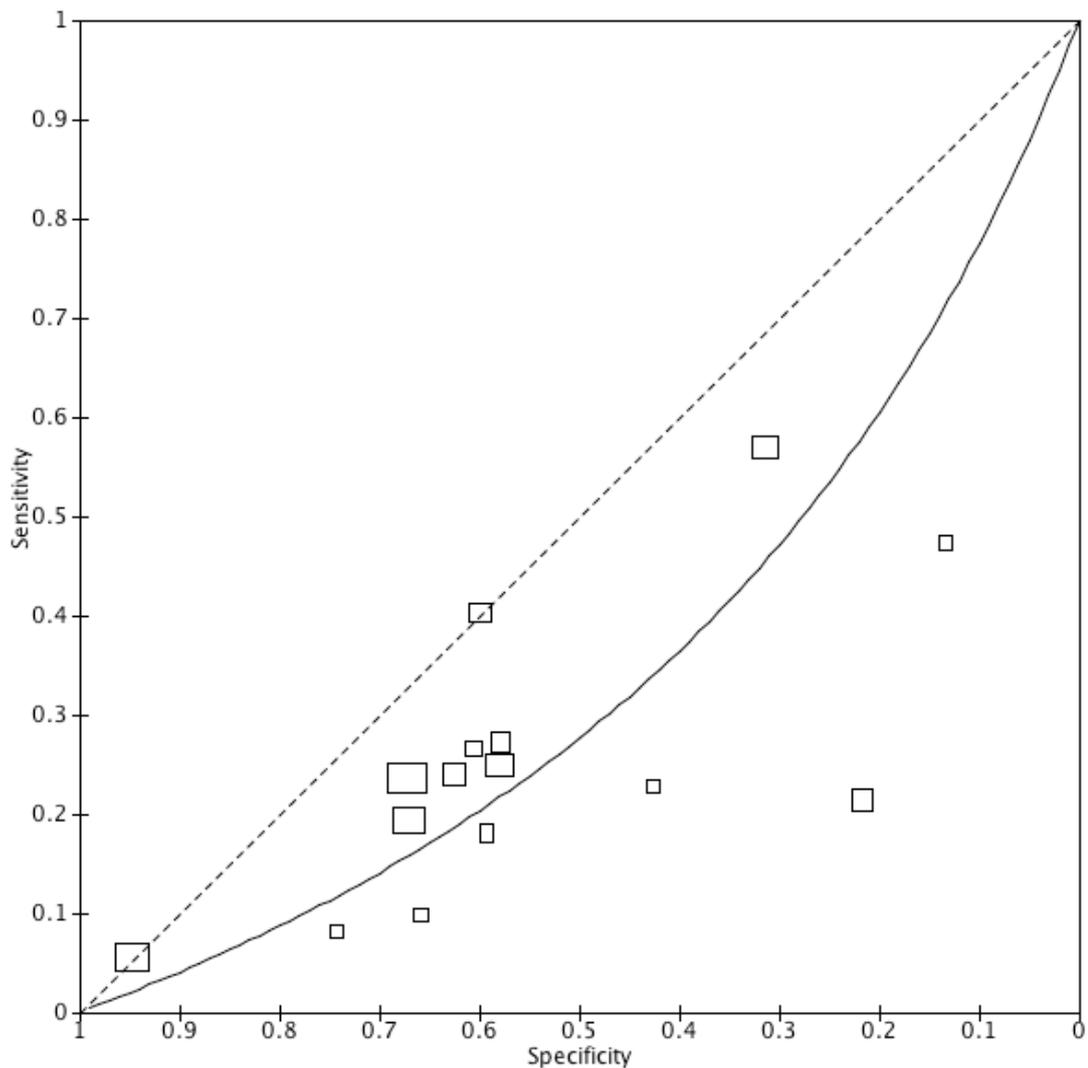


Figure 6-35 Summary ROC Curve No History of COPD

Only one study, Klemen et al. (2009) demonstrated high levels of sensitivity and specificity. The other included studies demonstrated little overall diagnostic value and this is shown by the relatively flat ROC curve. This study was carried out in a pre-hospital setting although all of the patients were subsequently admitted to hospital. This may mean that it represents a different population from the other studies. However, eliminating this study still leaves marked heterogeneity between the studies that would not appear to be explained by a positivity threshold effect. With the exception of the

single excluded study, there does not seem to be any significant diagnostic value to this historical variable.

### 6.2.18 History of Ischaemic Heart Disease

Twenty-four studies provided information about a history of ischaemic heart disease. This category was taken to include all patients described as having ischaemic heart disease or a history of myocardial infarction or angina. A total of 9,902 patients were included in this analysis. The calculated values for sensitivity and specificity from each study are provided in Figure 6-36. There is considerable variation that suggests a significant degree of heterogeneity within the studies. Constructing a summary ROC curve demonstrates an approximate curved shape overall that may suggest a threshold effect as shown in Figure 6-37. It is reasonable to expect a threshold effect due to variations in the definition of ischaemic heart disease. No studies provided a sensitivity or specificity high enough to suggest a diagnostic role for heart failure.

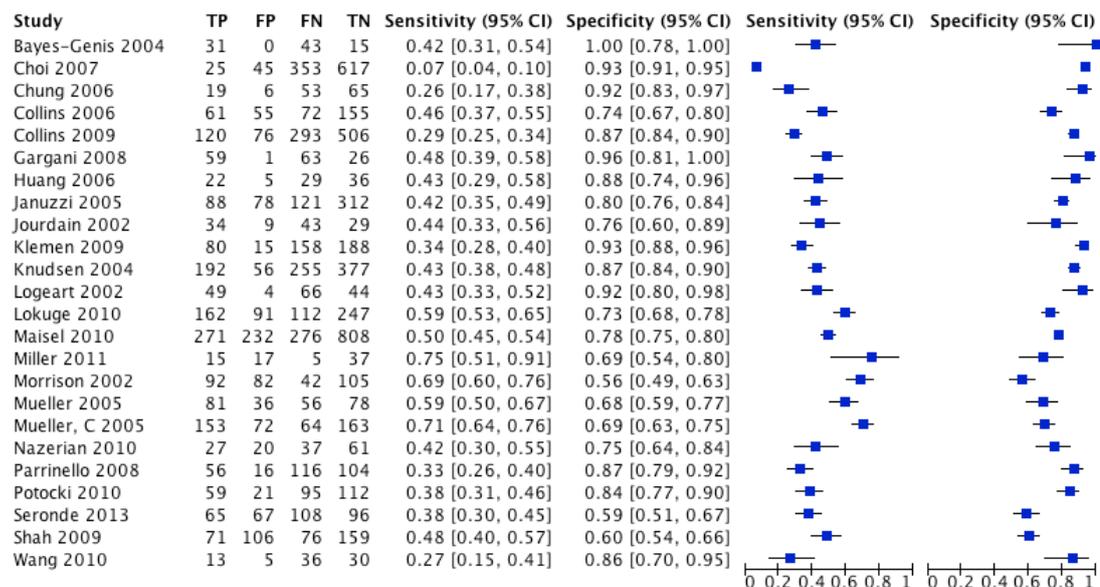


Figure 6-36 Sensitivity and Specificity of IHD as a Predictor of Heart Failure

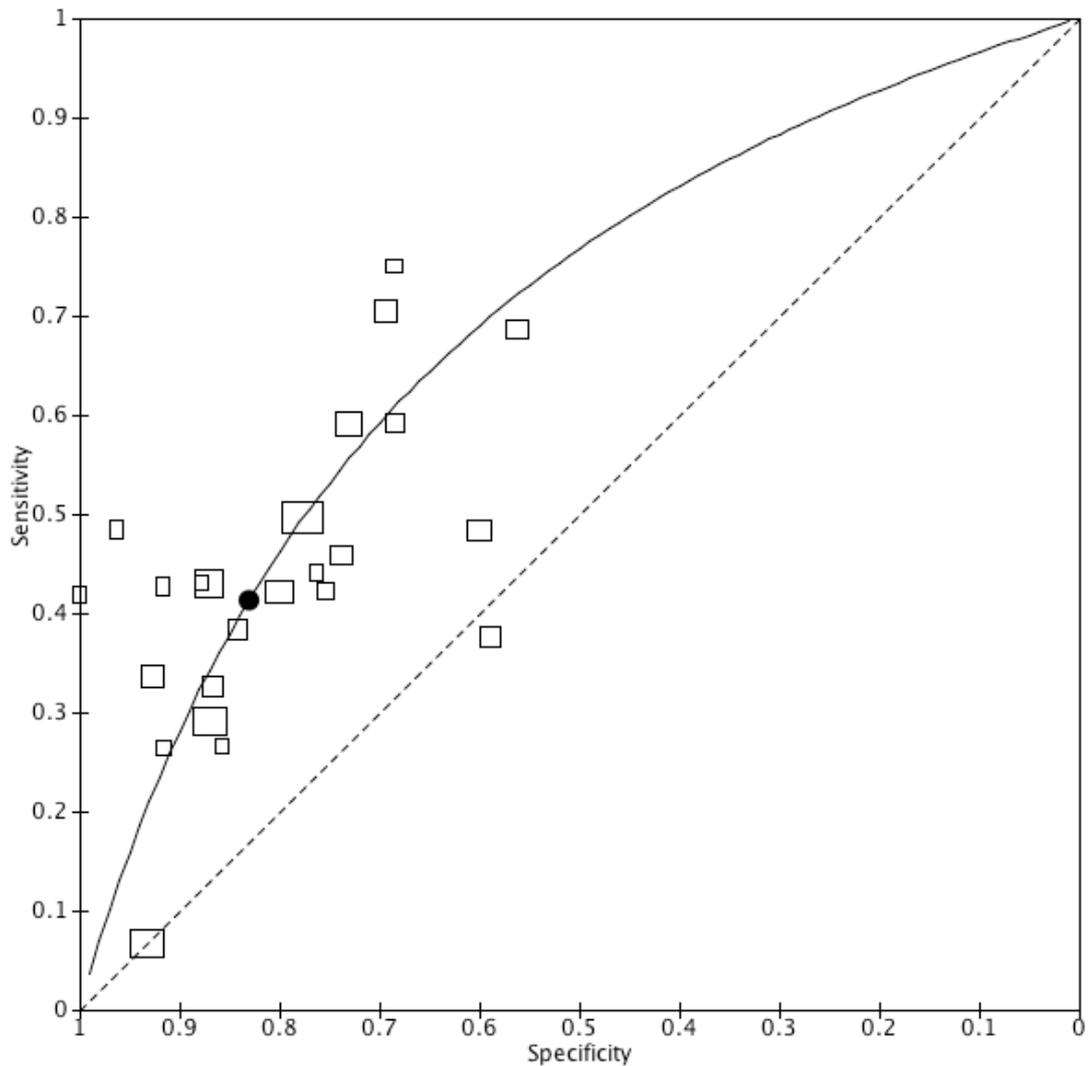


Figure 6-37 Summary ROC Curve – IHD

### 6.3 Variables with Limited Applicability

Five variables achieved a Youden's Index of 0.4 or greater suggesting at least reasonable diagnostic utility but have been excluded from further analysis on the basis that they are not generally available and could not be incorporated into standard emergency medicine practice without specific training or equipment.

One study, Lo et al. (2007), looked at impedance cardiography and another study, Parrinello et al. (2008), examined bioelectrical impedance. While these studies had impressive results suggesting useful diagnostic function, each

investigation is represented by single study representing novel technology that is not routinely available in emergency departments. Therefore, the usefulness of the method clinically is limited.

A study by Gegenhuber et al. (2006), demonstrated good sensitivity and specificity for mid-range proANP but the assay for this potential biomarker is not available for use in emergency departments, and even in the reported study it was not superior to BNP or NT-proBNP.

Two studies, Gargani et al. (2008) and Liteplo et al. (2009), looked at the use of ultrasound and in particular, the presence of 'B' lines or comets, to detect pulmonary oedema. While these studies had promising results with high specificity and reasonable sensitivity, they used slightly different techniques and cut-off values to define a positive result and are therefore not suitable for combination for further analysis. As the presence of ultrasound becomes more ubiquitous in emergency departments it may have a diagnostic role in the future but cannot be recommended as part of a decision pathway at the current time.

### 6.3.1 Impedance Cardiography

One study examined the use of impedance cardiography to diagnose heart failure in the breathless patient. In this process a small current is passed through the patient's thorax and the changes in current are used to calculate aortic blood volume and flow over time. This data can be used to calculate stroke volume and cardiac index. The study involved 52 patients. The results of this study are promising but the interpretation needs to be cautious due to the small size of the study.

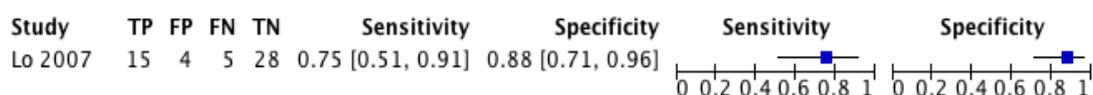


Figure 6-38 Forest Plot for Impedance Cardiography

### 6.3.2 Bioelectrical Impedance Analysis

The measurement of bioelectrical impedance can be carried out non-invasively using an impedance plethysmograph. This passes an electric current through specified areas of the body and measures the impedance to the current flow. The associated theory is that increased body water due to fluid overload will decrease the bio-impedance. A study by Parrinello et al. (2008) has looked at the diagnostic utility of this and has provided diagnostic data for two cut-off values.

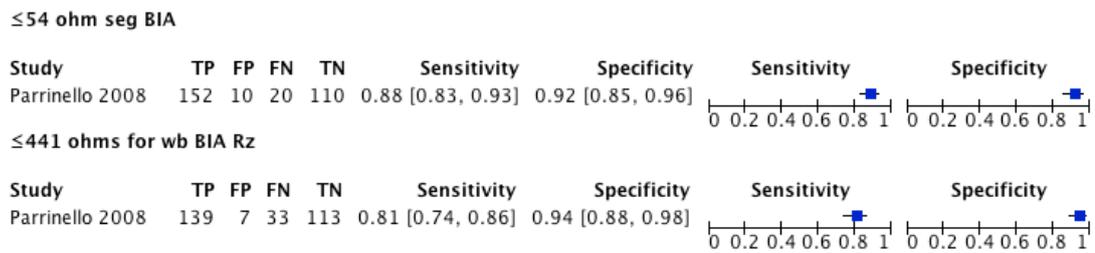


Figure 6-39 Forest Plot for Bioelectrical Impedance Analysis

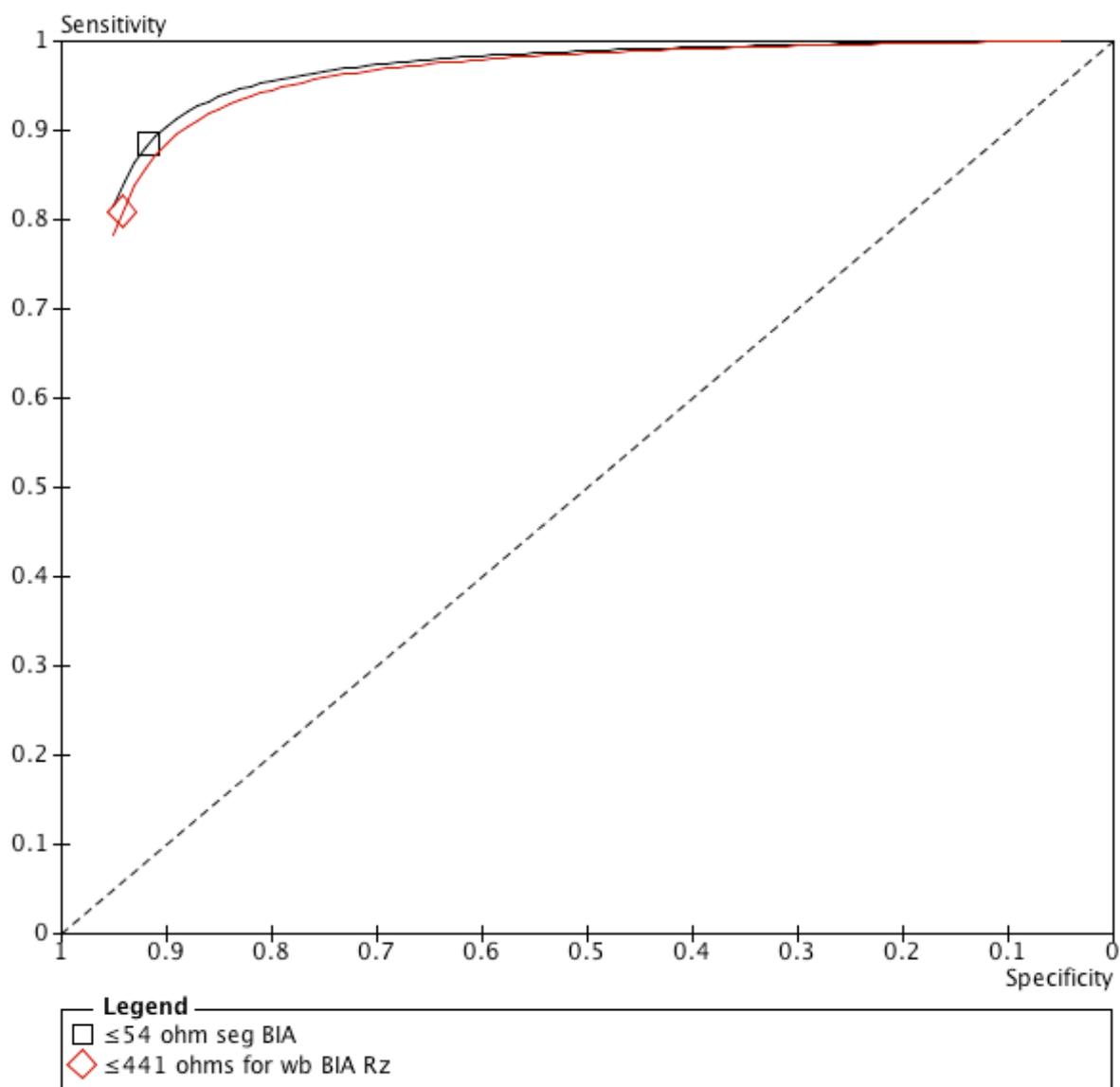


Figure 6-40 ROC Curves for Bioelectrical Impedance Analysis

This is very suggestive of significant diagnostic utility but was based on a single study that was used to derive the cut-off values for this diagnostic test. The technique used appears to be relatively simple and is non-invasive but it is not standard practice in the hospital setting and validation would be required before advocating this practice.

### 6.3.3 Acoustic Third Heart Sound

Two papers, by the same authors, Collins, Lindsell, Peacock, Hedger, et al. (2006) and (Collins et al., 2009), have looked at the utility of electronically recording the presence of a third heart sound in the diagnosis of heart failure. The equipment uses microphones to record the heart sounds while recording the patient's electrocardiograph. The auscultation of the heart sounds is analysed along with the electrocardiograph using a computer algorithm and these are combined to provide a diagnosis. The two studies involved a total of 1,338 patients.

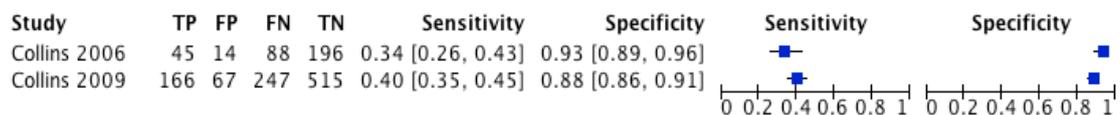


Figure 6-41 Forest Plot of Acoustic Third Heart Sound

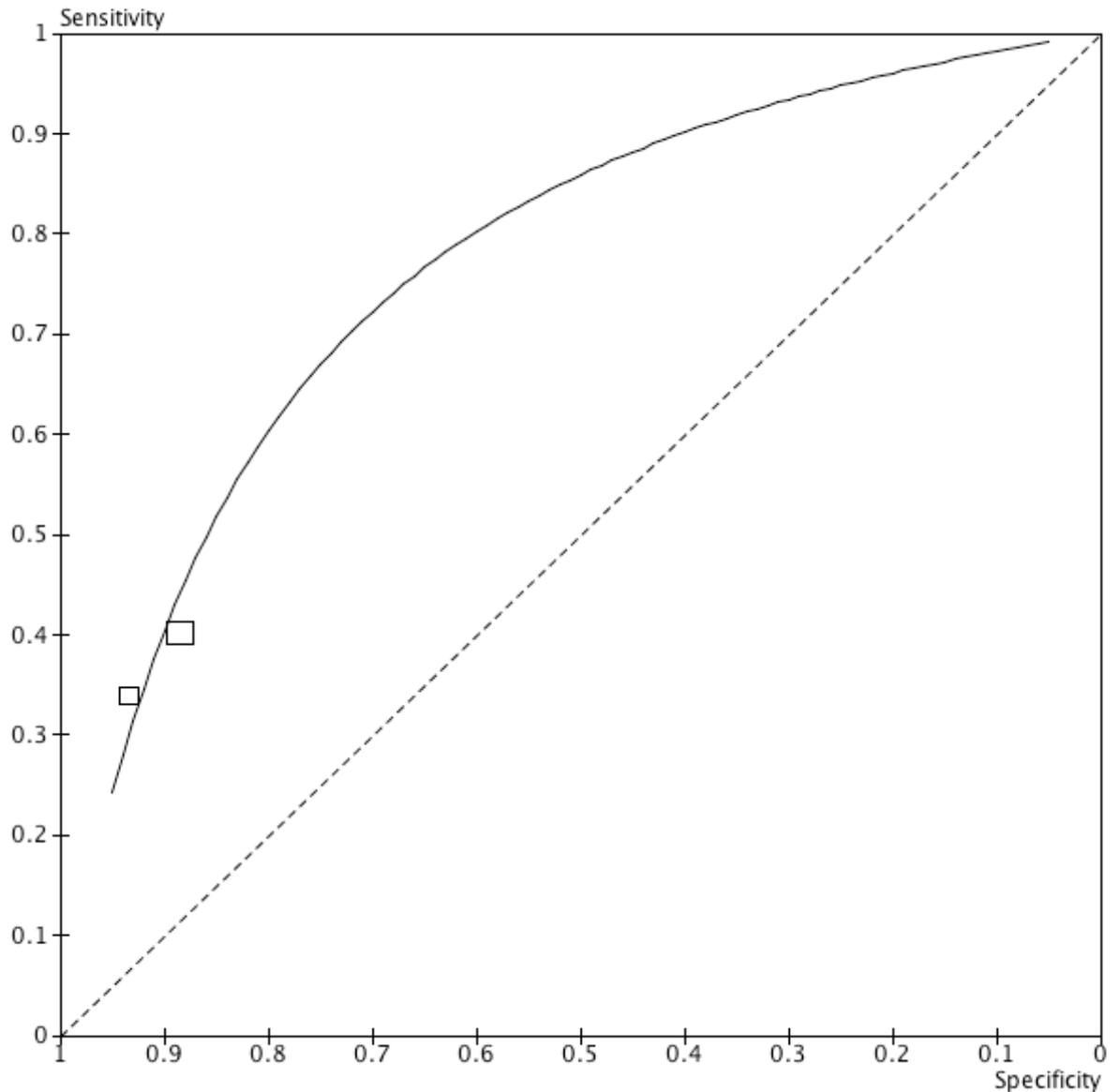


Figure 6-42 Summary ROC Curve for Acoustic Third Heart Sound

The studies suggest that there is good specificity for this test, especially compared with standard auscultation, but further validation is required.

### 6.3.4 Mid-regional pro-Atrial Natriuretic Peptide

Gegenhuber et al. (2006) explored the diagnostic value of 'N' terminal pro-Atrial Natriuretic Peptide (NT-proANP) for the condition of acute decompensated heart failure using an immunoassay that was targeted at the mid-regional epitopes of this molecule, this was also used by Maisel et al. (2010). The data from these studies appear promising but further examination

and validation of this test is required before widespread clinical use could be advocated.

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Gegenhuber 2006	130	50	7	64	0.95 [0.90, 0.98]	0.56 [0.47, 0.65]		
Maisel 2010	551	430	17	643	0.97 [0.95, 0.98]	0.60 [0.57, 0.63]		

Figure 6-43 Forest Plots for MR-proANP

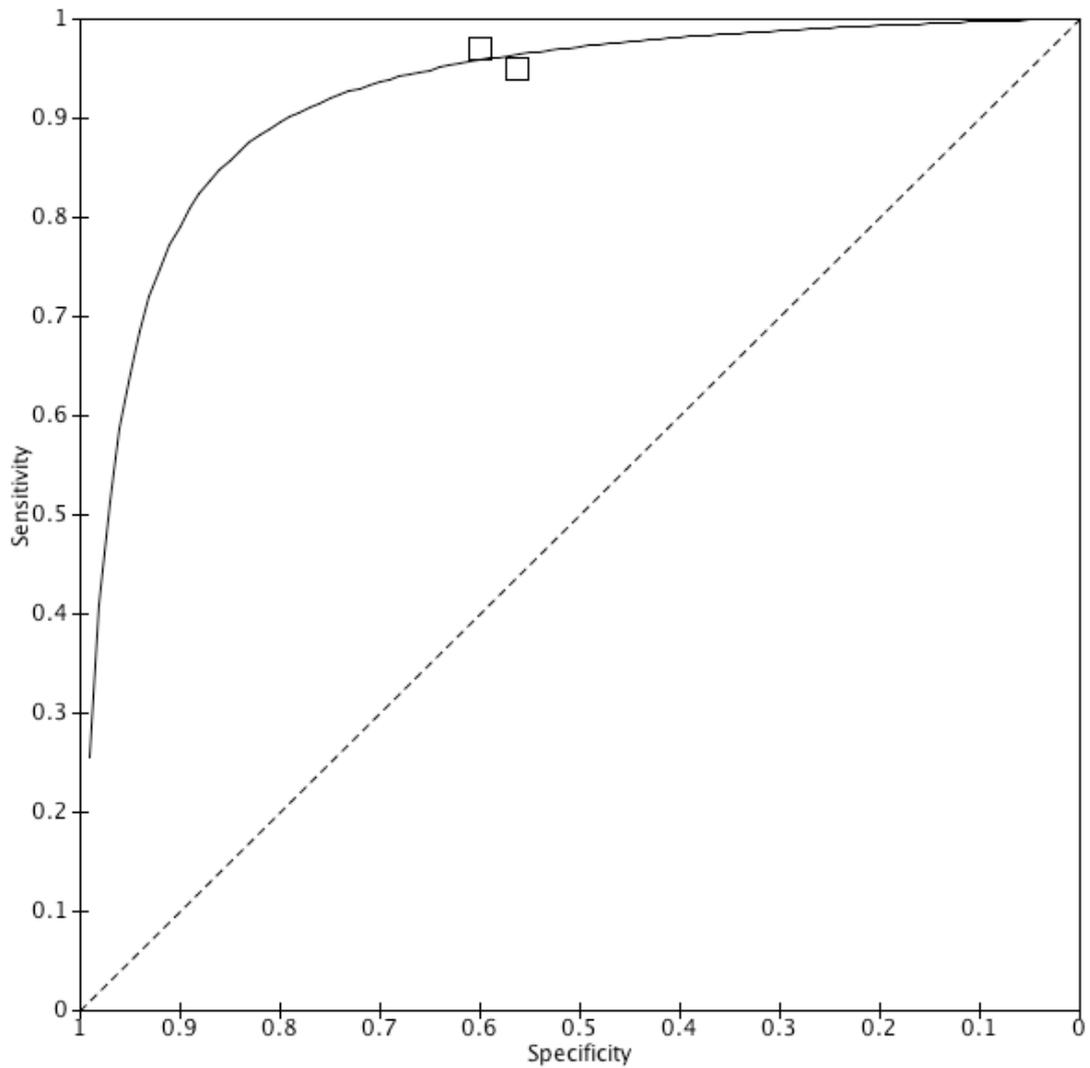


Figure 6-44 ROC Curves for MR-proANP

### 6.3.5 Use of Ultrasound

The presence of extra fluid within the lung tissue causes linear artefacts to be present when using ultrasound to scan lungs. These artefacts are referred to

as 'B' lines or 'ultrasound lung comets'. The presence of these have been demonstrated to correspond with NYHA functional class and signs on chest x-rays in-keeping with pulmonary oedema (Jambrik, 2004; Frassi et al., 2007). Two studies, Gargani et al. (2008) and Liteplo et al. (2009), involving 243 patients, reported on this finding as shown in Figure 6-45 and Figure 6-46.

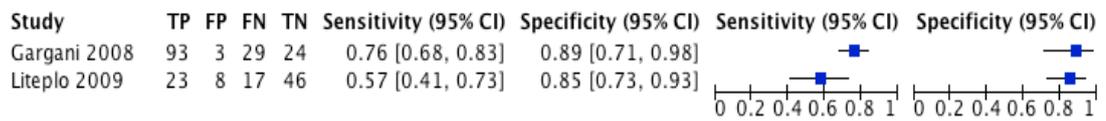


Figure 6-45 Forest Plot for Performance of Ultrasound

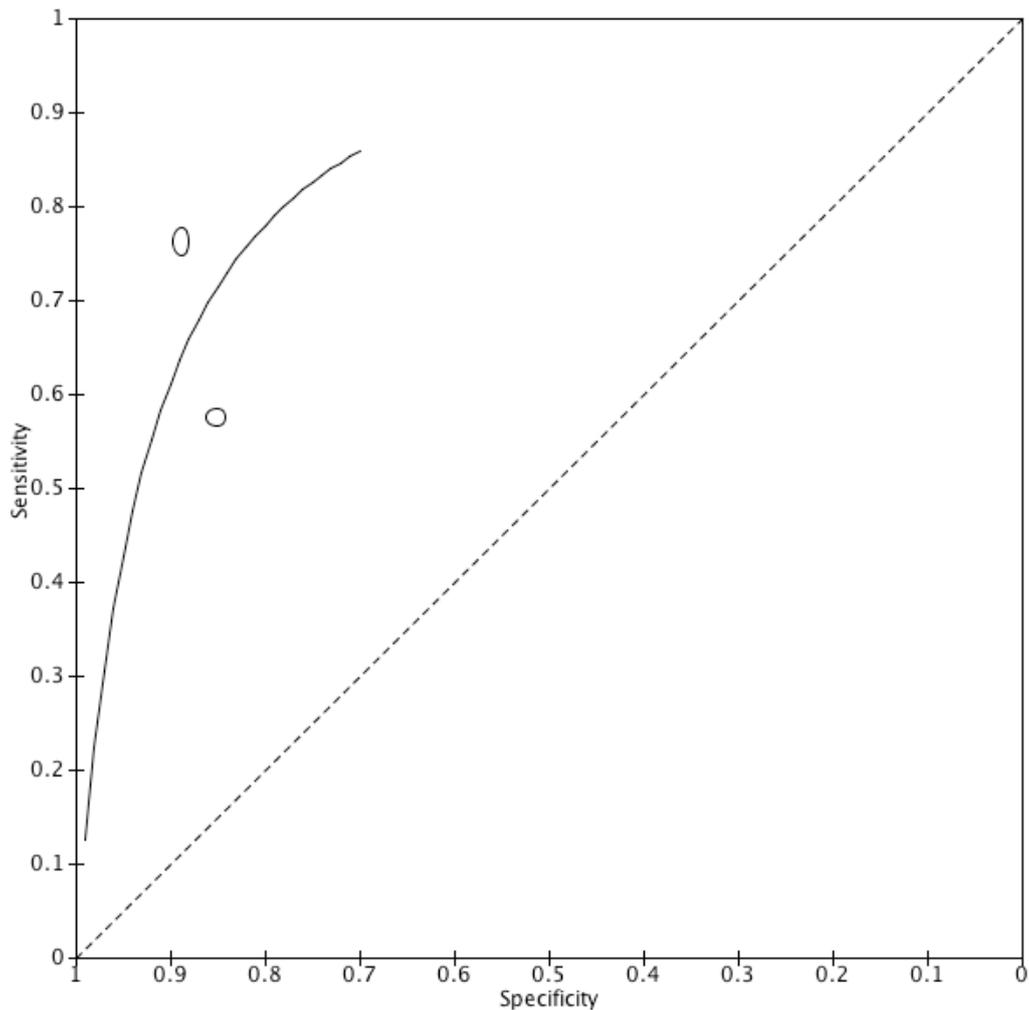


Figure 6-46 Summary ROC Curve for Performance of Ultrasound

## 6.4 Variables with High Specificity

Ten variables had very high specificity but low sensitivity. This means that this finding dramatically increases the likelihood of heart failure although the absence of this variable does not rule the condition out. Only a small proportion of the patients in each study had the variable present so the clinical utility remains limited. Summary points have been calculated using bivariate meta-analysis and the positive likelihood ratios calculated for each variable.

### 6.4.1 History of Paroxysmal Nocturnal Dyspnoea

This symptom is described in textbooks as being pathognomonic of heart failure. The patient typically describes waking at night with a sudden feeling of acute dyspnoea one to two hours after lying down. It is considered that this is due to increased venous return when the patient maintains a supine position (O'Connor, 2005). Eight studies provided information about this symptom that was derived from 2,639 patients as shown in Figure 6-47.

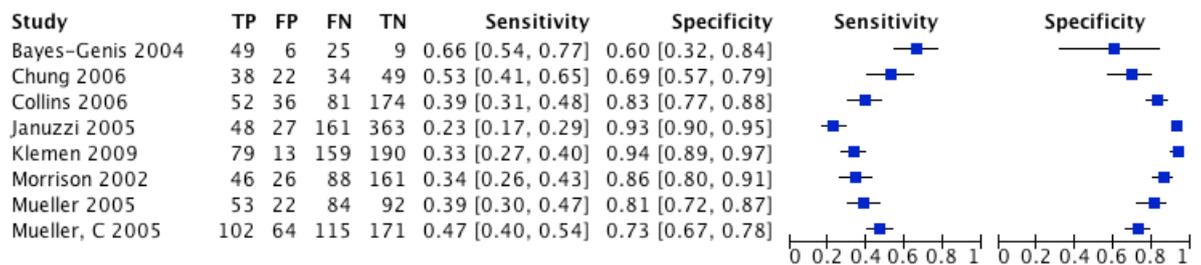


Figure 6-47 Sensitivity and Specificity PND

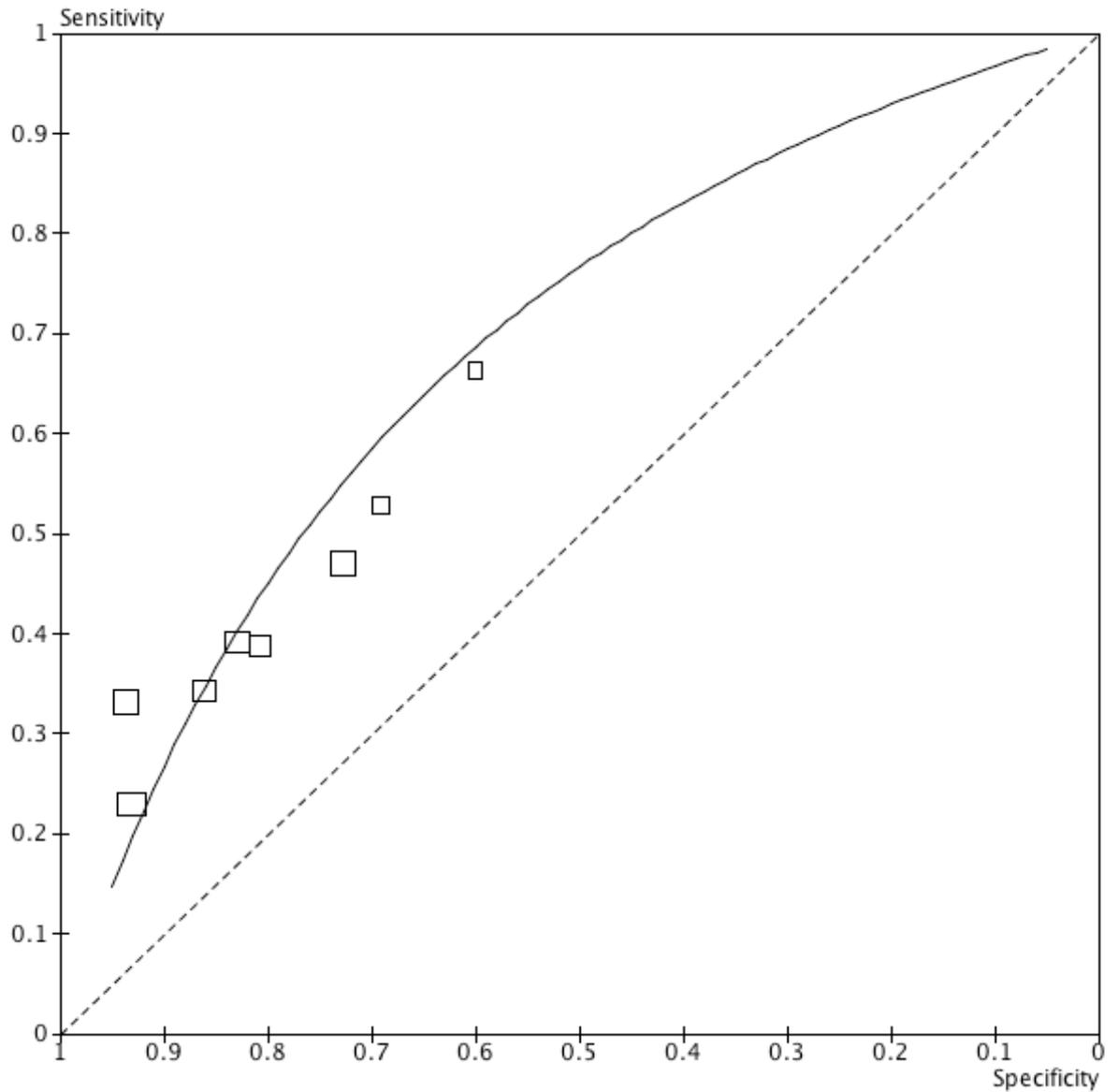


Figure 6-48 Summary ROC Curve PND

The following results were obtained from Meta-DiSc:

Heterogeneity chi-squared = 10.68 (d.f. = 7) p=0.153

Inconsistency ( $I^2$ ) = 34.5

Estimate of between-study variation (Tau-squared) = 0.0432

Despite the moderate inconsistency score the summary ROC curve displays a distribution that would be consistent with a threshold effect.

Bivariate meta-analysis using MetaAnalyst software provided the following summary point:

Sensitivity = 0.409 (95% CI 0.316 – 0.511)

Specificity = 0.816 (95% CI 0.697 – 0.900)

Although some studies suggested high specificity for this variable, the summary point suggests limited diagnostic value, and that is reflected in an overall positive likelihood ratio value of 2.28.

### 6.4.2 Presence of diabetes

Having diabetes increases the risk of developing cardiac disease. Sixteen studies with a total of 5,861 patients provided information about a history of diabetes in patients with heart failure. This provided reasonable specificity but poor sensitivity for the presence of heart failure as illustrated in Figure 6-49 and Figure 6-50.

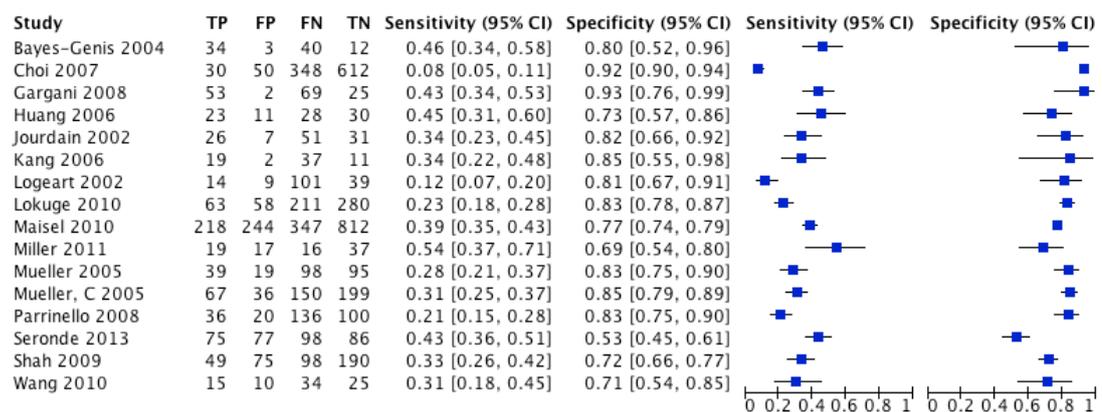


Figure 6-49 Forest Plot of Presence of Diabetes

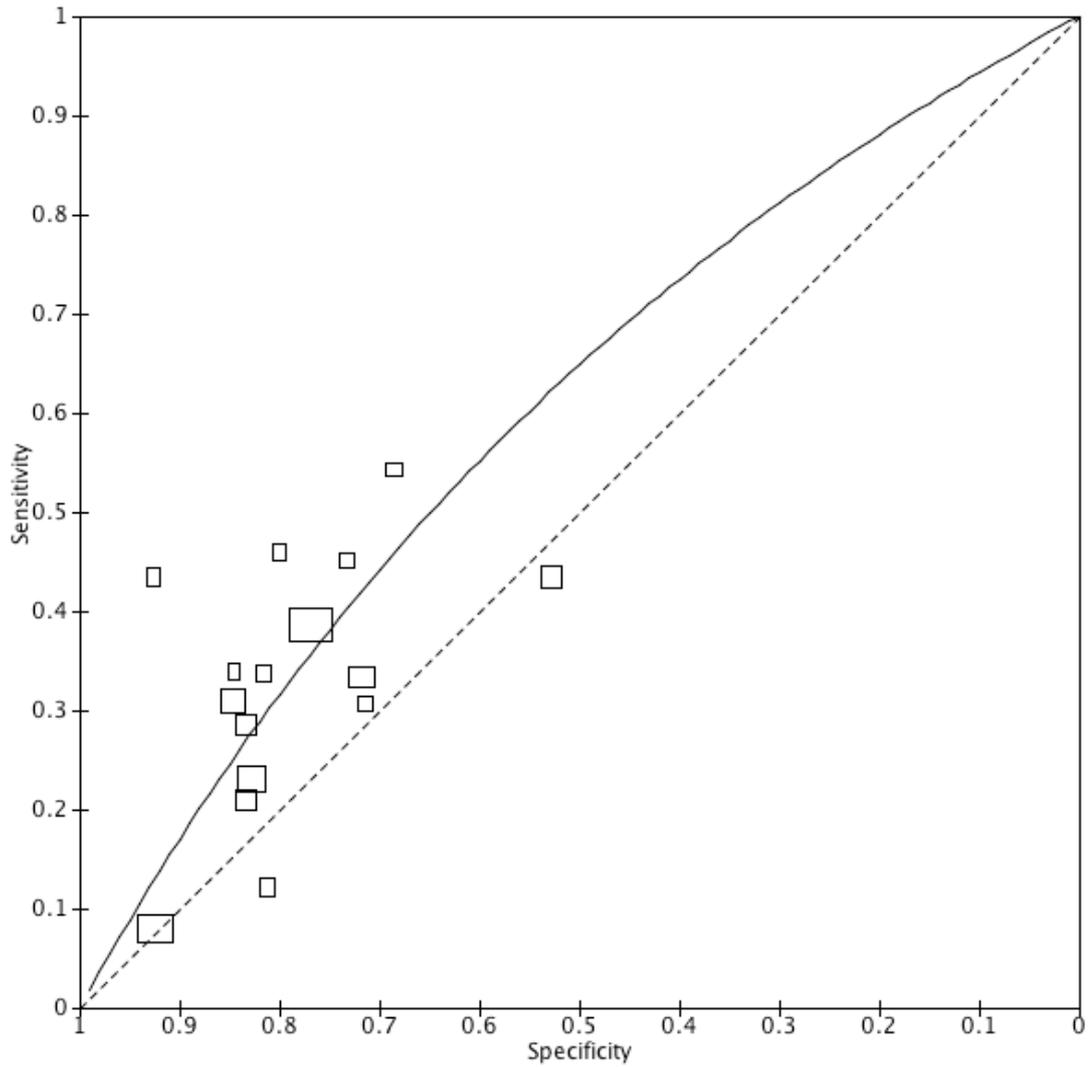


Figure 6-50 Summary ROC Curve of Presence of Diabetes

After collating the data into Meta-DiSc software the following results were obtained:

Heterogeneity chi-squared = 36.07 (d.f. = 14)  $p=0.001$

Inconsistency ( $I^2$ ) = 61.2%

Estimate of between-study variation (Tau-squared) = 0.1187

This suggests moderate heterogeneity and the studies appear widely distributed when their descriptive variables are plotted in ROC space.

Bivariate meta-analysis using MetaAnalyst software provided the following summary point:

Sensitivity = 0.310 (95% CI 0.244 – 0.382)

Specificity = 0.783 (95% CI 0.724 – 0.834)

This suggests that although some studies have a very high specificity, the overall value does not provide sufficient specificity to conclude that a patient definitely has heart failure. The positive likelihood ratio of 1.42 confirms this.

### 6.4.3 Patients with Atrial Fibrillation

Atrial fibrillation, where there is chaotic and uncoordinated depolarisation of the cardiac atria, can occur at any age but is more common as patients become older. There are a wide variety of causes but the commonest aetiology is as a result of ischaemic heart disease (Lip et al., 1995). If atrial fibrillation occurs in isolation, it may be well tolerated, but in the presence of a heart that is already compromised in some way, it can predispose the patient to the clinical manifestations of heart failure. Eight studies provided information about the presence or absence of atrial fibrillation in 1,996 patients. Although there is reasonable specificity associated with this finding the sensitivity is poor as shown in Figure 6-51.

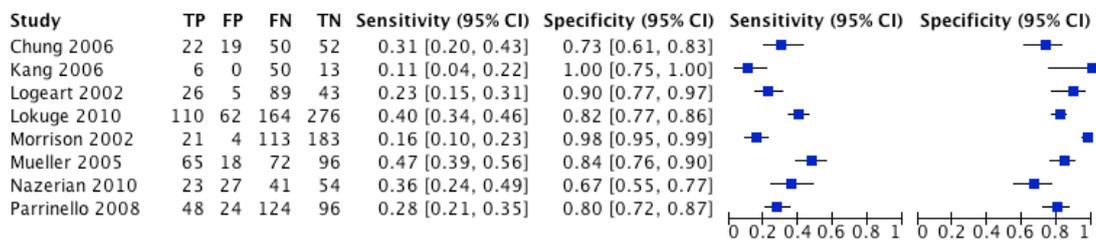


Figure 6-51 Forest Plot for Presence of Atrial Fibrillation

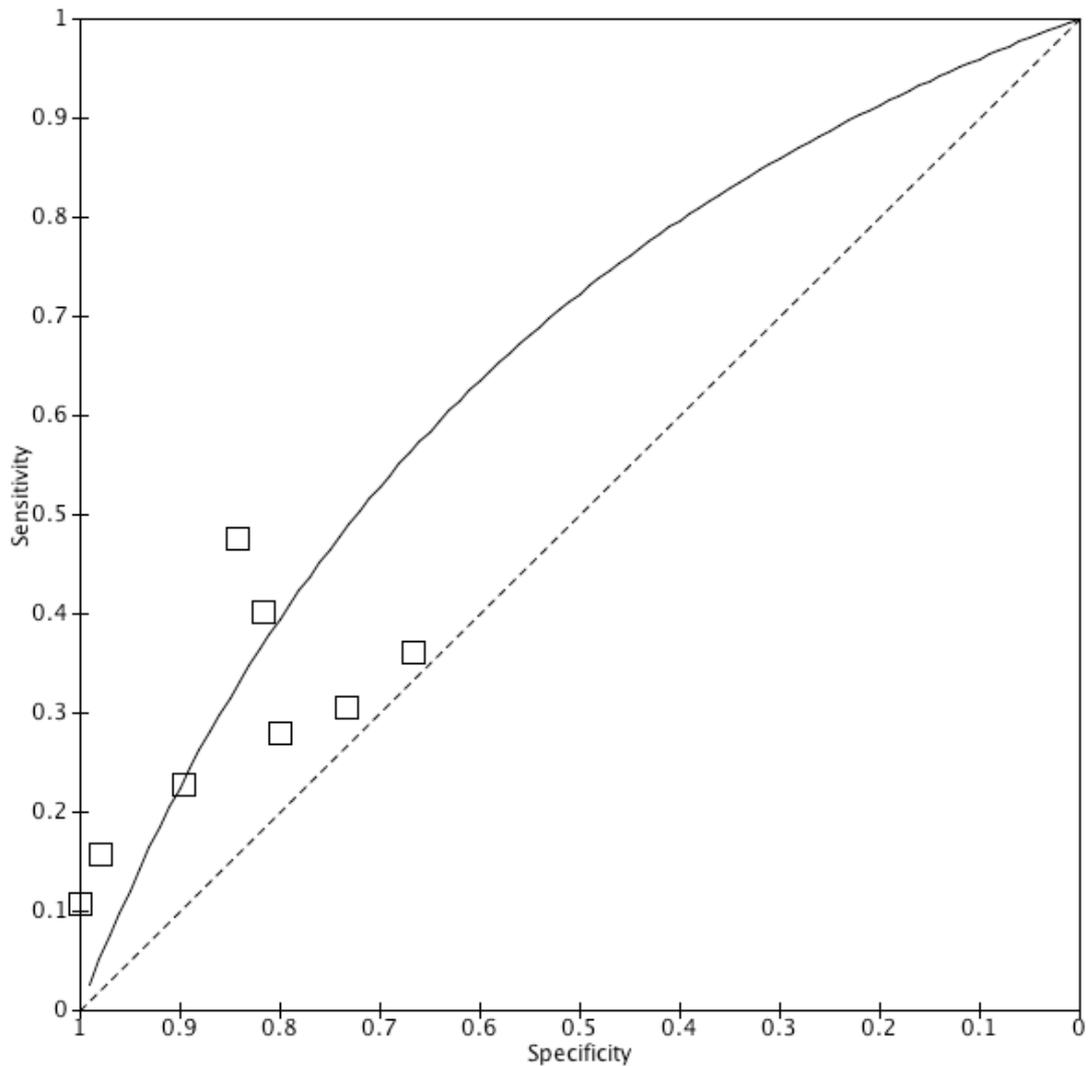


Figure 6-52 Summary ROC Curve for the Presence of Atrial Fibrillation

Analysis of the data using Meta-DiSc software provided the following results:

Heterogeneity chi-squared = 16.66 (d.f. = 6)  $p = 0.011$

Inconsistency ( $I^2$ ) = 64.0%

Estimate of between-study variation (Tau-squared) = 0.2197

Bivariate analysis using MetaAnalyst software provided the following summary point:

Sensitivity = 0.286 (95% CI 0.199 – 0.381)

Specificity = 0.871 (95% CI 0.750 – 0.955)

As with the other variables already discussed in this section, although this variable demonstrates reasonable specificity, suggesting a positive diagnosis of heart failure as the cause of the patient’s symptoms if it is present, the absence of atrial fibrillation does not rule out the possibility of heart failure. The positive likelihood ratio for this variable was 2.22.

#### 6.4.4 History of Renal Impairment

There is an association between renal disease and cardiac disease and these two conditions also share many of the same risk factors such as diabetes and hypertension. Ten studies have looked at the existence of renal dysfunction and presentation with acute dyspnoea in 5,873 patients. There is some variation in the definition of renal impairment. One study, Choi (2007), was from Korea and had a very low prevalence of renal impairment compared to the other studies, the definition of renal impairment in this paper was not provided. This study was the only one to demonstrate specificity greater than 95% but had a sensitivity of almost zero and therefore did not provide any useful diagnostic information.

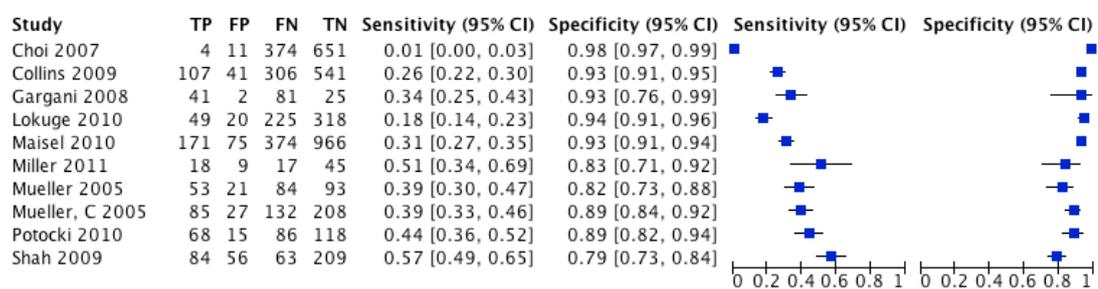


Figure 6-53 Forest Plot - Renal Impairment

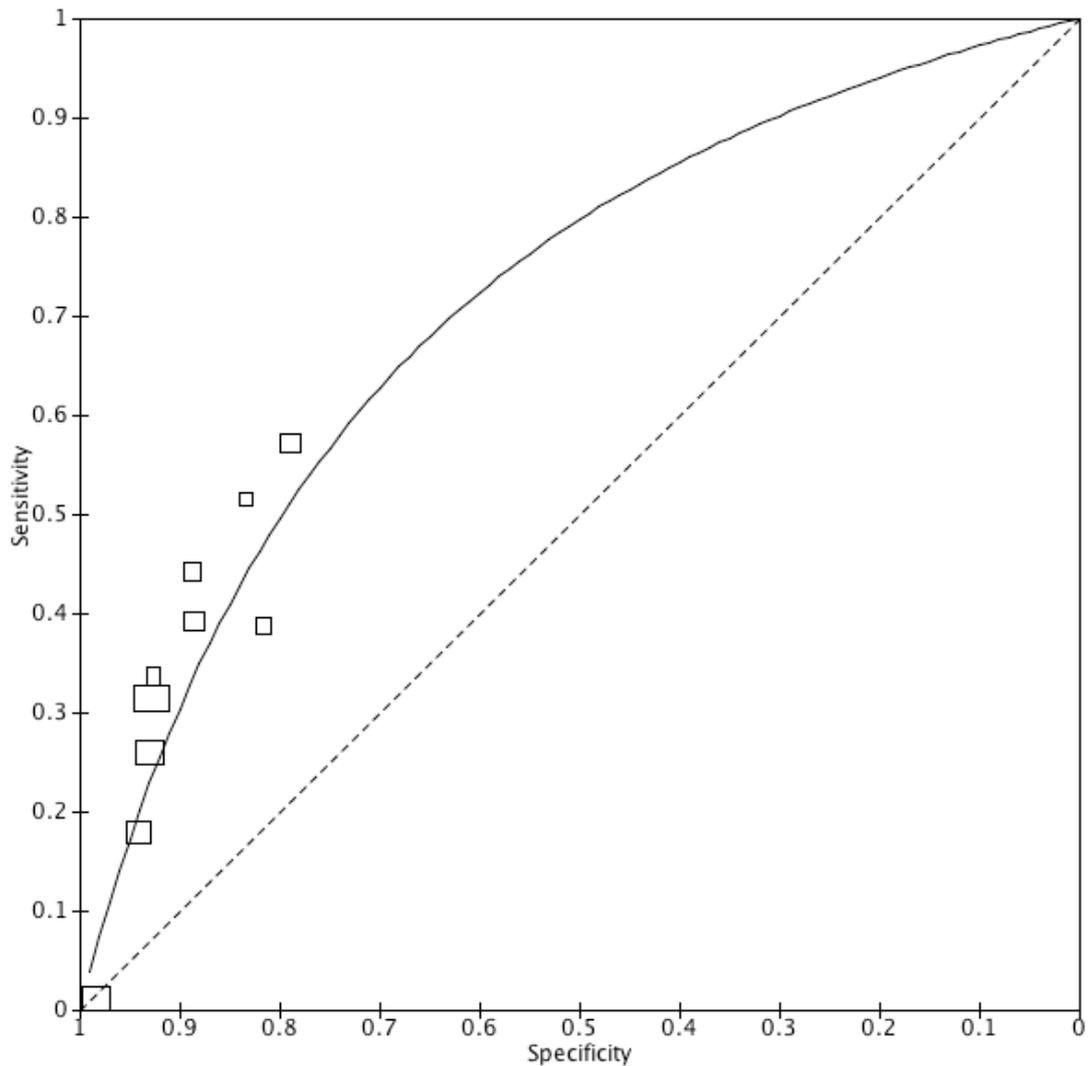


Figure 6-54 Summary ROC Curve - Renal Impairment

The data provided from Meta-DiSc was as follows:

Heterogeneity chi-squared = 18.38 (d.f. = 8) p=0.019

Inconsistency ( $I^2$ ) = 56.5%

Estimate of between-study variation (Tau-squared) = 0.0922

This suggests significant heterogeneity although most of the study summary points lie close to the estimated summary ROC curve. Bivariate meta-analysis provided the following results:

Sensitivity = 0.299 (95% CI 0.157 – 0.484)

Specificity = 0.907 (95% CI 0.849 – 0.947)

Even with this high specificity this variable does not provide conclusive evidence of heart failure as the sensitivity becomes very low as the specificity improves. The positive likelihood ratio calculated from the overall summary point is 3.22.

#### **6.4.5 Presence of a raised Jugular Venous Pressure**

The jugular venous pressure (JVP) can be estimated by examining the patient for the venous pulsation within the internal jugular vein. This vein is in direct continuation with the right atrium and so provides a measure of the pressure within it. The presence of a failing heart will usually lead to a raised jugular venous pressure though other conditions, for example superior vena cava obstruction can also cause this. The pressure can be measured invasively or can be approximated by lying the patient down at 45° and seeing if the jugular venous waveform can be visualised in the internal jugular vein. Although this test is useful when the waveform is visible, it is often difficult to discern in the acute unwell patient who is likely to be tachycardic, tachypnoeic and may not tolerate lying at this angle. Nineteen studies, with a total of 7,877 patients, examined the utility of this clinical finding. As can be seen in the forest plot in **Error! Reference source not found.**, the test has good specificity but poor sensitivity as a marker for acute decompensated heart failure.

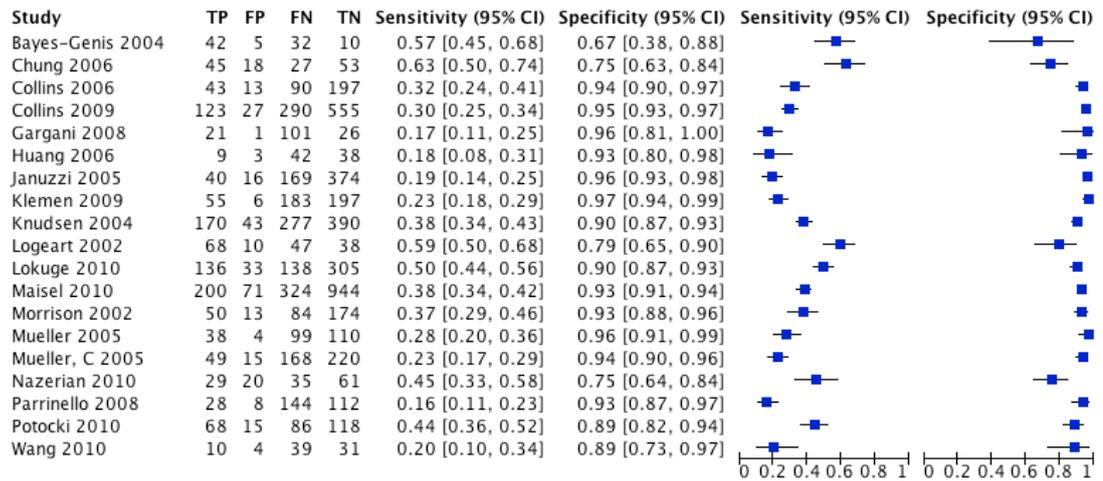


Figure 6-55 Forest plot for presence of raised JVP

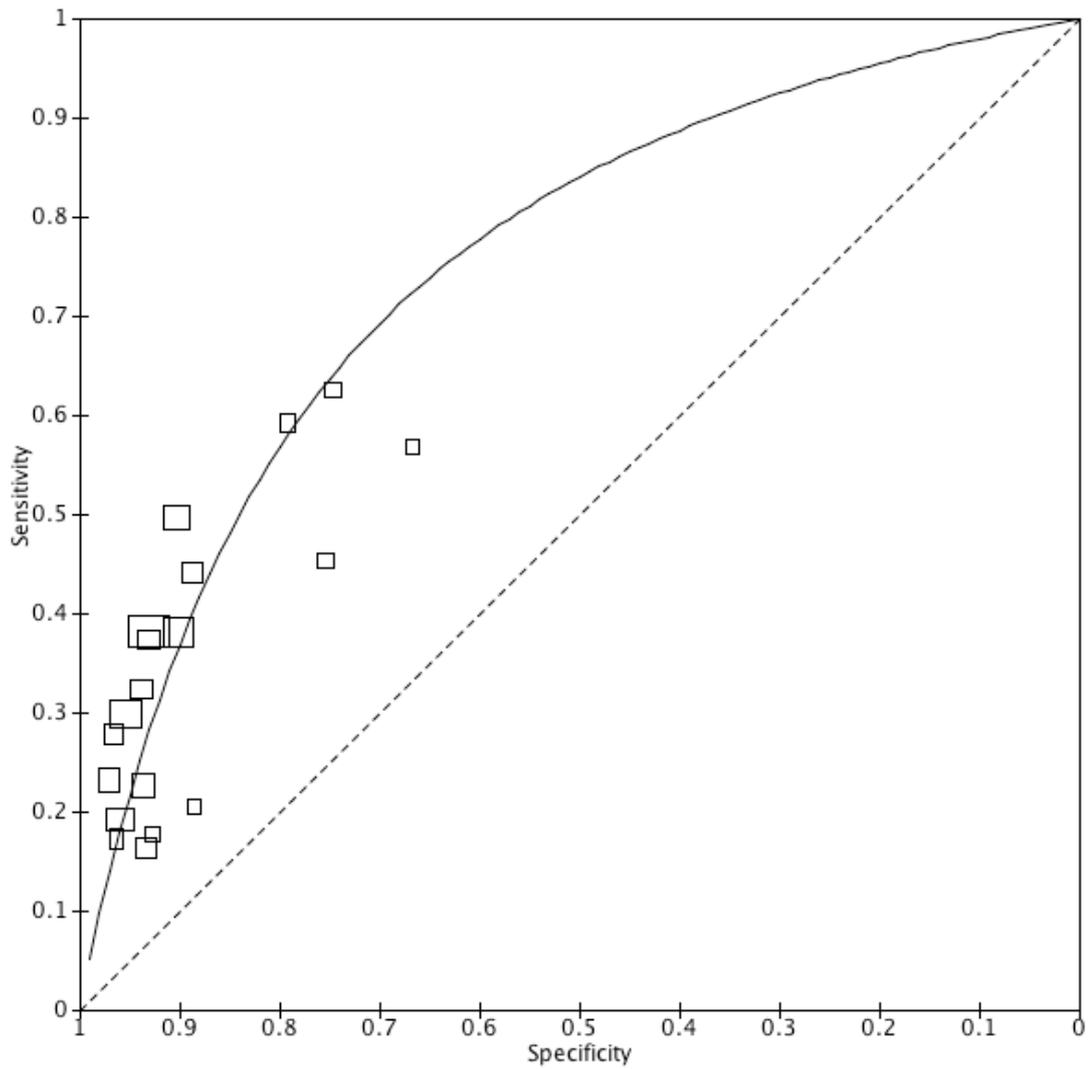


Figure 6-56 Summary ROC Curve for presence of raised JVP

Analysis of the data using Meta-DiSc produced the following results:

Heterogeneity chi-squared = 27.62 (d.f. = 17)  $p = 0.056$

Inconsistency ( $I^2$ ) = 37.4%

Estimate of between-study variation (Tau-squared) = 0.0563

This suggests a low level of heterogeneity.

Bivariate analysis provided the following results:

Sensitivity = 0.341 (95% CI 0.276 – 0.412)

Specificity = 0.920 (95% CI 0.889 – 0.945)

This corresponds to a positive likelihood ratio of 4.26 so the presence of clinically high JVP is very suggestive of heart failure, though as the specificity is below 95% it is not advisable to use this result in isolation.

#### **6.4.6 Presence of a Cardiac Murmur**

Heart murmurs are audible due to the turbulent flow of blood through the heart. Although some heart murmurs are benign and can be a normal finding in young patients, the presence of other murmurs suggests valvular disease, the presence of abnormal foramen within the heart or abnormal flow rate over a normal valve. Detecting heart murmurs can be difficult in a noisy environment, especially with an acutely unwell and tachypnoeic patient. Auscultation of the heart for the presence of murmurs can be enhanced by special manoeuvres or by breath holding; neither of these may be possible in the acutely decompensated patient. Six studies were found which looked at the presence or absence of audible heart murmurs in a total of 3,327 patients. The results of these studies are shown in Figure 6-57 and Figure 6-58 with the bivariate meta-analysis results shown in Figure 6-59.

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Chung 2006	32	9	40	62	0.44 [0.33, 0.57]	0.87 [0.77, 0.94]		
Januzzi 2005	40	27	169	363	0.19 [0.14, 0.25]	0.93 [0.90, 0.95]		
Klemen 2009	69	10	169	193	0.29 [0.23, 0.35]	0.95 [0.91, 0.98]		
Maisel 2010	156	98	395	956	0.28 [0.25, 0.32]	0.91 [0.89, 0.92]		
Morrison 2002	39	21	95	166	0.29 [0.22, 0.38]	0.89 [0.83, 0.93]		
Wang 2010	29	8	100	81	0.22 [0.16, 0.31]	0.91 [0.83, 0.96]		

Figure 6-57 Forest Plot for Presence of Heart Murmur

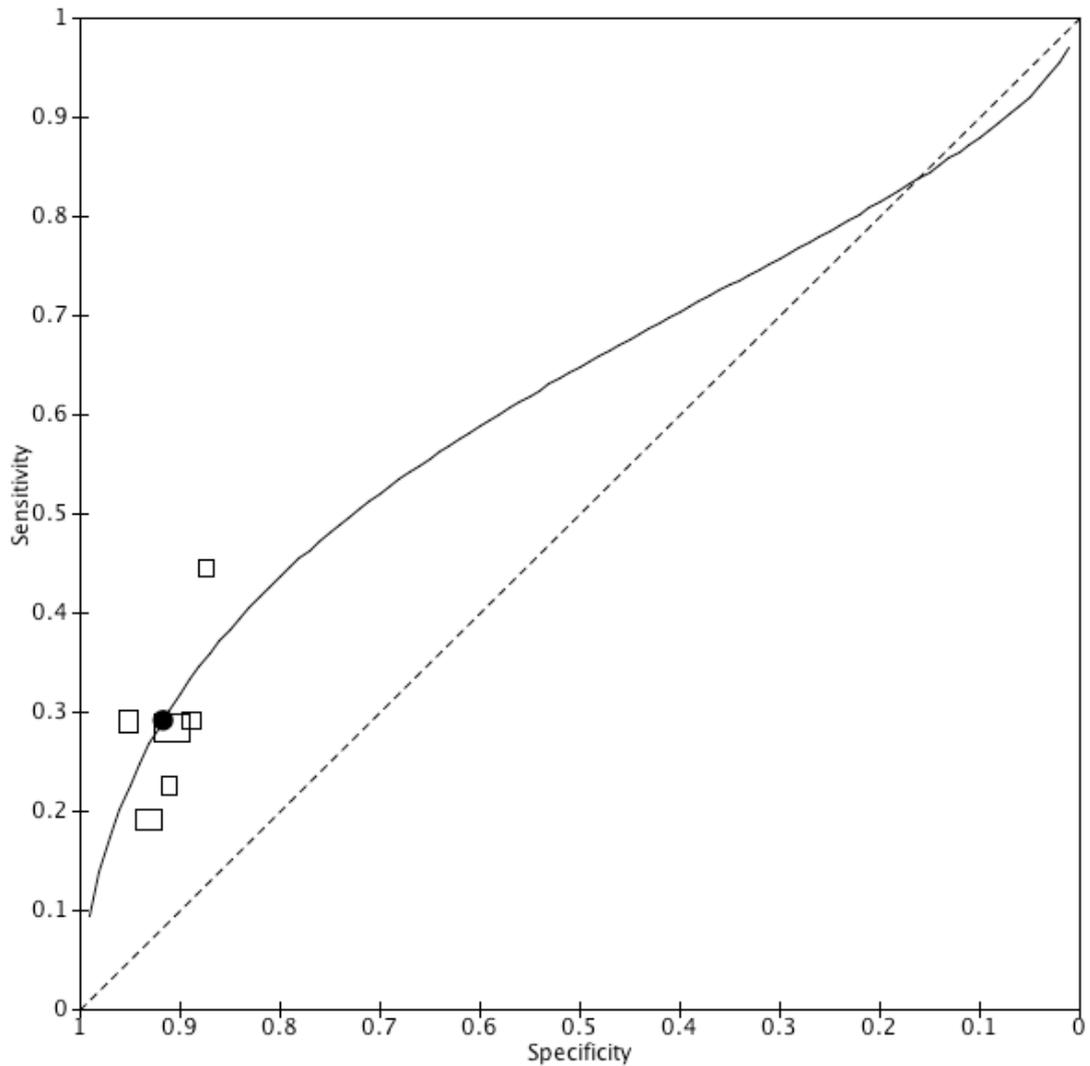


Figure 6-58 Summary ROC Curve for the Presence of a Heart Murmur

Analysis of the data using Meta-DiSc provided the following results:

Heterogeneity chi-squared = 5.58 (d.f. 4) p = 0.233

Inconsistency ( $I^2$ ) = 28.3%

Estimate of between-study variance (Tau-squared) = 0.0283

This suggests moderate heterogeneity. Bivariate meta-analysis presented the following summary point:

Sensitivity = 0.279 (95% CI 0.219 – 0.351)

Specificity = 0.912 (95% CI 0.870 – 0.943)

This provides a positive likelihood ratio of 3.2 and, although the presence of a murmur is strongly suggestive of a diagnosis of heart failure, it is not sufficiently specific to rule confirm this condition as an isolated finding.

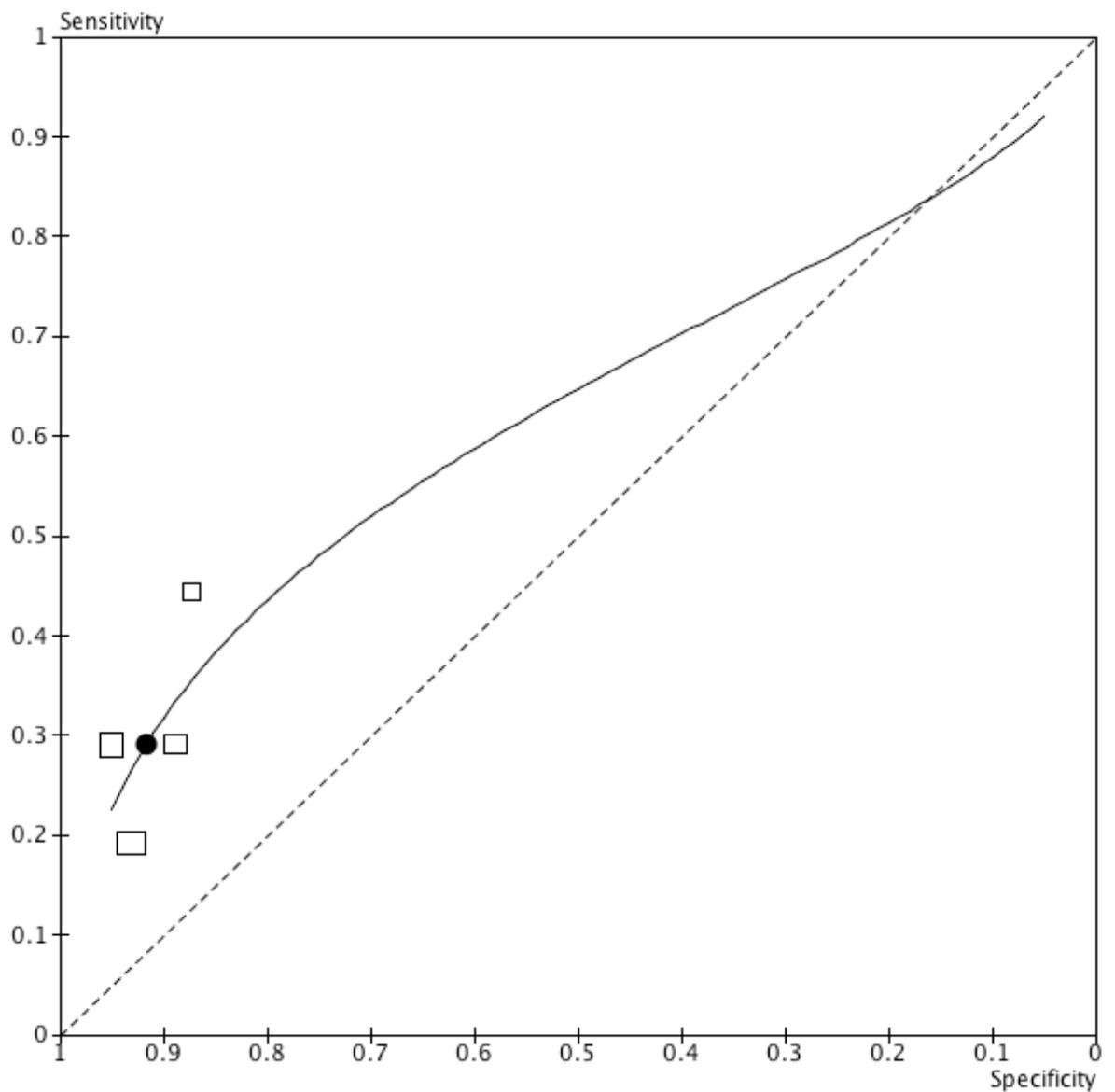


Figure 6-59 Summary ROC Curve Heart Murmur with Bivariate Meta-analysis Data

### 6.4.7 Troponin Levels

Troponin is a very specific cardiac marker, forming part of the contractile mechanism in the cardiac myocytes that is released into the circulation following cell death. Levels tend to be higher in patients with some degree of cardiac dysfunction and this may provide prognostic information (Felker et al., 2012). A prospective study has shown that troponin levels are raised in

around 50% of patients who present with acute decompensated heart failure and a further 20% develop a troponin rise during their admission (Gheorghiade et al., 2005). Three studies looked at this biomarker in 1,258 patients. As can be seen with reference to Figure 6-60 and Figure 6-61, this test is very specific but has poor sensitivity for the diagnosis of heart failure.

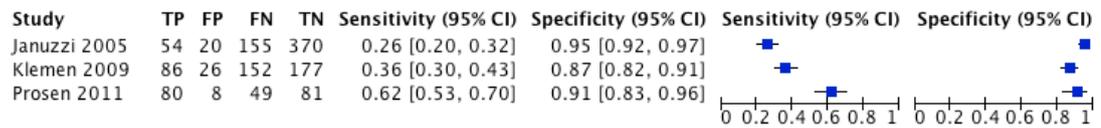


Figure 6-60 Forest Plot for Troponin Rise

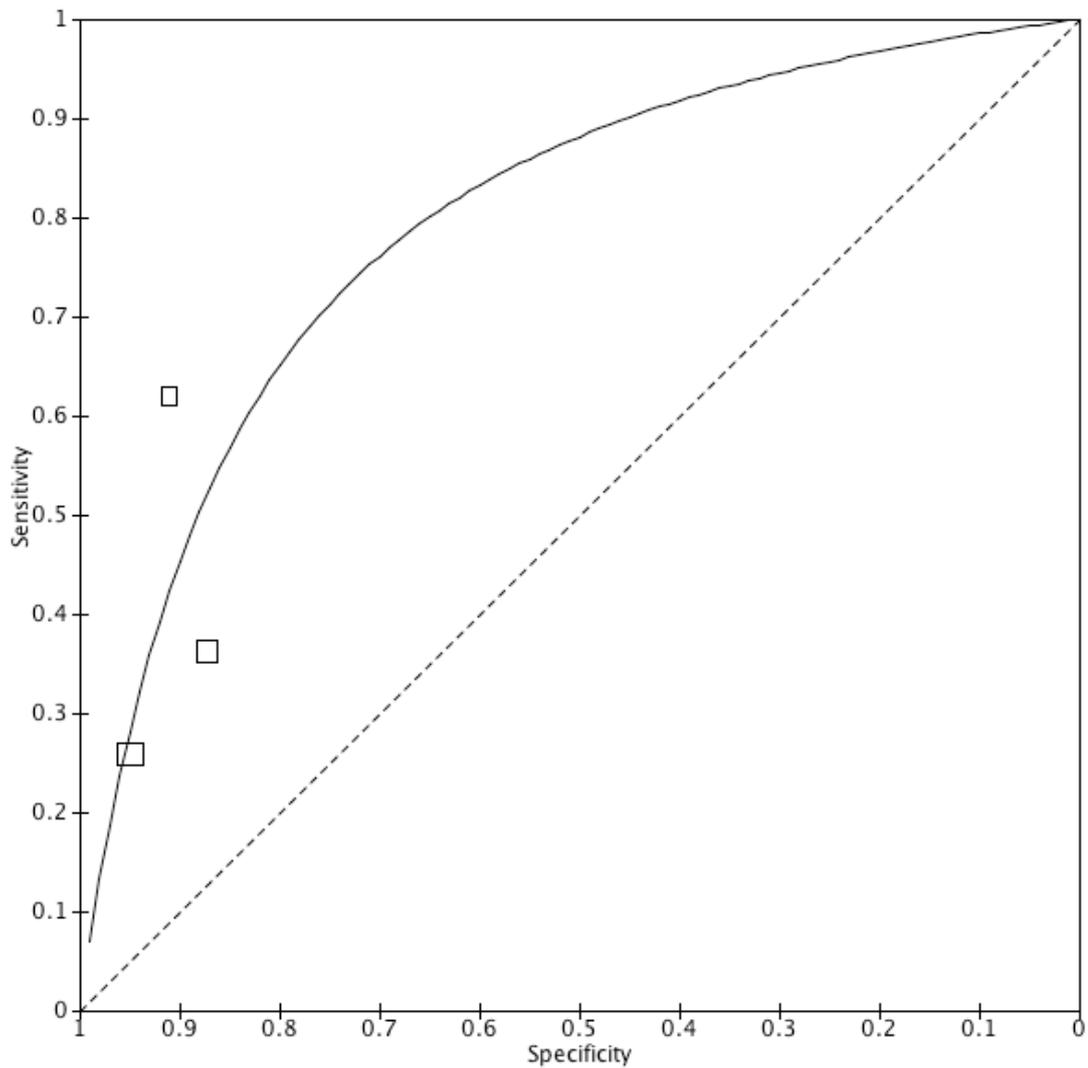


Figure 6-61 Summary ROC Curve for Troponin Rise

Pooling the available data using MetaAnalyst software provided the following values:

Sensitivity = 0.408 (95% CI 0.234 - 0.629)

Specificity = 0.908 (95% CI 0.802 - 0.970)

Again, this finding is strongly suggestive of heart failure with a positive likelihood ratio of 4.4, but the presence of a high troponin level does not confirm the diagnosis by itself. Other advantages of this test include its

ubiquity in the UK Emergency Department setting, the objectivity of the test and the fact that it is rapidly available.

### 6.4.8 Hepato-jugular Reflux

The hepato-jugular reflux is elicited by pressing in the right upper quadrant of the abdomen and observation to record any rise in jugular venous pressure. This was first described by Rondot (1898) as a diagnostic test for heart failure but subsequent studies have not confirmed its diagnostic value (Singh & Haider, 1973). Two studies with a total of 500 patients provided data about this clinical test. The results are provided in Figures 6-62 and 6-63.

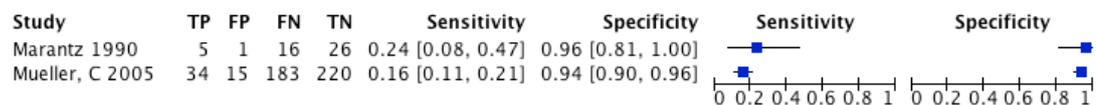


Figure 6-62 Forest Plot for Hepato-jugular Reflux

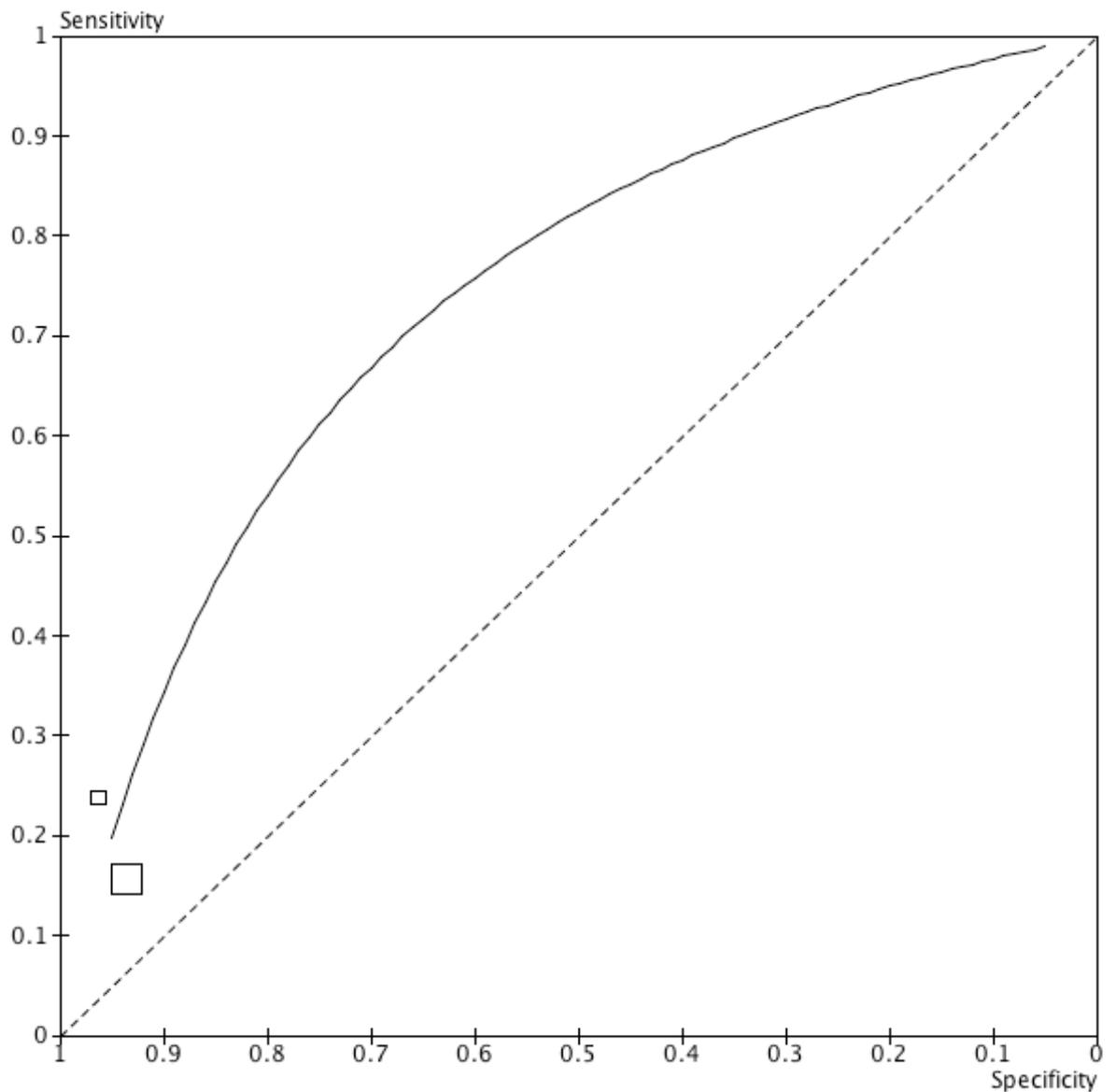


Figure 6-63 Summary ROC Curve for Hepato-jugular Reflux

Pooled results suggest a very good degree of specificity in the small proportion of patients who had the positive finding with poor sensitivity.

Results pooled using Meta-DiSc as shown:

Sensitivity = 0.164 (95% CI 0.119 - 0.217)

Specificity = 0.939 (95% CI 0.903 - 0.965)

Positive Likelihood Ratio = 2.69

### 6.4.9 Presence of a Third Heart Sound

The presence of a third heart sound occurs in early diastole and is thought to be due to abnormal ventricular filling. It is usually a pathological finding though it can be normal in young adults and during pregnancy. The finding is suggestive of heart failure. Twelve studies included the presence of this factor and the resulting data is derived from 6,072 patients. The resulting forest plot and summary ROC curve demonstrate that this test is extremely specific but has low sensitivity for the diagnosis of heart failure.

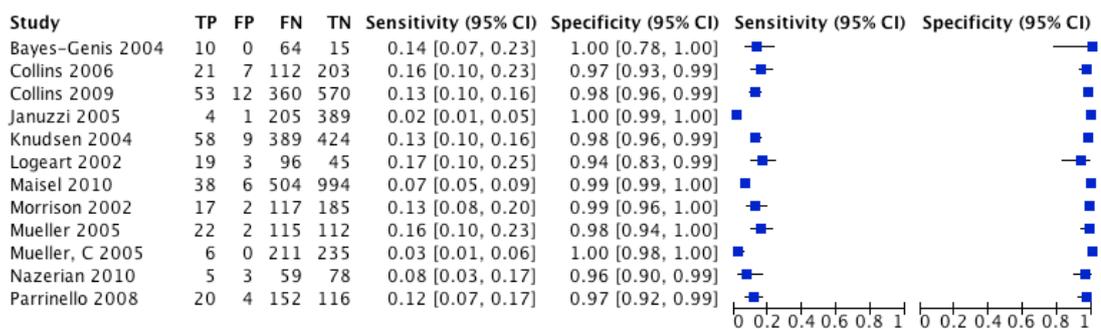


Figure 6-64 Forest Plot - Presence of Third Heart Sound

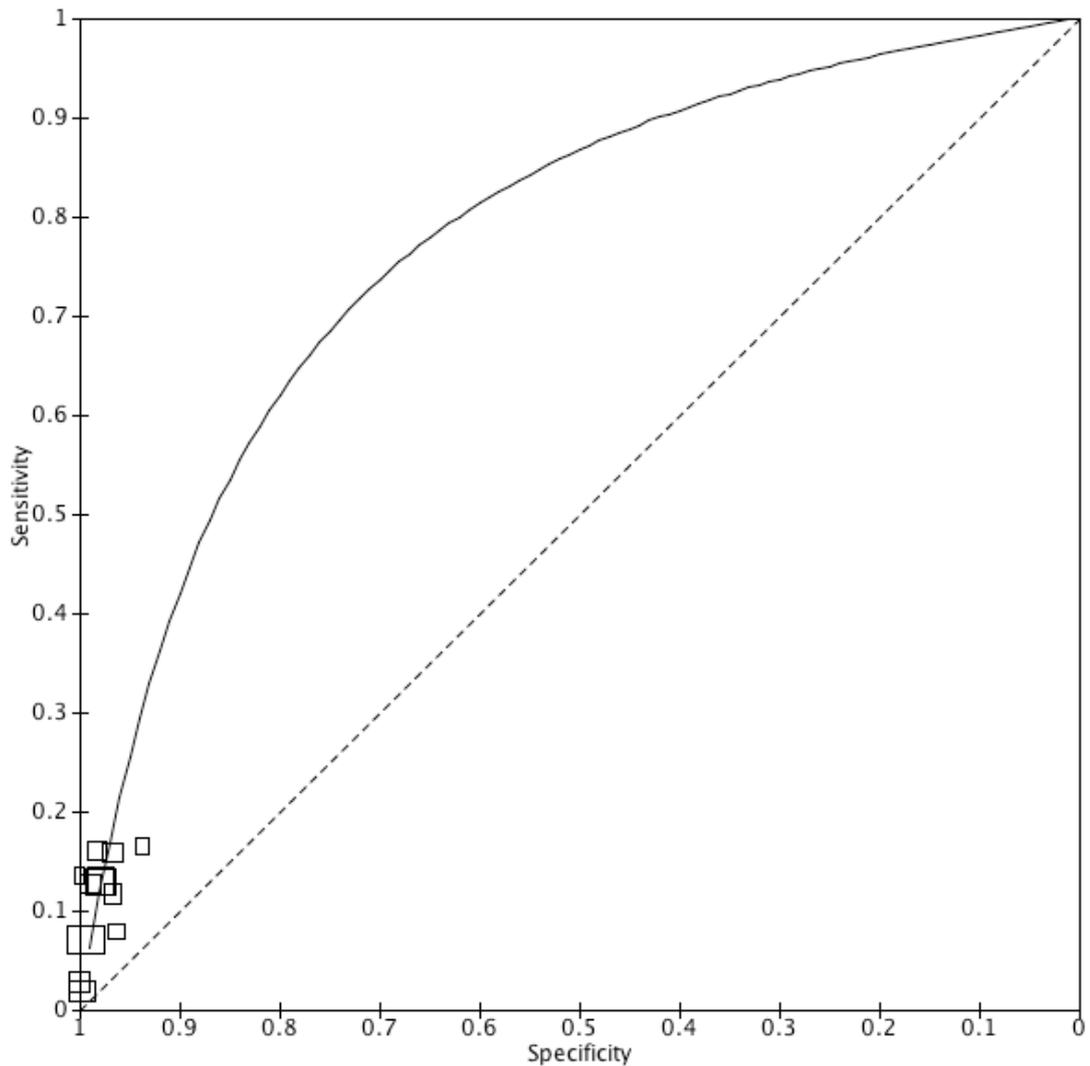


Figure 6-65 Summary ROC Curve - Presence of Third Heart Sound

In contrast with many of the other variables examined, the amount of statistical heterogeneity was minimal as is reflected in the following results:

Heterogeneity chi-squared = 4.34 (d.f. 9)  $p = 0.888$

Inconsistency ( $I^2$ ) = 0.0%

Estimate of between-study variance (Tau-squared) = 0.000

Bivariate meta-analysis provided the following summary point:

Sensitivity = 0.100 (95% CI 0.070 – 0.138)

Specificity = 0.985 (95% CI 0.973 – 0.993)

Although the specificity of this test is excellent, the sensitivity is very poor, in keeping with the very low incidence of this clinical sign and so the diagnostic value is limited despite a likelihood ratio of 6.66.

An example is provided to put this result in context: if this test were applied to a population of 1,000 patients with a prevalence of heart failure of 40% then the following results would be obtained.

Table 6-7 - Example Data for Third Heart Sound - 1

	Heart Failure	Not Heart Failure	Totals
3 <sup>rd</sup> Heart Sound	40	9	49
No 3 <sup>rd</sup> Heart Sound	360	591	951
Totals	400	600	1000

Therefore nine out of the forty-nine patients would not have acute heart failure despite having an audible third heart sound. However, if the patients are subjected to a diagnostic process that increases the prevalence or pre-test probability of heart failure to 70% then a positive finding becomes much more useful for ruling the condition in as shown below.

Table 6-8 Example Data for Third Heart Sound - 2

	Heart Failure	Not Heart Failure	Totals
3 <sup>rd</sup> Heart Sound	70	5	75
No 3 <sup>rd</sup> Heart Sound	630	295	925
	700	300	1000

### 6.4.10 Presence of a Fourth Heart Sound

The fourth heart sound is thought to be due to the increased non-compliance of the ventricle walls in the presence of heart failure. On contraction of the atria, the extra blood causes a sudden increase in pressure rather than volume, and this usually signifies abnormal cardiac function. Four studies derived data about this finding from 2,258 patients. As with the presence of a third heart sound, this clinical finding was highly specific for heart failure but demonstrated poor sensitivity as shown in Figure 6-66 and Figure 6-67.

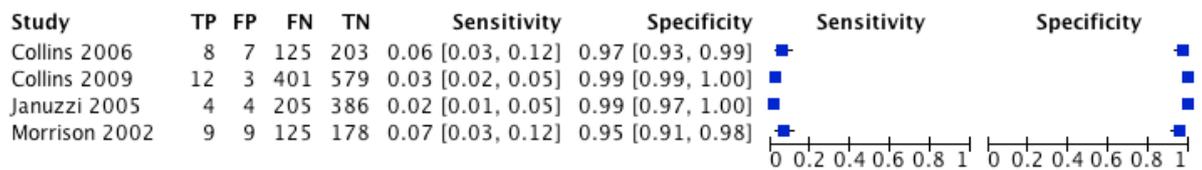


Figure 6-66 Forest Plot for Presence of Fourth Heart Sound

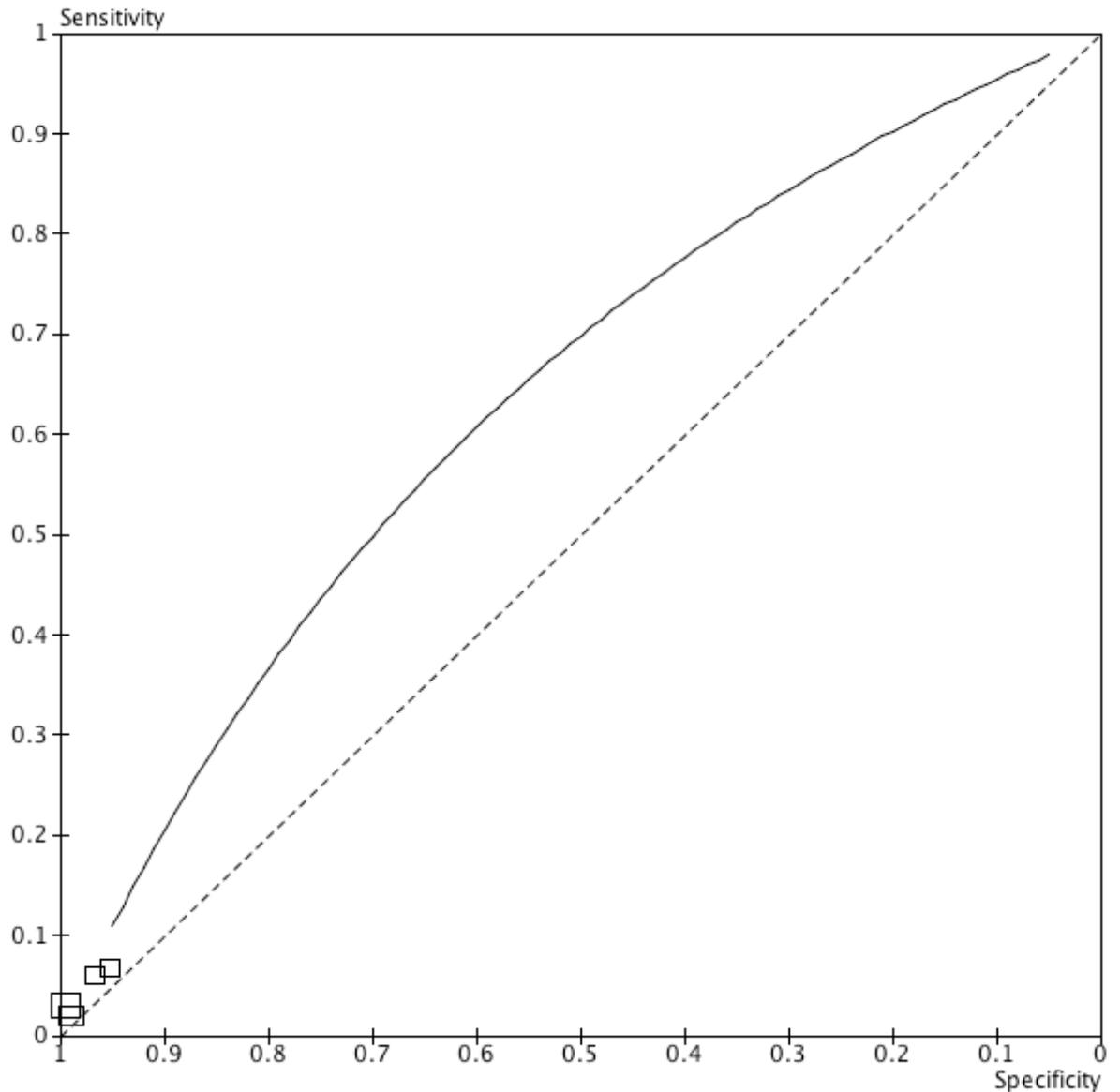


Figure 6-67 Summary ROC Curve for the Presence of a Fourth Heart Sound

Using Meta-DiSc software to analyse the available data the following results were obtained:

Heterogeneity chi-squared = 3.18 (d.f. 3)  $p = 0.364$

Inconsistency ( $I^2$ ) = 5.8 %

Estimate of between-study variance (Tau-squared) = 0.0207

Pooling of the data provided the following summary point:

Sensitivity = 0.037 (95% CI 0.026 – 0.052)

Specificity = 0.983 (95% CI 0.975 – 0.989)

As with the presence of third heart sound there is little statistical heterogeneity between the reported studies but an excellent specificity is linked to a very poor specificity which limits the diagnostic utility of this finding and this is reflected in a positive likelihood ratio of 2.18.

## 6.5 Variables with Good Diagnostic Value

### 6.5.1 History of Heart Failure

Given that heart failure rarely resolves entirely, that acute exacerbations are not associated with a 100% mortality rate and that it tends to follow a course of remission and exacerbation, it is not surprising that having a history of heart failure in the past, predicts further episodes in the future. The utility of this variable as a diagnostic tool is limited in the patient presenting with dyspnoea for the first time, but in patients with numerous admissions for heart failure in the past, the pre-test probability of heart failure would be expected to be substantially higher than that of the general population. Twenty-three studies examined this historical variable in 9,080 patients. The findings are displayed in Figure 6-68. The studies suggest that the history of heart failure is a reasonably specific finding for diagnosing heart failure although there is greater variation in the reported sensitivity.

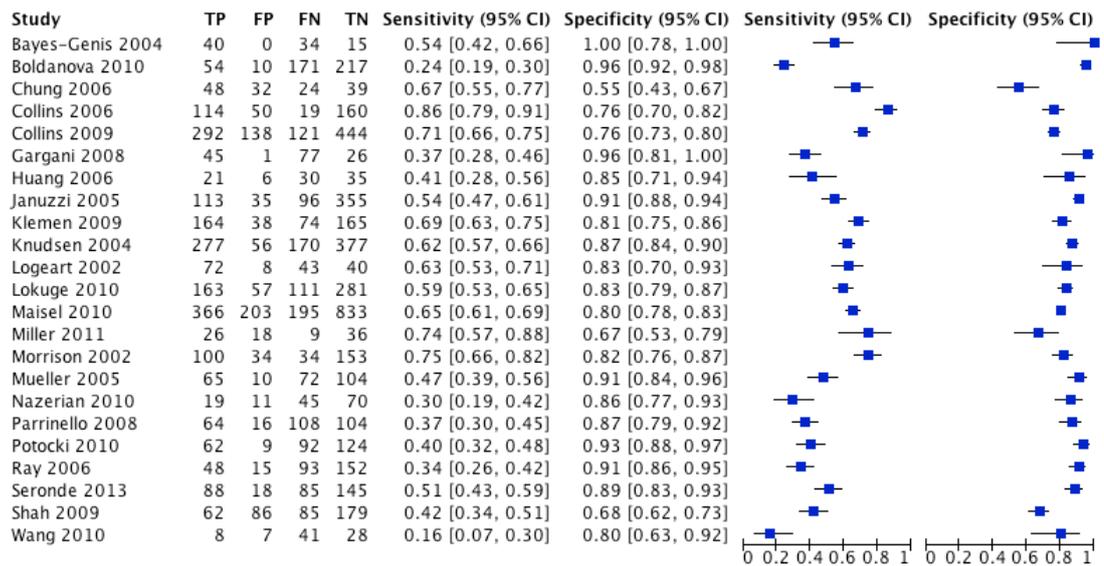


Figure 6-68 Sensitivity and Specificity of a History of Heart Failure

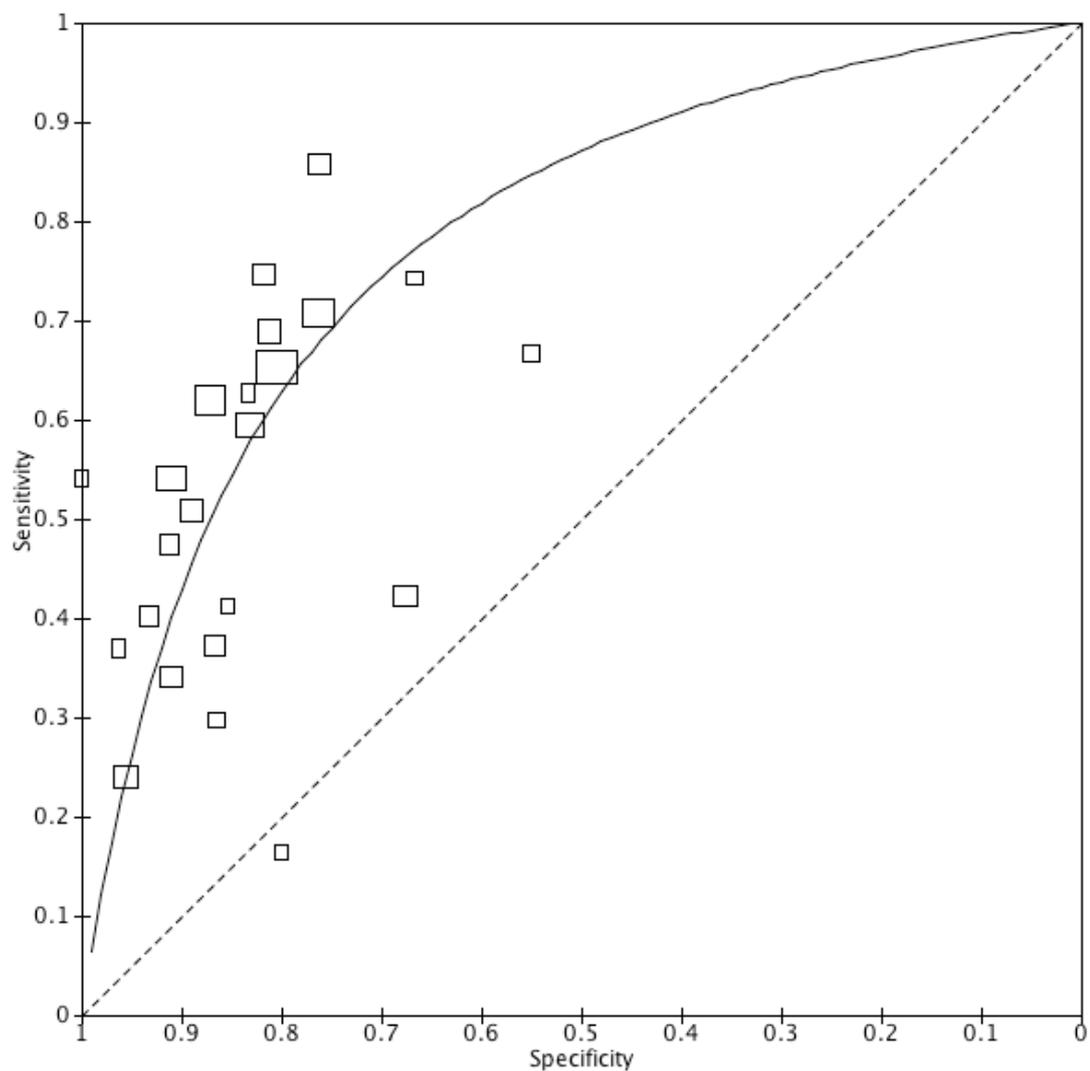


Figure 6-69 Summary ROC Curve - History of Heart Failure

Analysis of heterogeneity using Meta-DiSc software provided the following information:

Heterogeneity chi-squared = 122.50 (d.f. = 22) p = 0.000

Inconsistency = 82.0%

Estimate of between-study variance (Tau-squared) = 0.3163

(Added  $\frac{1}{2}$  to any cells with zero).

MetaAnalyst provided the following bivariate meta-analysis results.

Mean specificity = 0.864 (95% CI 0.819 – 0.903)

Mean sensitivity = 0.514 (95% CI 0.431 – 0.600)

The overall result suggests reasonable specificity although the results of the sensitivity are more variable and limit the overall diagnostic value. Marked statistical heterogeneity is present although the distribution in ROC space suggests a threshold effect.

Two studies are outliers on examining the summary ROC curve Shah (2009) and Chung (2006). The reason for this discrepancy is not obvious. In the Shah (2009) study 32.4% of the group of patients who were judged not to have acute decompensated heart failure had a history of heart failure. Excluding both of these studies from the analysis did not improve the heterogeneity scores and did not markedly improve the diagnostic utility:

Heterogeneity chi-squared = 54.02 (d.f. = 20) p=0.000

Inconsistency = 63.0%

Estimate of between-study variance (Tau-squared) = 0.1190

(Added  $\frac{1}{2}$  to any cells with zero).

Overall estimated data by Meta-Analyst for data with two studies excluded:

Specificity = 0.880 (95% CI 0.845 – 0.911)

Sensitivity = 0.512 (95% CI 0.420 – 0.610)

As may have been expected, the test is specific as patients who have been diagnosed with heart failure before are more likely to present this condition, than patients who have never been diagnosed with this malady. The low sensitivity may be due to a significant number of patients presenting with decompensated heart failure as an initial presentation. There may be some bias in applying the reference standard to this group, as patients with a known history of heart failure may be more likely to receive this as an end diagnosis.

### **6.5.2 Clinical Opinion**

Information regarding 'clinical opinion' was available from nine studies. In order to be included the clinical opinion had to be the opinion of the treating physician, recorded while the patient was still in the emergency department. The opinion needed to be based on the historical, clinical and investigative findings that would usually be available at that point in time. Data from eleven studies, where the clinician has indicated that it was highly likely or more than 80% likely that the patient had heart failure, has been included. A total of 5,357 patients were involved in these studies. One study, Lo (2007), is an obviously outlier on the summary ROC curve as shown in Figure 6-71, with values far different from the other studies. The other studies follow a reasonably smooth curve that would be consistent with a 'threshold' effect.

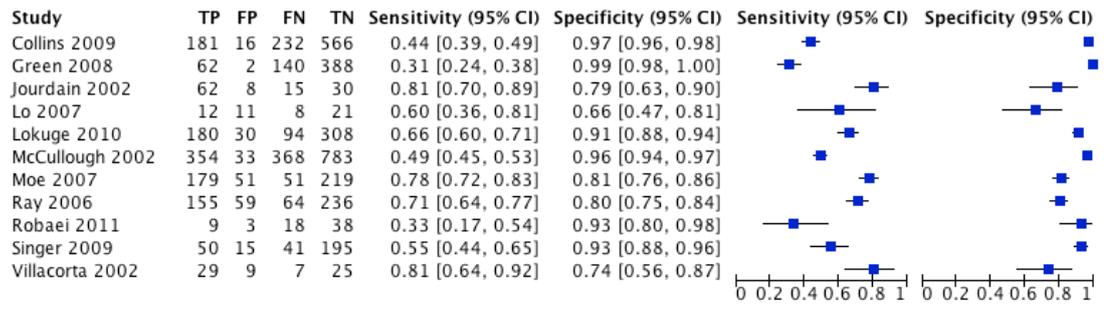


Figure 6-70 Forest Plot of Clinical Opinion

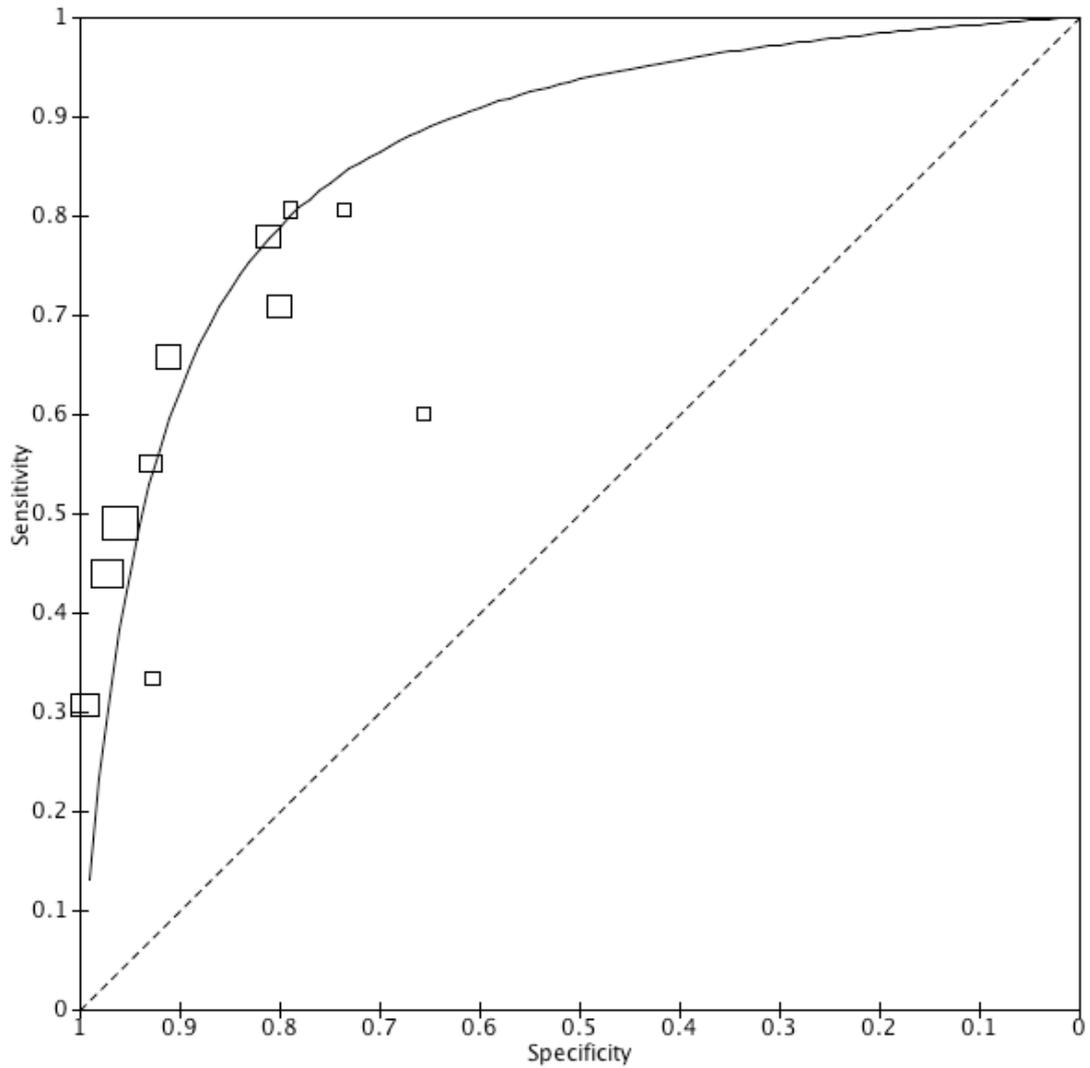


Figure 6-71 Summary ROC Curve - Clinical Opinion

Although the individual components that make up clinical opinion may have been shown to have limited diagnostic value, clinical opinion overall is reasonably accurate. There is considerable variation, and as may be expected with such a subjective variable, the pattern of distribution of the results in ROC space is strongly suggestive of a threshold effect.

Entering the data and an analysis using Meta-DiSc software provided an inconsistency score ( $I^2$ ) of 66.9%, suggesting a high level of heterogeneity. It can be seen on the summary ROC curve that one study, Lo (2007), is an obvious outlier. The study involved a low number of patients selected as a convenience sample with wide confidence intervals and lies far from the summary curve compared with the other studies.

Excluding this study improves the  $I^2$  score to 57.3%, which remains reasonably high, although this score does not account for heterogeneity due to the positivity threshold effect. A revised summary ROC curve, excluding Lo 2007, demonstrates that the lie of the plotted studies is similar the SROC curve suggesting a threshold effect.

Bivariate meta-analysis using MetaAnalyst software and excluding Lo (2007) provides the following summary point:

Sensitivity = 0.603 (95% CI 0.483 – 0.713)

Specificity = 0.919 (95% CI 0.852 - 0.965)

It may be difficult to apply this result directly to clinical practice due to the subjectivity involved, but examination of the data is useful. It is reassuring that, overall, clinical opinion is a reasonably good test of whether or not a person has heart failure. It has to be borne in mind that this variable closely resembles the reference test that has been applied for these studies and will be based on many of the same findings.

In addition to the inter-observer and intra-observer variation that is likely to exist with such a subjective criteria, there is also the consideration of the

spectrum of the disease; there may be greater discrepancy in the diagnoses for patients with mild or moderate symptoms compared with patients with severe heart failure.

### 6.5.3 Presence of Cardiomegaly on the Chest X-ray

The presence of cardiomegaly can be suggested by a widening of the cardiothoracic ratio on a posterior-anterior chest film. The normal width of the cardiac shadow is generally defined as less than 50% of the internal thoracic wall when measured at the widest point. An increase in the size of the cardiac shadow on x-ray is suggestive of heart failure in the symptomatic patient. The cardiac shadow can be measured either on a physical film or on computer image and a ratio between the cardiac shadow and the thoracic diameter can be calculated. The estimate of cardiac size is much less reliable on anterior-posterior (AP) films as the heart is further away from the radiation source and so projects a larger shadow. Unfortunately, as patients with acute dyspnoea are often too unwell to be transferred to the radiology department, they tend to have a portable radiograph that can only be performed as an AP film. Ten studies have looked at this investigative finding in a total of 3,194 patients. The summaries of the findings from these studies are presented in Figure 6-72 and Figure 6-73.

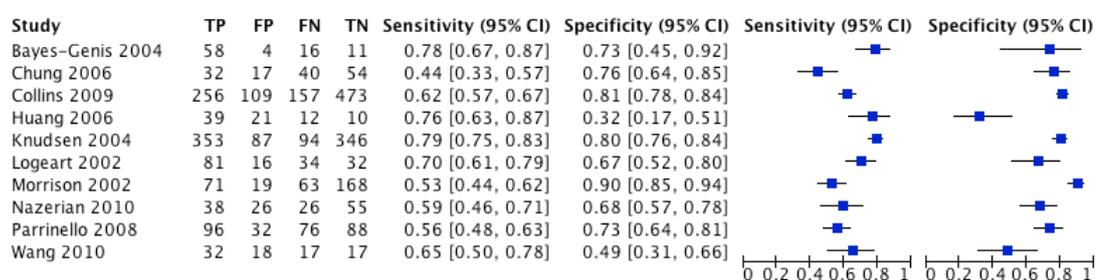


Figure 6-72 Sensitivity and Specificity Presence of Cardiomegaly

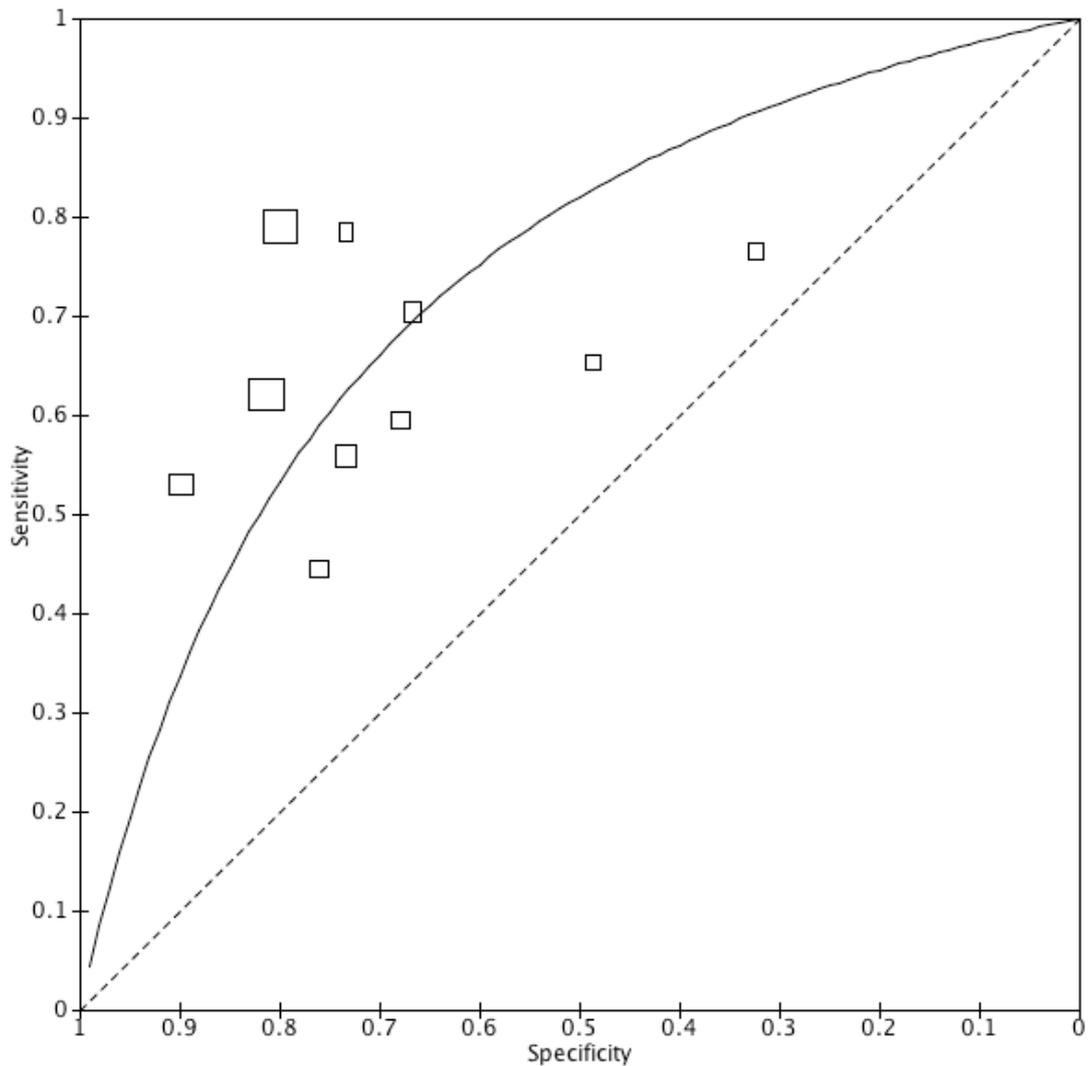


Figure 6-73 Summary ROC Curve - Presence of Cardiomegaly

The data extracted from the studies was analysed using Meta-DiSc software. The following results were obtained suggesting marked heterogeneity.

Heterogeneity chi-squared = 52.89 (d.f. = 8)  $p = 0.000$

Inconsistency ( $I^2$ ) = 84.9%

Estimate of between-study variance (Tau-squared) = 0.4114

The study Huang (2006) is a notable outlier on the summary ROC curve. The study involved small numbers of patients and has wide confidence intervals. The study was set in Taiwan and excluded over ten per cent of potentially

eligible patients due to poor echocardiography windows. Exclusion of this study allowed calculation of the following results:

Heterogeneity chi-squared = 43.85 (d.f. = 7) p=0.000

Inconsistency ( $I^2$ ) = 84.0%

Estimate of between-study variance (Tau-squared) = 0.3527

This still suggests significant heterogeneity between the studies although the plotted results do not appear to stray too far from the summary ROC curve on visual inspection. The next obvious outlier is Chung (2006). The inclusion criteria for this included patients with severe dyspnoea and whom the attending physician felt needed admission to hospital. This suggests that this study population may have been a subgroup of the patients who attended hospital with acute dyspnoea, and different from the other study populations in this respect. However, elimination of this study did little to improve the estimated heterogeneity.

The best diagnostic result was obtained in the Knudsen (2004) study. This is a subgroup analysis from a larger study, Maisel (2002), in which patients who had their chest x-ray reported by a radiologist while in the emergency department. It is possible that selection bias occurred in this population as the process by which this occurred it is not clear from the study description.

Heterogeneity chi-squared = 27.75 (d.f. = 5) p = 0.000

Inconsistency ( $I^2$ ) = 82.0%

Estimate of between-study variance (Tau-squared) = 0.2650

Bivariate meta-analysis was performed using MetaAnalyst software on the available data excluding Huang 2006 and provided the overall summary point:

Sensitivity = 0.640 (95% CI 0.554 – 0.718)

Specificity = 0.75 (95% CI 0.652 – 0.825)

The studies do not provide information about whether the x-rays were taken as AP or PA films. Excluding the Knudsen (2006) study makes the summary point slight less diagnostic as shown:

Sensitivity = 0.617 (95% CI 0.540 – 0.688)

Specificity = 0.736 (95% CI 0.616 – 0.833)

The presence of cardiomegaly may be suggestive of chronic heart failure, rather than an acute episode but given the remitting and relapsing nature of this condition, there is likely to be some diagnostic value of the recognition of this finding. This is reflected in the summary result. This long period that is required to develop this feature may mean that it is not be independent of other variables, such as a history of a previous diagnosis of heart failure.

#### **6.5.4 Oedema on Chest X-ray**

Chest x-rays are performed on almost all patients presenting to emergency departments due to shortness of breath. The investigation is fast, portable, rapid, non-invasive, and painless and can help, not only with the diagnosis of heart failure, but can also aid in the diagnosis of other conditions, such as pneumonia or pneumothorax. Pulmonary oedema develops in heart failure due to the increased pulmonary venous pressure causing increased hydrostatic pressure leading to capillary leakage. This process was first described by Starling (1896). Fluid first tends to accumulate around the pulmonary vessels with blurring of their appearance on x-ray, subsequently Kerley lines and sub-pleural effusions develop. This is referred to as interstitial pulmonary oedema. Further increases in pressure result in fluid filling the alveoli which causes diffuse opacity of the lower parts of the lungs or occasionally the peri-hilar regions (Gluecker et al., 1999). Two studies specified whether interstitial oedema or alveolar oedema was present on the x-ray; the other studies just had a section for general pulmonary oedema. These studies have been considered together for the purposes of this

analysis. Twelve studies involving 4,162 patients reported on this finding. The results are shown in Figure 6-75

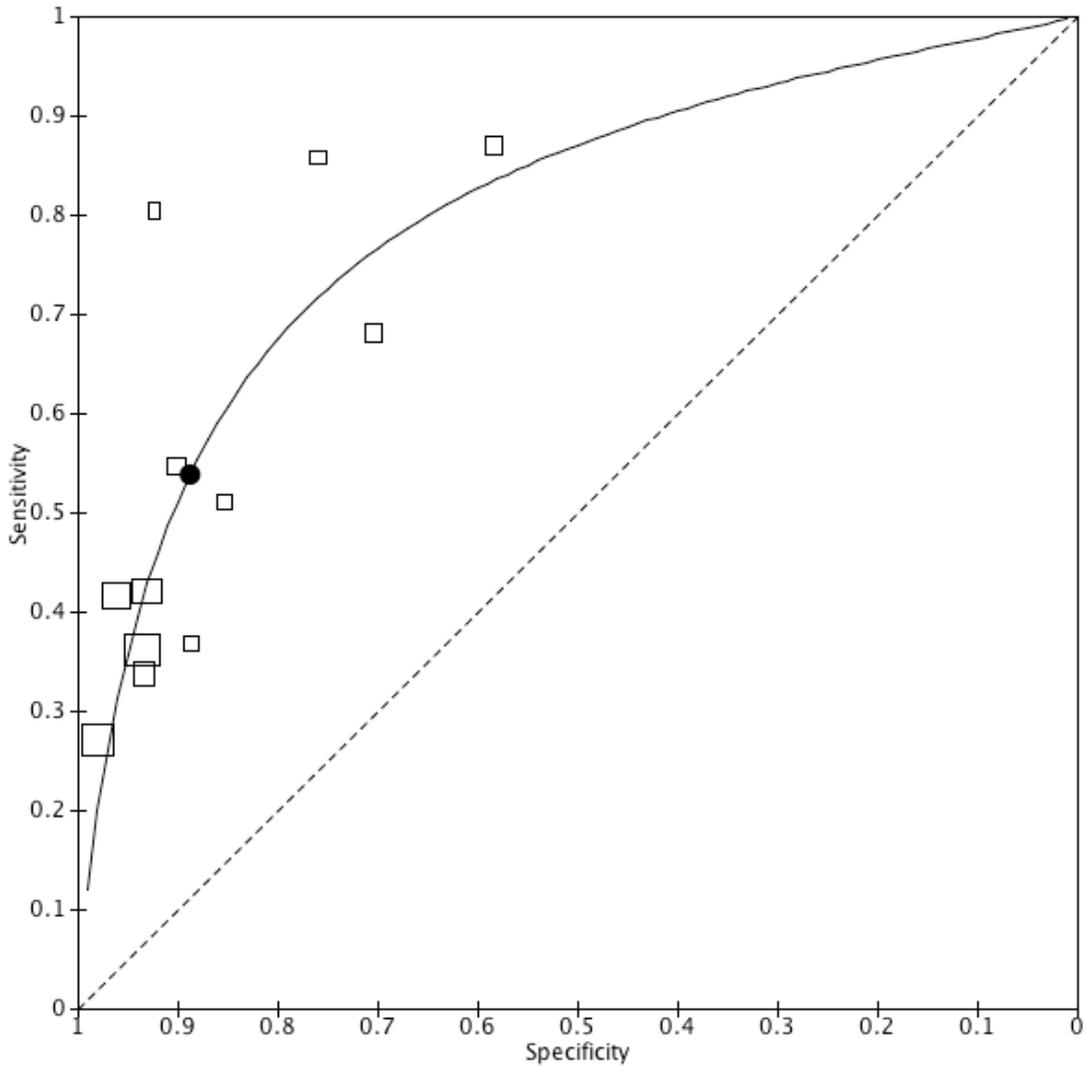


Figure 6-74 Presence of pulmonary oedema on chest x-ray

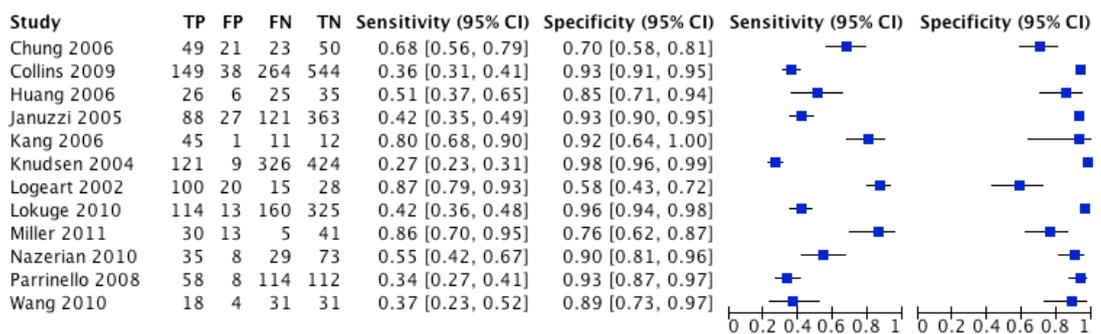


Figure 6-75 Forest Plot - Presence of Pulmonary Oedema on CXR

Analysis of the data using Meta-DiSc provided the following analysis

Heterogeneity chi-squared = 17.54 (d.f. = 11) p = 0.093

Inconsistency ( $I^2$ ) = 37.3%

Estimate of between-study variance (Tau-squared) = 0.0791

All of the studies apart from Kang (2006) appear to follow the estimated summary ROC curve suggesting a positivity threshold effect. In the study reported by Knudsen et al (2004), in which the chest x-rays were reported by radiologists in the emergency department, rather than the attending physician, there was higher specificity than the other studies. This study still lay close to the summary ROC curve as it demonstrate a relatively low sensitivity. The study by Kang et al (2006) was set in Korea and involved a small number of patients. The results suggested very good specificity and reasonable sensitivity but with very wide confidence intervals. Exclusion of this study provided the following results:

Bivariate meta-analysis of the data using MetaAnalyst software provided the following overall estimates:

Sensitivity = 0.543 (95% CI 0.412 – 0.667)

Specificity = 0.890 (95% CI 0.819 – 0.940)

This demonstrates reasonably good specificity with moderate sensitivity. Although this finding does not demonstrate a definitive diagnostic role in isolation, it may have value as part of a diagnostic pathway.

### **6.5.5 Cephalisation of Vessels on Chest X-ray**

Pulmonary vascular congestion occurs as a result of impaired pumping by the left ventricle, which leads to increasing left atrial pressure, which leads, in turn, to pulmonary hypertension. In a person with normal cardiac and pulmonary function the blood flow within the lungs is not distributed evenly. Due mainly to gravity, the blood tends to be present in the lower vessels with

a reduction in relative perfusion of the upper lung compared with the lower lung. In pulmonary hypertension this pattern of distribution is lost as the pressure increases and the upper lobe vessels dilate as a result. This may be visible on viewing a chest x-ray due to the relative dilatation of the upper lung vessels. Four studies examined this investigative finding in 1,633 patients as summarised in **Error! Reference source not found.** and Figure 6-78.

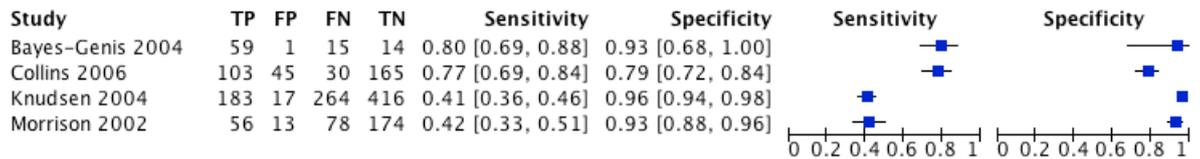


Figure 6-76 Sensitivity and Specificity of Cephalisation of the Vessels on CXR

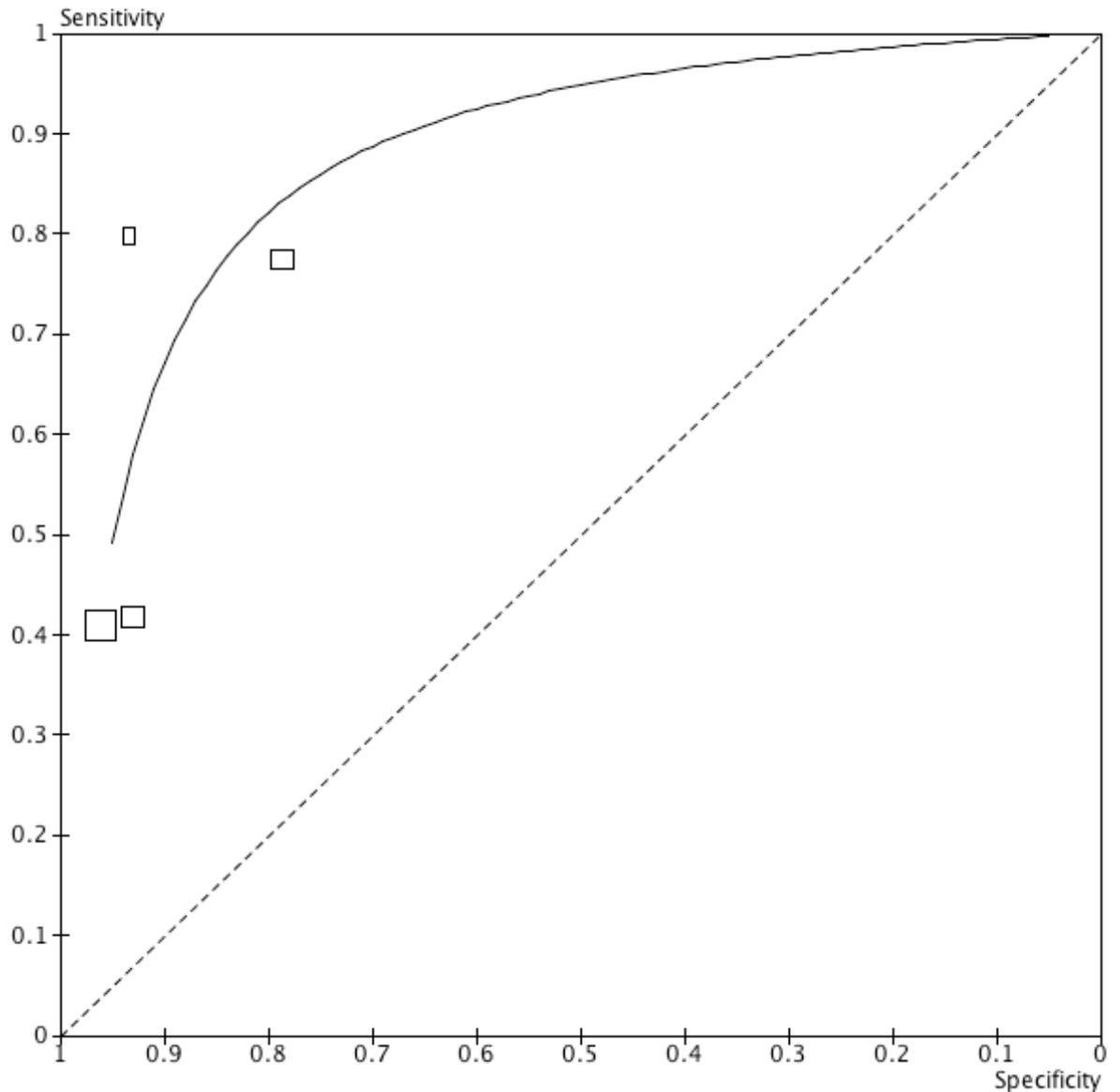


Figure 6-77 Summary ROC Curve for Cephalisation of Vessels

Analysis of the four studies for heterogeneity using Meta-DiSc provided the following results:

Heterogeneity chi-squared = 3.59 (d.f. = 3)  $p=0.309$

Inconsistency ( $I^2$ ) = 16.4%

Estimate of between-study variance (Tau-squared) = 0.0233

This shows relatively low heterogeneity. Two of the studies, Morrison (2002) and Knudsen (2004), show data from the same centre although one of the

authors has confirmed in correspondence that there was no overlap in recruited patients (Schwam, 2004).

Meta-analysis of the results using Meta-Analyst provided the following results.

Sensitivity = 0.603 (95% CI 0.368 – 0.797)

Specificity = 0.894 (95% CI 0.751 – 0.967)

Applying the resultant data to RevMan software provided the following summary ROC curve suggesting that the finding may have a diagnostic role. The finding is unlikely to be independent of the presence of pulmonary oedema on the x-ray.

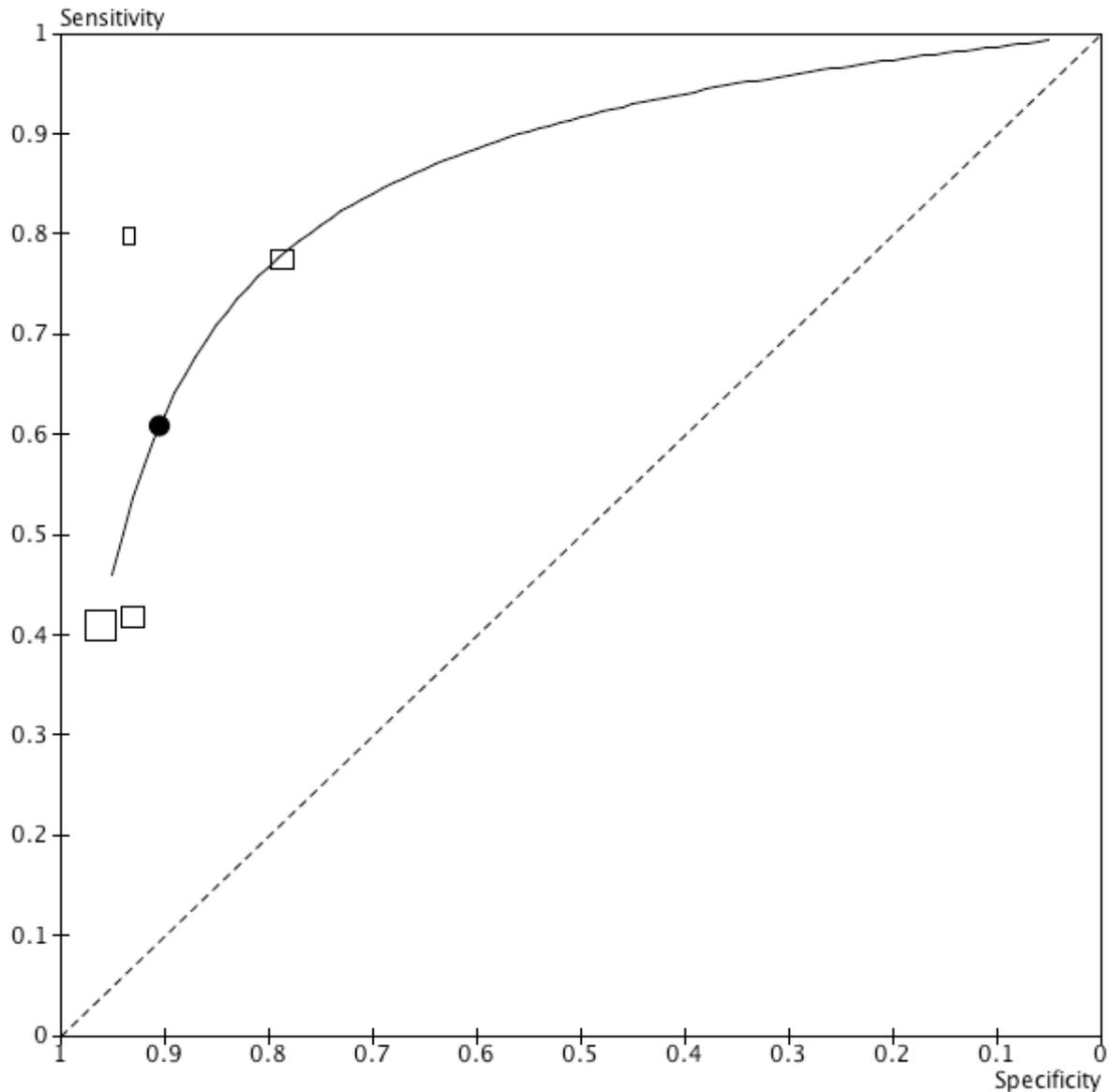


Figure 6-78 Summary ROC Curve with Bivariate Data Cephalisation Vessels

### 6.5.6 Brain Natriuretic Peptide 100 pg/ml

Nineteen studies provided information about a total of 27 different cut-off values for Brain Natriuretic Peptide (BNP) with several studies providing data at various cut-off thresholds. There was also more than one type of BNP assay used. For this analysis the value of 100pg/ml as a cut-off was chosen as this is the level recommended by both Biosite Inc. and Abbott Laboratories to rule out heart failure. These manufacturers produce the assays have been used in the vast majority of the published studies. A total of sixteen studies

provided information about this threshold for BNP derived from 7,659 patients as shown in Figure 6-79 and Figure 6-81. This cut-off point was deliberately chosen in order to rule-out heart failure and has therefore been chosen to maximise sensitivity. The sensitivity is correspondingly high in all of the selected studies, though the specificity is more variable.

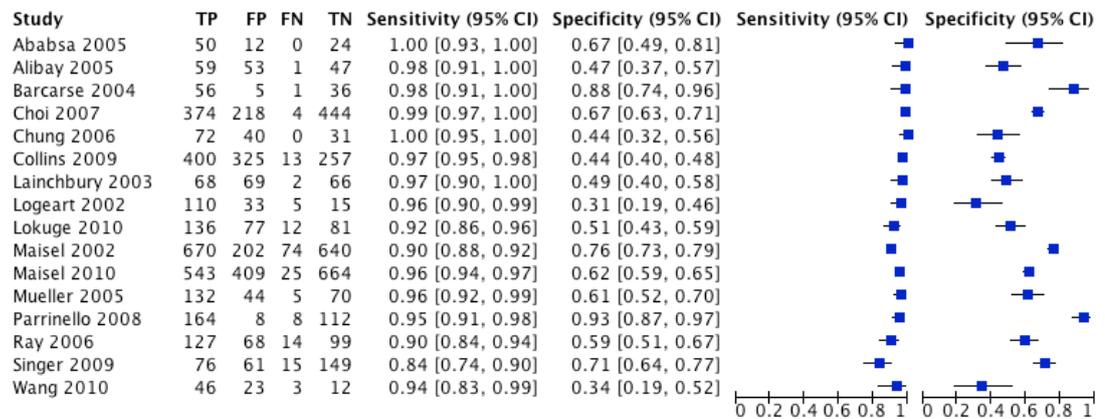


Figure 6-79 Forest Plot - Brain Natriuretic Peptide 100pg/ml

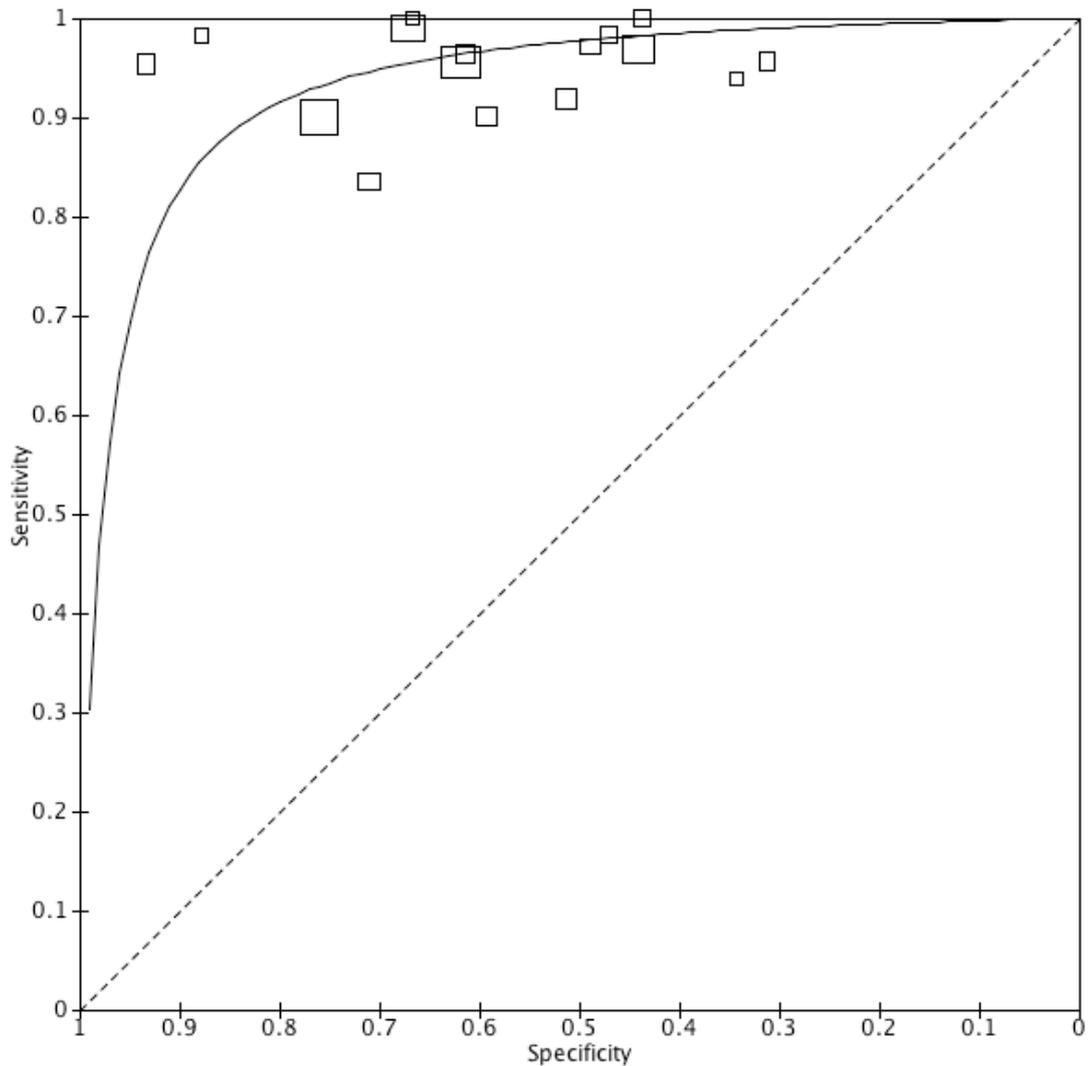


Figure 6-81 Summary ROC Curve for BNP 100pg/ml

The majority of studies used the Biosite assay while two studies, Mueller (2005) and Ababsa (2005), used the Abbot Laboratories assay. Testing for heterogeneity provided the following results:

Heterogeneity chi-squared = 69.24 (d.f. = 14)  $p = 0.000$

Inconsistency ( $I^2$ ) = 79.8%

Estimate of between-study variance (Tau-squared) = 0.5360

Eliminating studies on the basis of size, assay used or various measures of quality failed to significantly reduce the inconsistency score. This variation

may be due to the variety in the cut-off value of the applied reference standard. There is no obvious correlation between the prevalence of heart failure in the studies and the level of specificity of the test.

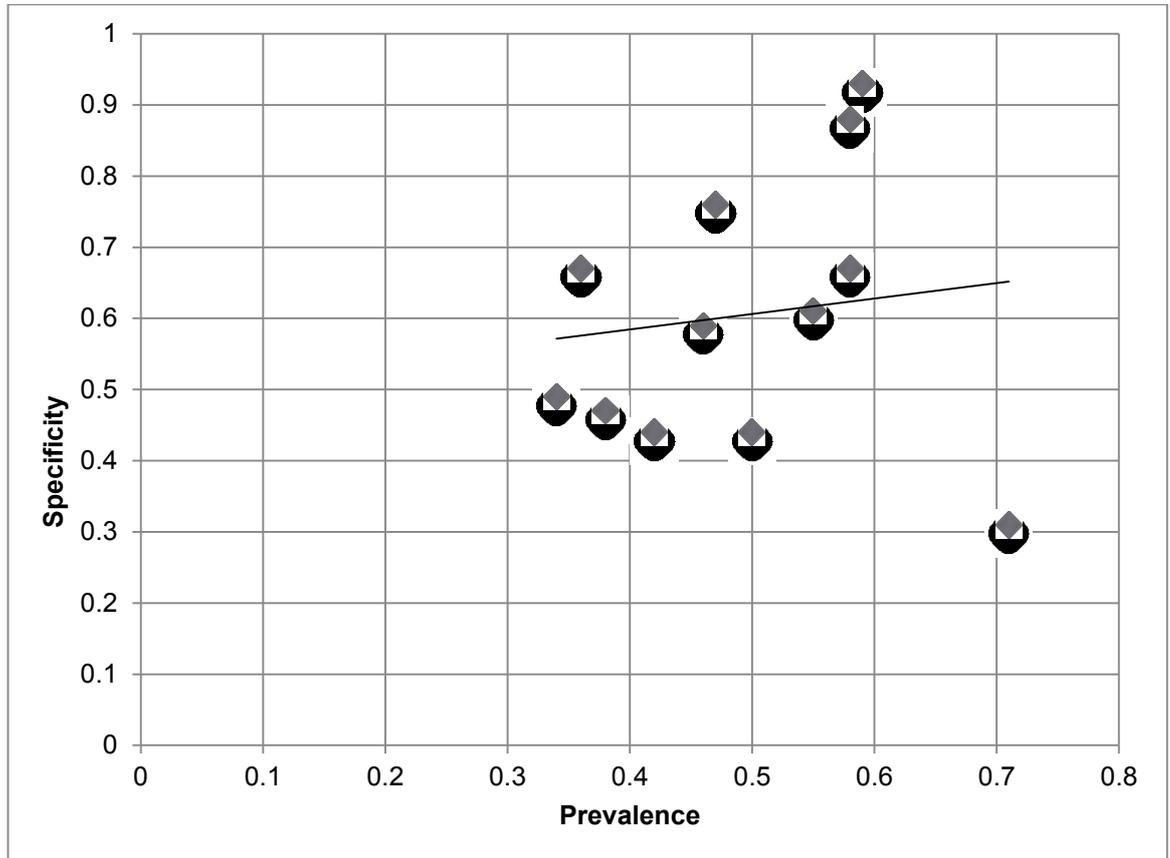


Figure 6-82 - Prevalence of Heart Failure Versus Specificity for BNP Studies

Bivariate meta-analysis of this data gave the following results:

Sensitivity = 0.958 (95% CI 0.939 – 0.974)

Specificity = 0.608 (95% CI 0.499 – 0.707)

Six of the studies included the BNP result as part of the reference standard. Analysis of the studies where BNP did not form part of the reference standard was also performed. This analysis estimated lower heterogeneity and actually improved sensitivity.

Heterogeneity chi-squared = 21.84 (d.f. = 8)  $p = 0.005$

Inconsistency ( $I^2$ ) = 63.4%

Estimate of between-study variance (Tau squared) = 0.2146

Sensitivity = 0.965 (95% CI 0.948 – 0.979)

Specificity = 0.551 (95% CI 0.441 – 0.652)

### 6.5.7 Brain Natriuretic Peptide 400 pg/ml

The cut-off recommended by Biosite Inc. to rule in heart failure is 400 pg/ml. Six studies, involving 2,564 patients were found to provide data about this value as illustrated in **Error! Reference source not found.** and Figure -6-84. As this cut-off value was chosen to maximise specificity it is unsurprising that the specificity of this test is excellent, although the sensitivity also remains reasonable at this threshold.

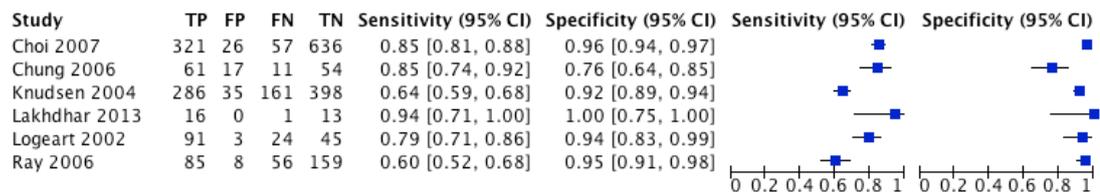


Figure 6-83 Forest Plot for BNP with 400 pg/ml Cut-off

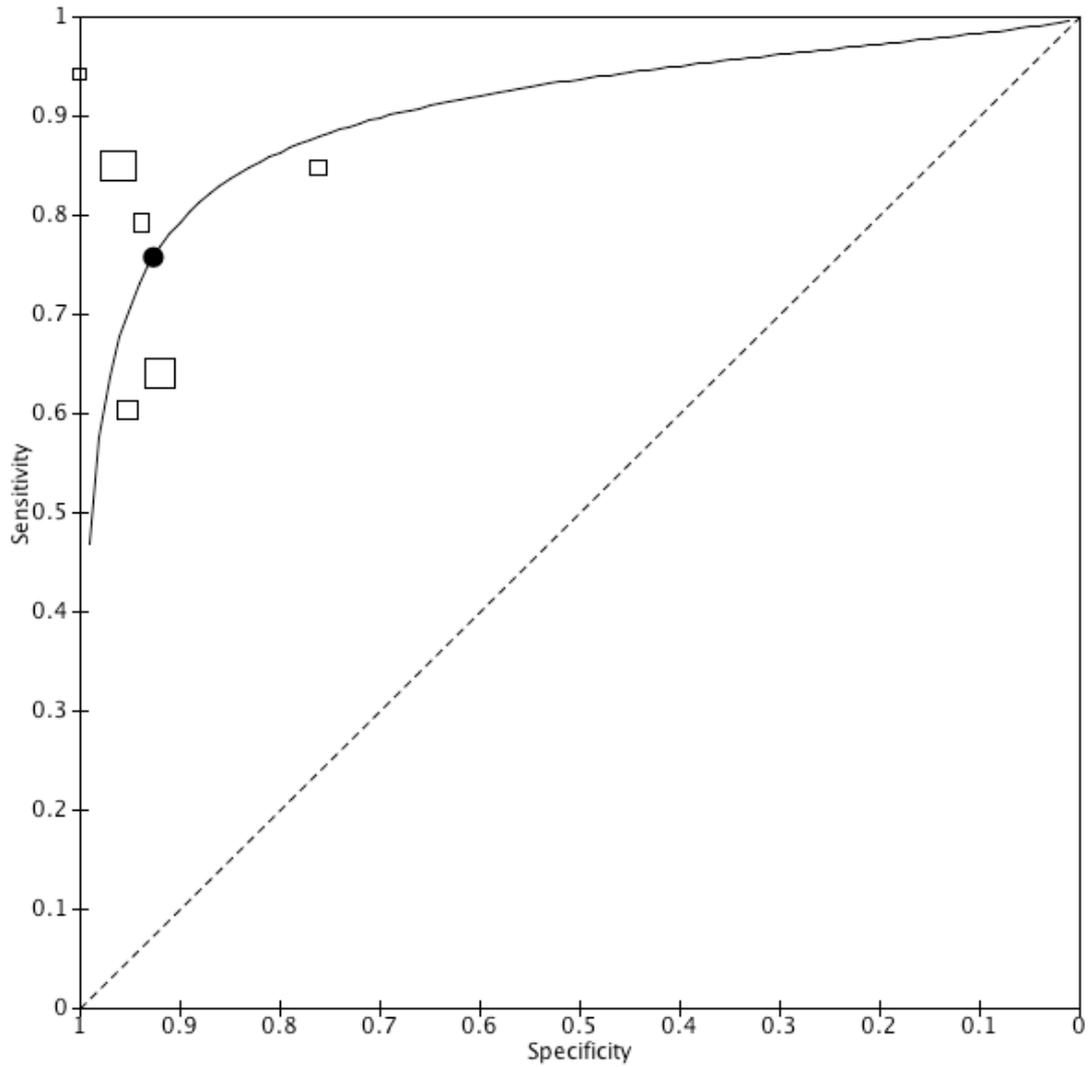


Figure -6-84 Summary ROC Curve - BNP 400 pg/ml

Analysis of this data using Meta-DiSc provided the following estimates:

Heterogeneity chi-squared = 42.53 (d.f. = 5) p = 0.000

Inconsistency ( $I^2$ ) = 88.2%

Estimate of between-study variance (Tau-squared) = 0.9784

Meta-analysis of the data using MetaAnalyst software provides the following summary point:

Sensitivity = 0.771 (95% CI 0.664 – 0.865)

Specificity = 0.927 (95% CI 0.850 – 0.975)

This would suggest significant heterogeneity although this is not apparent on looking at the summary ROC curve. The  $I^2$  score is considered to exaggerate the level of heterogeneity when a small number of studies are used (Hardy & Thompson, 1998).

The obvious outlying study in terms of specificity is Chung et al (2006). The inclusion criteria for this study included severe dyspnoea that the attending physician felt necessitated hospital admission; this may have been a different population from the other studies that included all attenders with acute dyspnoea.

Excluding this study from the estimate of heterogeneity failed to improve the estimated level of statistical heterogeneity as shown:

Heterogeneity chi-squared = 38.71 (d.f. = 4)  $p = 0.000$

Inconsistency ( $I^2$ ) = 89.7%

Estimate of between-study variance (Tau-squared) = 1.0959

Excluding the Chung (2006) study improved the specificity as shown:

Sensitivity = 0.761 (95% CI 0.640 – 0.873)

Specificity = 0.950 (95% CI 0.916 – 0.979)

### **6.5.8 'N' Terminal proBNP (NT-proBNP)**

NT-proBNP also has two cut-off levels recommended by Roche Diagnostics who are the main manufacturers of this assay. Patients with levels below 300 pg/ml are considered very unlikely to have acute heart failure and patients with levels greater than 1800 pg/ml are considered very likely to have acute heart failure. Patients with levels in between these two values are considered to have moderate likelihood of acute heart failure. Three studies provided data about the cut-off value of 300 pg/ml and a further study, Mueller (2005),

provided data from the cut-off value of 292 pg/ml, which has been included in this group. This data represents very high sensitivity as demonstrated in **Error! Reference source not found..**

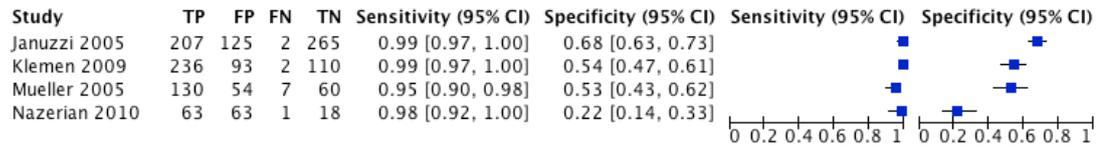


Figure 6-85 Forest Plot Using NT-proBNP Cut-off Value of 300 pg/ml

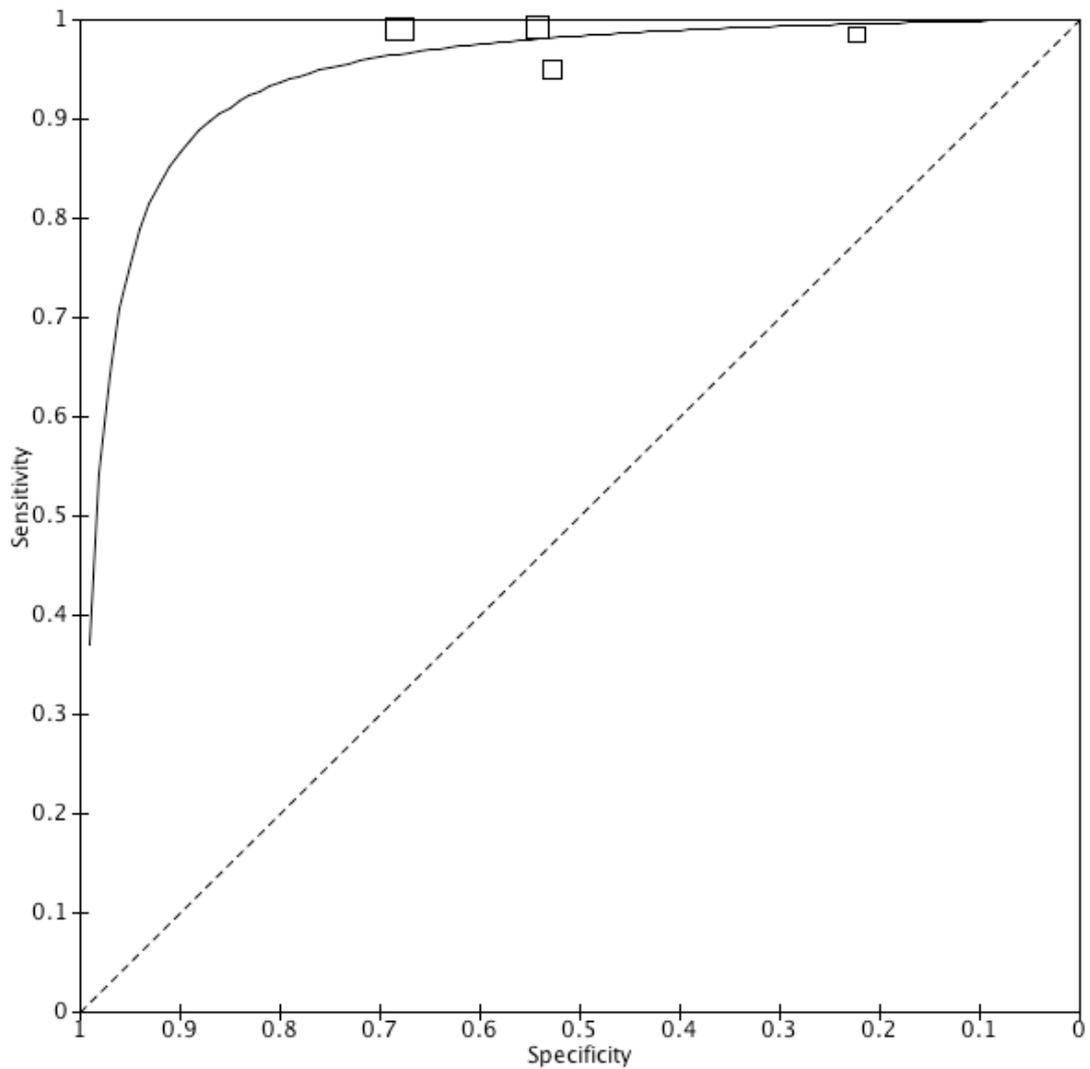


Figure 6-86 Summary ROC Curve for NT-proBNP with Cut-off Value of 300 pg/ml

Only four studies provide information about this cut-off level. Meta-analysis is not appropriate with so few studies. Pooling of the data provides an approximate summary value. The data was pooled using Meta-DiSc software to give the following results:

Pooled sensitivity = 0.981 (95% CI 0.968 – 0.990)

Pooled specificity = 0.575 (95% CI 0.539 – 0.610)

The manufacturers also recommend a higher cut-off value of 1800 pg/ml. No studies were found that used this cut-off value to rule-in the condition of heart failure but one study used 2000 pg/ml, Klemen (2009) and a further study, Lainchbury (2003), used a cut-off of 240 pmol/l which is equivalent to 2029.7 pg/ml and these have been grouped together for the purpose of the analysis. This cut-off provides good sensitivity and specificity as displayed in

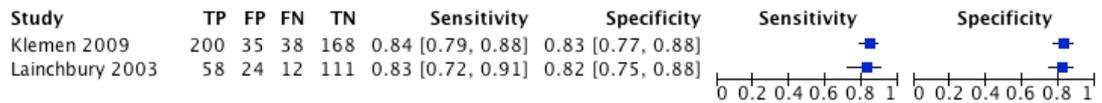


Figure 6-87 and Figure 6-88.

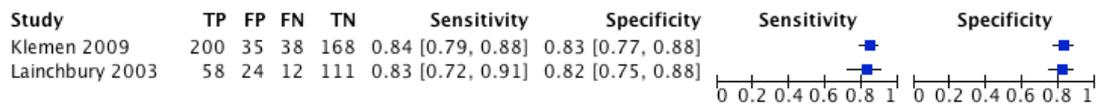


Figure 6-87 Forest Plot Using NT-proBNP with Cut-off Value of 2000 pg/ml

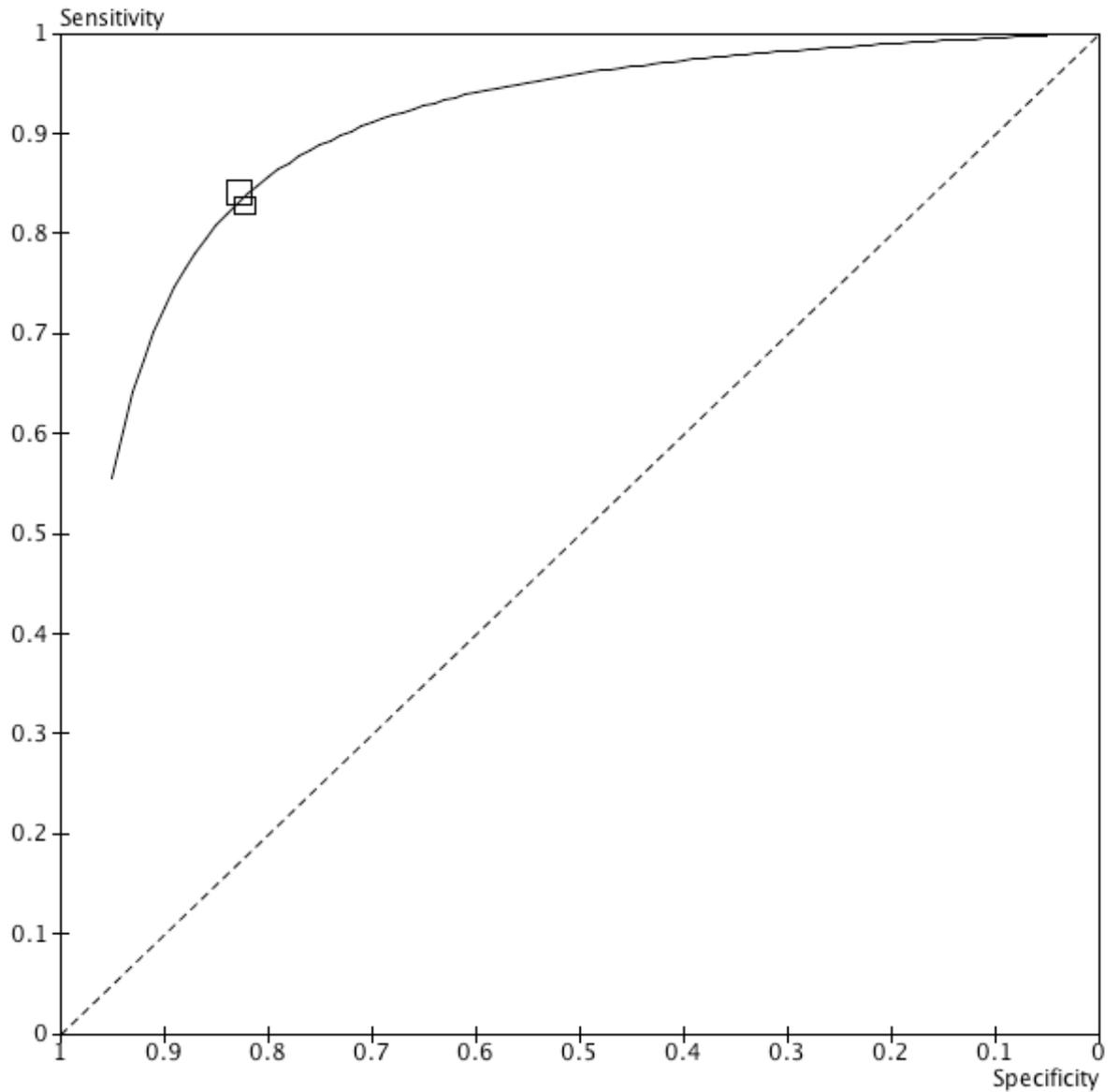


Figure 6-88 Summary ROC Curve Using NT-proBNP with Cut-off Value of 2000 pg/ml

Only two studies provided information about this cut-off. The results have been pooled using Meta-DiSc software.

Pooled sensitivity = 0.838 (95% CI 0.792 – 0.877)

Pooled specificity = 0.825 (95% CI 0.781 – 0.864)

## 6.6 Discussion

“Mankind always sets itself only such problems as it can solve; since, looking at the matter more closely, it will always be found that the task arises only when the material conditions for its solutions already exist or at least are in the process of formation.” - Karl Marx (1859)

A summary of the optimum diagnostic variables is provided in Table 6-9. It can be seen from the results of the meta-analysis that the natriuretic peptides, BNP and NT-proBNP, are superior to any other individual diagnostic test and even appear to out-perform the clinical assessment by the attending physician.

Although a search of the literature dating back to 1950 was performed, the majority of papers concerning the diagnosis of heart failure have been published since the discovery of BNP, suggesting that the difficulties in making this diagnosis were not acknowledged, at least in medical literature, until this superior test was discovered.

The variables tend to perform well in terms of either sensitivity or specificity, with the perfect diagnostic test that is both sensitive and specific, remaining elusive. The use of upper and lower cut-off values for the natriuretic peptides, optimised for specificity and sensitivity respectively, bypass this issue to some extent, but still leave a zone of intermediate risk. Unfortunately, there is some evidence that this zone overlaps with the region of clinical diagnostic uncertainty (Schwam, 2004).

Having selected the strongest candidates for inclusion in a clinical decision rule (CDR) and having derived summary values to represent their diagnostic performance it would appear to be a simple process to combine the sensitive and specific tests to derive a robust CDR. Unfortunately, by examining each of the tests in isolation there is no assessment of the independence of the tests. In order to derive the optimum combination of diagnostic variables to synthesise the CDR, individual patient data must be used so that it is possible

to perform multivariable analysis and assess the utility of the potential components in this situation.

## 6.6.1 Summary of Findings

Table 6-9 Summary Table - Diagnostic Utility of Selected Variables

Diagnostic Variable	Summary Sensitivity	Summary Specificity	Positive Likelihood Ratio (LR+)	Negative Likelihood Ratio (LR-)
History of HF	0.514 (95% CI 0.431 – 0.600)	0.864 (95% CI 0.819 – 0.903)	2.02	0.23
Clinical Opinion	0.603 (95% CI 0.483 – 0.713)	0.919 (95% CI 0.852 – 0.965)	6.94	0.43
Large cardiac shadow on CXR	0.640 (95% CI 0.554 – 0.718)	0.75 (95% CI 0.652 – 0.825)	2.76	0.5
Pulmonary oedema	0.543 (95% CI 0.412 – 0.667)	0.890 (95% CI 0.819 – 0.940)	4.47	0.53
Cephalisation of vessels	0.603 (95% CI 0.368 – 0.797)	0.894 (95% CI 0.751 – 0.967)	5.69	0.44
BNP 100 pg/ml	0.958 (95% CI 0.939 – 0.974)	0.608 (95% CI 0.499 – 0.707)	2.15	0.05
BNP 400 pg/ml	0.771 (95% CI 0.664 – 0.865)	0.927 (95% CI 0.850 – 0.975)	12.79	0.29

NT-proBNP 300 pg/ml	0.981 (95% CI 0.968 – 0.990)	0.575 (95% CI 0.539 – 0.610)	2.55	0.03
NT-proBNP 2000 pg/ml	0.838 (95% CI 0.792 – 0.877)	0.825 (95% CI 0.781 – 0.864)	4.79	0.20
Impaired Renal Function	0.299 (95% CI 0.157 – 0.484)	0.907 (95% CI 0.849 – 0.947)	2.82	0.81
Raised JVP	0.341 (95% CI 0.276 – 0.412)	0.920 (95% CI 0.889 – 0.945)	4.52	0.75
Presence of heart murmur	0.279 (95% CI 0.219 – 0.351)	0.912 (95% CI 0.870 – 0.943)	3.30	0.77
Raised Troponin	0.408 (95% CI 0.234 - 0.629)	0.908 (95% CI 0.802 - 0.970)	3.31	0.78
Hepato- jugular Reflux	0.164 (95% CI 0.119 - 0.217)	0.939 (95% CI 0.903 - 0.965)	2.69	0.89
Presence of 3 <sup>rd</sup> HS	0.100 (95% CI 0.070 – 0.138)	0.985 (95% CI 0.973 – 0.993)	6.56	0.91
Presence of 4 <sup>th</sup> HS	0.037 (95% CI 0.026 – 0.052)	0.983 (95% CI 0.975 – 0.989)	2.18	0.98

## 6.6.2 Limitations of the Systematic Review

The search has designed to be as open and inclusive as possible to ensure that no studies concerned with the diagnosis of acute heart failure in the Emergency Department setting were missed. Publication bias in diagnostic studies is an unknown quantity (Song et al., 2002). It is very likely to exist, as the same pressures that encourage the publication of positive results for interventional studies apply to diagnostic studies but this matter has not been studied. The publication bias for interventional trials has been quantified by the requirement to register the trials before they commence but no such registers have been created for the diagnostic studies. The paucity of new data generated by the grey literature search did not suggest a wealth of unpublished material.

It was recognised that the quality of the included studies was variable. The decision was made not to exclude studies directly on the basis of poor quality but to assess the impact of these studies on the sum of knowledge and explore the effects of including these studies.

Likewise it was recognised that considerable heterogeneity exists even in studies purporting to study the same variable, with the same threshold, in the same population. This is to be expected given the lack of an objective reference standard and variations in definitions and populations that are likely to exist in a variety of geographic and social environments. Heterogeneity has not been grounds for exclusion of studies and the possible reasons for its presence and the effects on the results have been explored in the analysis.

Most of the variables that have been examined may also form part of the reference standard. The results of this inclusion bias are difficult to ascertain. In theory, if a certain variable were to be considered pathognomonic for a diagnosis, then the presence a positive finding would confirm both the diagnosis and the importance of this variable for making the diagnosis. In practice, none of the variables examined demonstrated this degree of diagnostic certainty, and the only variables that approached this optimal performance, the natriuretic peptides, retained this diagnostic utility even when the blinding was in place.

It was recognised that the ideal clinical decision rule, combining several independent predictor variables, which each demonstrated good diagnostic utility, was unlikely to be attained.

The variation in the applied reference standards for the included studies is likely to be one of the sources of heterogeneity in this review. In the absence of a recognised gold standard all that can be done is to acknowledge that this is not ideal. The quality of the reference standards used was considered as part of the inclusion criteria for the selected studies, and the reference standard applied has been considered in attempting to explain apparent heterogeneity. Information about the reference standard used has been provided for each of the included studies so that the reader can make an informed interpretation of the provided analysis. This is a pragmatic solution as, in real-life practice, the diagnosis has to be made and treatment initiated, often with less evidence than is available in these studies. While hindsight allows greater clarity of vision, these patients require immediate treatment, and the challenge is to ensure that the immediate treatment is also the optimal treatment.

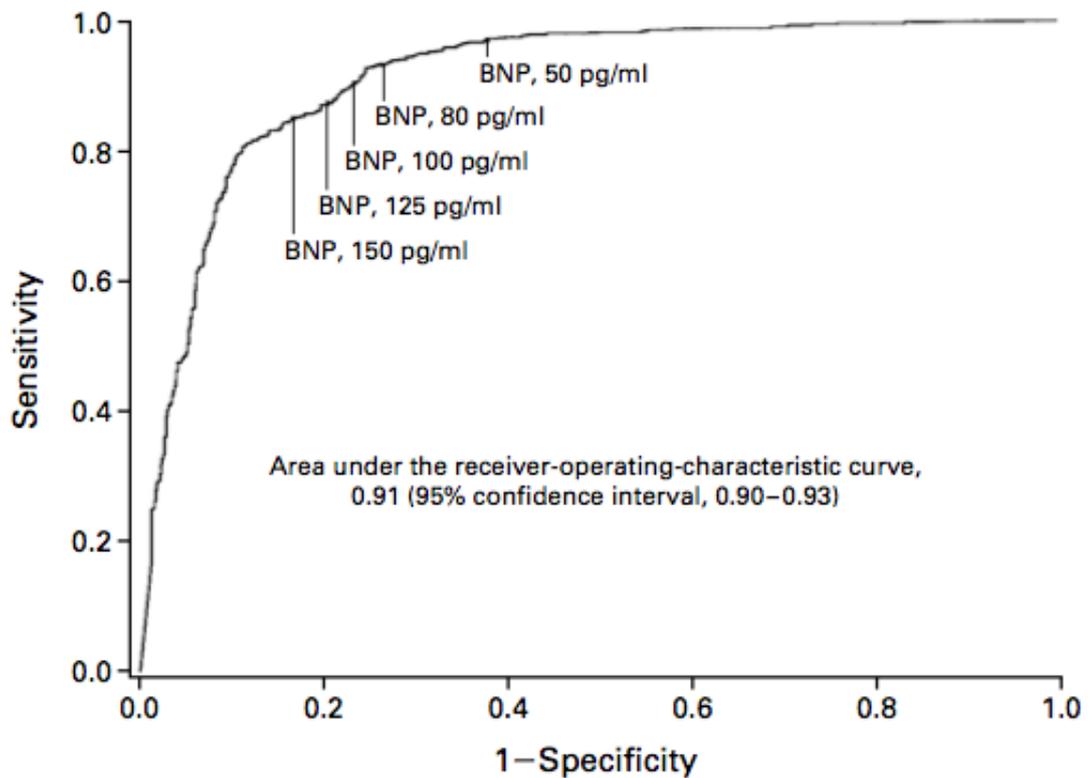
One issue with this review is that the tests are being examined and analysed in isolation; this does not reflect clinical practice. It has been suggested that diagnostic tests perform better in combination than they do individually (Irwig et al., 2006). The variable of clinical opinion overcomes this to some degree, although this will also have inclusion bias due to the similarity this will bear to the end-diagnosis. If a patient is confidently diagnosed as having heart failure in the emergency department and the specific treatment for this condition is commenced and continued then this may influence the end-diagnosis even when the person making the end-diagnosis is blinded to the documented opinion of the emergency physician.

Another limitation of this systematic review is that it does not assess the independence of the diagnostic variables. This information is essential if one intends to create a diagnostic pathway as it optimises the value of each

applied diagnostic test. One would also like to know how the diagnostic variables perform in combination as it there may a synergistic effect where the combined tests perform better than would be expected from combining their individual utility values.

### **6.6.3 Limitations of Brain Natriuretic Peptide**

A recognised problem with the use of the natriuretic peptide levels is the presence of a diagnostic 'grey area'. A very low result rules out heart failure with excellent sensitivity and a very high level of BNP has sufficient specificity to confirm the diagnosis with confidence. This leaves a range of results between these two cut-off values where the diagnostic power is less definitive, although it still compares well with any other individual diagnostic test. The two studies that provide data about both cut-off values, Choi (2007) and Logeart (2002), had around 25% of the included patients in this diagnostic grey zone. Maisel (2002) presents a ROC curve for the BNP at various thresholds that is shown in Figure 6-89. This demonstrates that even with values in the middle of the diagnostic 'grey-zone' there is still reasonable sensitivity and specificity. A diagnostic pathway could use a range of values rather than just a positive or negative dichotomy in order to maximise the information provided by this result.



BNP pg/ml	SENSITIVITY	SPECIFICITY (95 percent confidence interval)	POSITIVE PREDICTIVE VALUE	NEGATIVE PREDICTIVE VALUE	ACCURACY
50	97 (96-98)	62 (59-66)	71 (68-74)	96 (94-97)	79
80	93 (91-95)	74 (70-77)	77 (75-80)	92 (89-94)	83
100	90 (88-92)	76 (73-79)	79 (76-81)	89 (87-91)	83
125	87 (85-90)	79 (76-82)	80 (78-83)	87 (84-89)	83
150	85 (82-88)	83 (80-85)	83 (80-85)	85 (83-88)	84

Figure 6-89 ROC Curve and Data for BNP Taken From Maisel 2002

BNP and NT-proBNP are cleaved from the same molecule and so should be created in the same proportion at the same time. In fact NT-proBNP levels rises more than BNP in the acute setting. This is probably due to differences in the clearance method with NT-proBNP being eliminated by renal clearance while BNP is broken down by endo-peptidases. The diagnostic utility of these molecules is very similar as can be seen by combining ROC curves for BNP at 100 pg/ml and NT-proBNP at 300 pg/ml as shown in Figure 6-90. The summary ROC curves are almost identical. These variables are unlikely to be

independent, as the values would be expected to correlate with each other in any particular patient.

There are known limitations for the use of brain natriuretic peptide as a marker for acute heart failure; levels are known to be affected by age, renal function, obesity, sex and other conditions that can cause ventricular strain such as pulmonary embolism. These factors are explored in more detail in Chapter 3 of this thesis. The reasons for false positive results were not documented or explained in the included studies although it may be due to patients having with some degree of cardiac failure even if this is not the primary cause for their attendance. As a high BNP level is a significant predictor of all-cause mortality, it is probably prudent to recommend admission and further investigation of these patients, even if heart failure is not felt to be the cause of their symptoms (Hammerer-Lercher et al., 2010).

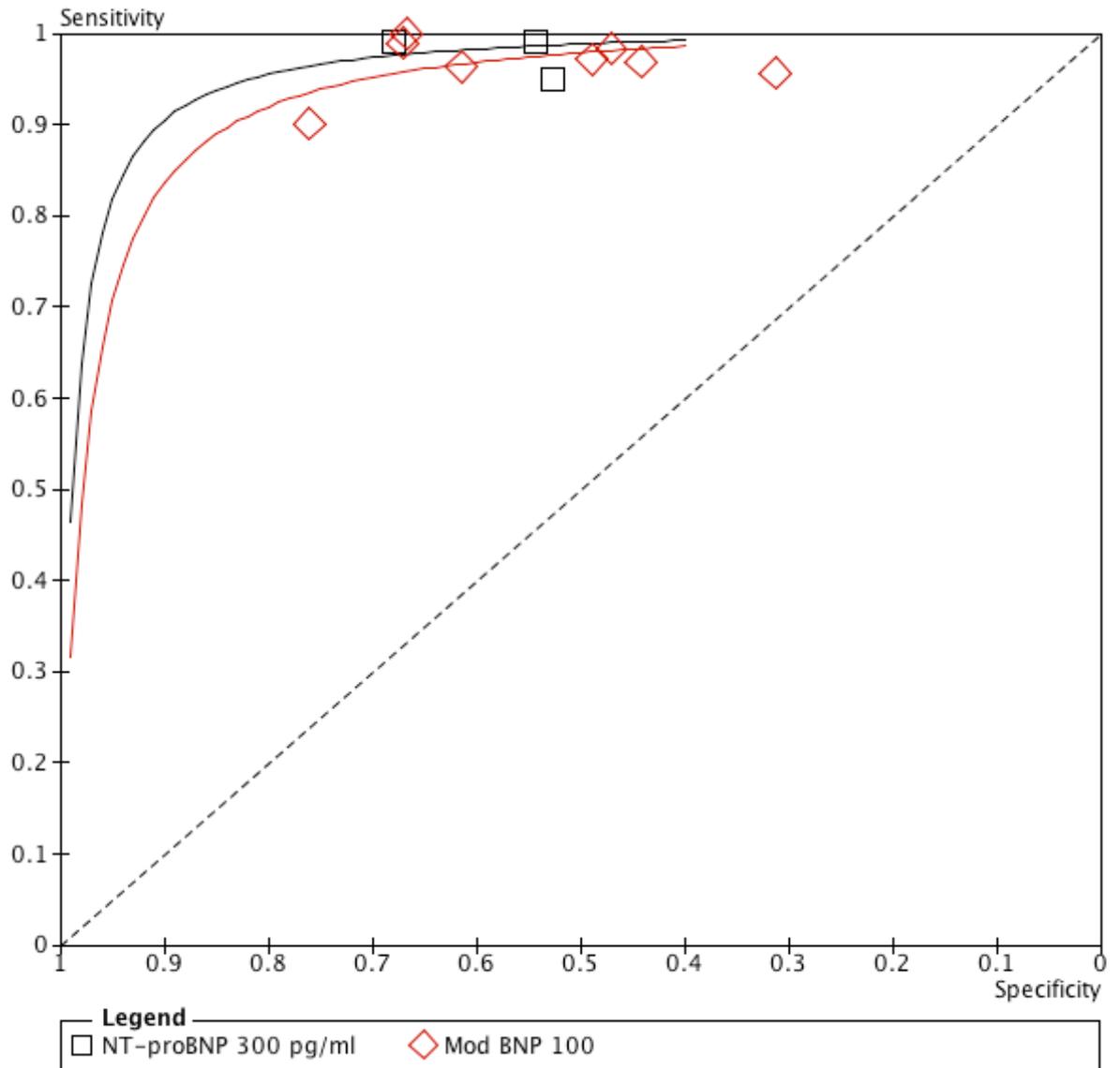


Figure 6-90 Summary ROC Curve Comparing BNP 100 pg/ml data against NT-proBNP 300 pg/ml data

### 6.6.4 Other Reviews of the Literature

Other systematic reviews examining the diagnosis of heart failure have been conducted, although these have been concerned with the diagnosis of chronic heart failure and so have not been in the Emergency Department setting. Issues that have previously been explored have been concerned with the selection of appropriate patients for further elective investigation and specialist referral.

Khunti et al. (2000) looked performed a systematic review of the available literature in order to assess the value of clinical and historical findings in confirming a diagnosis of left ventricular systolic dysfunction in patients presenting in a primary care setting. The authors concluded that although some symptoms and signs were suggestive of heart failure, there were no findings conclusive enough to make a firm diagnosis without further investigation such as ECG or echocardiogram. Significant findings were the presence of dyspnoea, ankle swelling, a history of myocardial infarction, hypertension or angina, raised JVP, oedema, abnormal apical pulse, basal crepitations and a baseline pulse between 90 – 100 bpm.

Fonseca et al. (2004) looked at a random selection of 6300 patients over the age of 25 to examine the prevalence of heart failure within the population of Portugal. Patients who were suspected of having heart failure underwent chest radiography, ECG, blood tests and an echocardiogram. 17% of the patients were excluded, as an echocardiogram could not be performed. In this sub-group of selected patients, an abnormal ECG had a sensitivity of 80% and a specificity of 40%. An abnormal chest x-ray had sensitivity of 57% and a specificity of 78%. The most sensitive components of chest x-ray abnormality were the presence of cardiomegaly and an enlarged cardiothoracic ratio ( $>0.50$ ). Pulmonary vessel cephalisation and lung interstitial oedema had a sensitivity of 18%.

Craig et al. (2005) performed a Health Technology Assessment examining the use of natriuretic peptides in the investigation of patients with suspected heart failure. The authors performed a literature review and cost effectiveness assessment for the use of natriuretic peptide measuring in patients in primary care and in the acute setting. The main recommendations for the use of natriuretic peptide testing were as a rule-out test for General Practitioners who were not confident in ECG interpretation and in the acute setting where there remained diagnostic uncertainty.

Davenport et al. (2006) carried out a systematic review and meta-analysis to assess the accuracy of natriuretic peptides and ECG in the diagnosis of left ventricular systolic dysfunction. The authors selected studies using nuclear studies or echocardiography as a reference standard to confirm the dysfunction. The authors found that while ECG, BNP and NT-proBNP tended to be sensitive tests for heart failure there remained considerable variation in the published data. The specificity results tended to be lower and were even more heterogeneous for all three tests. The authors also found that combining BNP and ECG findings did not improve the diagnostic value beyond that of the individual tests.

Clerico et al. (2007) performed a systematic review and meta-analysis of studies comparing the diagnostic utility of BNP with that of NT-proBNP for patients with either chronic or acute heart failure. The authors included and analysed 15 studies concerned with diagnosing patients with chronic heart failure, and 9 studies examining the relative diagnostic values in patients with acute heart failure. The authors did not find any significant difference between the two compounds in terms of their diagnostic value. The diagnostic odds ratio (DOR) for the condition of chronic heart failure for BNP was 8.44 (95% CI 4.66 – 15.30) versus 23.36 (95% CI 9.38 – 58.19) for NT-proBNP; for acute heart failure the respective DORs were 16.46 (95% CI 10.65 – 25.43) and 18.61 (95% CI 12.99 – 26.65).

Mant et al. (2010) et al performed a systematic review and meta-analysis examining the diagnosis of heart failure within primary care on behalf of the Health Technology Assessment programme. The patient population used in this analysis is significantly different to the patients presenting to the Emergency Department in terms of the acute onset of the condition. The investigations available to the General Practitioner may be more limited and there will usually be a greater delay in receiving the results of these tests. The authors found that dyspnoea had high sensitivity but poor specificity for the diagnosis of heart failure, while a history of myocardial infarction, orthopnoea, oedema, cardiomegaly, added heart sounds and hepatomegaly all had high specificity but low sensitivity. The recommendation was that patients with

dyspnoea and a history of myocardial infarction, basal crepitations or ankle oedema should be referred for echocardiography. In patients without these findings, a natriuretic peptide level should be checked and depending on certain cut-offs which vary depending on the patient's sex and the presence or absence of ankle oedema, a decision should then be made about whether the patient requires echocardiography. BNP and NT-proBNP were found to be of equivalent diagnostic value.

The emergency department attracts a different population of patients than those attending their family doctor and these patients are more likely to represent the more severe end of the disease spectrum. Patients presenting with an acute onset of symptoms may not have had time to develop the stigmata of chronic heart failure. It may also be difficult to obtain a history of recent events or past medical history in the acutely unwell, dyspnoeic patients and it is unlikely that the Emergency Physician will have any prior knowledge of the patient. On the other hand, there is greater accessibility for tests and investigations with results available in minutes or hours rather than days or weeks.

### **6.6.5 Application of Study Findings**

In the consideration of the patient with undifferentiated acute dyspnoea, the patient still needs to be assessed in the traditional manner with regards to history taking, examination and the relevant investigations performed, although it is now possible to estimate the diagnostic utility of each of the components with regards to the diagnosis of heart failure. Individual components can be combined to produce a more definitive post-test probability to confirm or refute the diagnosis of heart failure.

It remains important to define the aetiology of the patients' symptoms and so, investigations that may impart information suggesting diagnoses other than heart failure should be continued. Informing a dyspnoeic patient that their symptoms do not arise from heart failure is likely to provide little comfort if an alternative diagnosis cannot be provided.

Although the systematic review and meta-analysis suggest that the natriuretic peptides bear significant diagnostic promise, questions that remain unanswered include which group of patients, and at what stage in the diagnostic pathway, should these tests be used?

### **6.6.6 The Need For Original Research**

A further study was felt to be necessary to confirm the findings of the systematic review in a prospective study in the UK setting, thus allowing a wide range of potential diagnostic variables to be reliably collected and then analysed. This would allow determination of the independence of the variables and facilitate the derivation of a clinical decision rule to guide clinicians towards an optimal and consistent diagnostic route. The methods and data analysis for the clinical study are presented in the next chapter.

## Chapter 7 Original Diagnostic Study

### 7.1 Introduction

#### 7.1.1 Research Background

The intention of this research was to devise a clinical decision rule for the diagnosis of heart failure, confirming or refuting its presence in patients presenting to an Emergency Department due to acute breathlessness. The rule had to be applicable to patients in the United Kingdom setting and only incorporate information that is routinely available at the time of patient's attendance in an Emergency Department.

#### 7.1.2 Considerations for this Study

In order for the research to be applicable to clinical practice it was important to use the clinical findings and parameters that would be readily available in most emergency departments in the UK. As these decisions are currently made in clinical practice in the emergency setting everyday, a pragmatic approach to this study was considered appropriate. The clinical decisions have to be made quickly as it benefits the patient to have rapid and appropriate treatment started as soon as possible.

In order to confirm the diagnosis in this study, each patient had an echocardiogram performed as part of the reference standard. This is not standard practice in UK Emergency Medicine although the use of general ultrasound has become more commonplace in UK departments and has recently been added to the College of Emergency Medicine Curriculum (CEM, 2014). This means that most departments in the UK have access to an ultrasound machine that is capable of at least limited echocardiography views and trainees with a potentially transferrable skill set. Other studies have examined the use of Emergency Department echocardiography in the assessment of patients with acute coronary syndrome (Mohler et al., 1998).

Although natriuretic peptides are used in the diagnosis of acute heart failure in some European and US centres, BNP and NT-proBNP testing do not form part of current standard practice in the UK setting. It was therefore considered reasonable to keep the attending physician blind to these investigation results for the purposes of this study.

High sensitivity Troponins now form part of the standard investigations for Acute Coronary Syndrome (ACS) but the utility of this test for the diagnosis of heart failure remains unknown. As this does not form part of current practice, it was considered reasonable to perform this test without informing the attending physician of the result. The cardiologists providing the reference standard were also blinded to these data in order to assess the absolute utility of these variables without bias.

## 7.2 Methods

### 7.2.1 Setting

This study was carried out in a large city-centre emergency department in a university teaching hospital that had around 175,900 emergency attendances per year during the duration of the study (HSCIC, 2011). The department staff levels at the time of this study consisted of ten Consultants, nine middle-grade doctors (Specialist Registrar or equivalent) and two Clinical Research Fellows with part-time clinical commitment.

Ethical approval was obtained through the local National Health Service (NHS) Research Ethics Committee as detailed in Appendix III.

The patients were recruited on a convenience basis between the start of March 2008 and the end of June 2009. Physicians working in the Emergency Department, including the lead researcher, recruited patients. The recruiting physicians were all either middle-grade doctors or consultants.

Roche Diagnostics provided the sample kits for the NT-proBNP testing without cost. Biosite provided the meter required to perform the point of care

testing for BNP without cost (Triage Meter Pro with Census Data Management Software). The appropriate testing kits were purchased from Biosite.

### **7.2.2 Inclusion/Exclusion criteria**

The inclusion criteria for patients presenting to the emergency department consisted of a primary complaint of dyspnoea in patients who were aged 50 years or older at the time of recruitment. It was not necessary for patients to be admitted to hospital to be included.

This age cut-off was chosen as it was considered unlikely for patients below this age to present with decompensated heart failure in the UK setting (Cowie et al., 1999). Younger patients are much less likely to have heart failure and those with heart failure are likely to be known to have a cardiac condition and present less of a diagnostic conundrum. The condition is so uncommon in patients of this age group that the prevalence rate is likely to be below the sensitivity of any feasible clinical decision rule.

Patients were excluded from the study if the attending physician suspected an acute myocardial infarction, if they had renal dysfunction to the point of requiring dialysis, if there was another obvious cause for the dyspnoea such as trauma, or if they were unable to give informed consent.

Patients with renal failure requiring dialysis were excluded as both renal dysfunction and dialysis are known to affect the levels of BNP and NT-proBNP and the measured levels may have been more reflective of the timing of dialysis than any acute disease process (Wahl et al., 2004). There is an important, though incompletely understood, relationship between impaired renal function and acute or decompensated cardiac failure and so a significant proportion of patients with acute heart failure have some degree of renal impairment. Excluding all patients with renal impairment would have

excluded a significant proportion of patients with heart failure and so the assessment of the utility of these markers in real clinical practice would have been compromised. Renal impairment is known to affect NT-proBNP and troponin levels and to a lesser extent BNP and so it was important to quantify this and how it affected the diagnostic utility in this population

The inclusion criteria were chosen in order to select the group of patients where there remained a quandary as to whether the patients' symptoms are due to heart failure. It was not necessary for patients to be admitted to hospital to be considered for inclusion in this study though the majority of patients were not discharged from the Emergency Department.

Table 7-1 Inclusion / Exclusion Criteria

<b>Inclusion Criteria</b>
---------------------------

<ul style="list-style-type: none"> <li>• Patient <math>\geq</math> 50 years of age</li> <li>• Primary complaint of acute dyspnoea</li> </ul>
<b>Exclusion Criteria</b>
<ul style="list-style-type: none"> <li>• Suspected AMI</li> <li>• Previous inclusion in this study</li> <li>• Trauma</li> <li>• Renal impairment requiring dialysis</li> <li>• Obvious cause of dyspnoea that is not heart failure e.g. pneumothorax</li> <li>• Unable to communicate in English</li> </ul>

### 7.2.3 Informed Consent

Informed consent was obtained using two separate consent forms provided in Appendix IV: one form documented consent to participate, the other form was for consent to provide and gift a blood sample. The ethics committee stipulated that two separate forms be used. In the event that the patient was too unwell to give written consent at the time of attendance, permission had been granted by the ethics committee to proceed on the basis of verbal consent until the patient was well enough to give written consent (Coats, 2006). This occurred with two patients who verbally confirmed that they were willing to take part in this research project and completed written consent the day after admission, when more stable.

Patients were presented with an information sheet with contact details for the lead investigator and a letter was sent to the patient's GP informing them that their patient had consented to take part in this trial.

### 7.2.4 Blood Sampling

Two blood samples were taken from each patient on attendance in addition to the blood samples required by the attending physician. Blood samples were taken into Becton, Dickinson and Company Vacutainer® containers. One sample was taken into a tri-potassium EDTA tube and used for the point-of-care BNP test in a Biosite meter within a few minutes being drawn. The other sample was taken into a tube containing a polymer gel and clot activator. This sample left for several minutes until the clot had formed and was then spun using a Heraceus Sepatech Labofuge 200 centrifuge at 3000 rpm for seven minutes. The serum was then drawn off using a pipette and labelled and frozen at -20C. The frozen samples were used for batch analysis for NT-proBNP using the Roche (Elecsys) machine and kit.

The frozen blood samples were also tested for High Sensitivity Troponin T to assess the potential diagnostic utility of this. This test was also performed on the Roche (Elecsys) machine using kits provided by Roche.

All included patients had bloods taken to measure urea and electrolyte levels and a full blood count. Any other blood tests including standard troponin T, C Reactive Protein, D-dimer, INR, glucose and arterial blood gases were performed at the discretion of the attending physician. All included patients underwent an electrocardiogram and had a chest x-ray performed. The majority of patients had a chest x-ray performed in the radiology department but, in patients that the attending physician felt that were too unwell or unstable to be transferred there, a portable chest x-ray was performed in the emergency department. The majority of patients, therefore, had posterior-anterior films while the minority of less stable patients had anterior-posterior films taken within the Emergency Department.

### **7.2.5 Recording of Patient Data**

The patient's demographic details, clinical history, current medications and examination findings were recorded on a standardised sheet by the attending

doctor and then this sheet was included as part of the patients' medical records. A copy of this sheet is included in Appendix V. This form was incorporated as part of the patient's medical clerk-in as well as providing data collection for this study.

The attending doctor was also asked to comment on the results of the chest x-ray, interpret the ECG and provide a clinical impression of whether or not this patient had heart failure. An anonymised copy of the ECG was made and stored with the patient details for presentation to the cardiologists.

The patients were also asked to provide a score from 1 to 10 to represent their dyspnoea when at its worst. At the beginning of the study, patients were also asked to fill in a visual analogue scale consisting of a 10cm line ranging from, 'not breathless' to 'extremely breathless'. The use of the visual analogue scale was abandoned during the study as many patients found the concept difficult to understand, and were unable to complete it. The scoring system was then used alone. The use of a numerical score for dyspnoea has been validated as being correlated to the respiratory rate, oxygen saturation, heart rate and systolic blood pressure in a study by (Saracino et al., 2008). There is also some evidence of a relationship with pulmonary congestion as assessed by ultrasound and natriuretic peptide levels (Weber et al., 2014). Relevant details from patients' medical history, clinical examination findings and the results of investigations carried out within the emergency department were entered into a database produced compiled within SPSS 18 © (SPSS, 2008). Analysis was subsequently performed on this data using SPSS© and other software.

Patient data was checked for up to three years following their original attendance in order to obtain data on re-attendance and subsequent procedures or events. The final discharge diagnosis as recorded in the medical notes was also obtained from the recorded patient discharge letter where available.

### **7.2.6 Echocardiography**

All patients had echocardiography performed within forty-eight hours of their attendance of the emergency department. In the original design of the study the intention was that this should be performed within seventy-two hours of the patient's attendance. This time period was selected on the advice of the cardiologist (RK) who felt that there was unlikely to be significant changes to the cardiac function within this time period. In practice, this was performed within a few hours of attendance in all but two cases when the study was performed the following day.

Echocardiography was performed using a Siemens Acuson X300 ultrasound machine with the P4-2 cardiac transducer and the supplied cardiac software. For each patient the intention was to record short and long axis para-sternal views with M-mode measurements of the left ventricle and valves. Apical four, five and two chamber views were also obtained. Colour Doppler flow was also recorded where possible and pulsed and continuous wave Doppler was used to record blood flow across the mitral and tricuspid valves. This data was presented anonymously to the cardiologists and contributed to the reference-standard.

In almost all of the patients, the lead researcher performed the echocardiogram. In one patient a cardiology registrar had already performed an echocardiogram and recorded his findings in the notes prior to the patient being recruited. Other patients went on to have echocardiography performed by the echocardiogram technicians or cardiologists while in-patients. When this occurred the results were made available for consultants making the final diagnosis in the form of a report.

## **7.2.7 Brain Natriuretic Peptide**

### *7.2.7.1 Biosite Triage Kit*

The Triage® BNP test is a point of care test that uses fluorescence immunoassay for the quantitative measurement of BNP. This takes the form

of a specific cartridge with the test reagents incorporated. Several drops of EDTA anti-coagulated blood were added to the cartridge, using a calibrated pipette, before it was inserted into the triage meter. The concentration of BNP is calculated by the amount of fluorescence. Murine BNP monoclonal antibodies and BNP polyclonal antibodies are labelled with fluorescent dye. The test takes around 15 minutes to produce a result. The meter printed out the result of the assay but this was not viewed until after the patient's management had been decided. The results were not made available to the receiving team if the patient had been referred, or to the attending physician. In patients recruited by the main researcher the results were not examined until all of the appropriate paperwork had been completed and the patient referred or discharged. In one case this led to the failure to obtain a BNP result as the blood sample was discarded before the results were reviewed, and there had been an error with the processing of this sample.

The Triage meter was tested using a quality control device on every day that the machine was used. Specimens are measurable using the Triage meter for concentrations ranging from 5 pg/ml to 5,000 pg/ml (Biosite, 2007). Linear regression analysis of data derived from measuring known concentrations of BNP in healthy volunteers' plasma show that the assay is linear throughout the measurable range.

The lowest detectable concentration distinguishable from zero was determined by using a zero calibrator 20 times using 3 lots of reagent and 5 different meters on 5 days. The average 95% confidence interval was less than 5 pg/ml. Haemoglobin, lipid, triglycerides or bilirubin concentrations were not shown to affect BNP levels. The haematocrit was also varied between 27% and 51% with no significant effect. A wide variety of drugs have been tested at normal therapeutic concentration with no effect on measuring the BNP concentration.

Various proteins and peptides were checked, including angiotensin I, II and III, and C-type natriuretic peptide with no significant interference on measured BNP levels. Average within day and total precision were determined using the

ANOVA model. The CV was between 9 and 12%. A comparison study was performed on EDTA whole blood vs. plasma. The correlation data shows  $r^2 = 0.9878$ ,  $y = 0.925x + 13.439$ .

The manufacturers state that there are no statistically significant changes in BNP concentration associated with hypertension, diabetes or chronic, obstructive pulmonary disease (Maeda et al., 1998).

In this study the lead researcher performed all of the samples analyses. The results were available within fifteen minutes of the sample being entered to the point-of-care machine. The machine provided a result on a digital display and a printout from an integrated printer. The results were collected after the patient had left the department.

### **7.2.8 'N' Terminal pro-Brain Natriuretic Peptide**

The concentration of NT-proBNP was measured using the Roche Elecsys proBNP II assay. This assay contains two monoclonal antibodies that recognise epitopes in the N-terminal part of the proBNP molecule. One of the antibodies is labelled with ruthenium, the other with biotin. The complex then binds to micro-particles coated with streptavidin. The micro-particles are then separated from the solution by using a magnetic electrode. Application of a voltage to the electrode then induces a chemiluminescence that is measured to provide a concentration value.

The measurement range for this assay is 5 to 35000 pg/ml (0.6-4130 pmol/l). The Intermediate Precision coefficient of variation (CV) was 4.6% for a solution of concentration 44.0 pg/ml and 3.8% for a solution with concentration of 33606 pg/ml. There were no specific cross-reactions with various tested compounds including aldosterone, angiotensin II, BNP, endothelin, NT-proANP, renin or urodilatin (Roche, 2010).

The collected samples were stored in a freezer within the laboratory and then analysed as a single batch at the end of the study.

### **7.2.9 High Sensitivity Troponin 'T'**

High sensitivity Troponin T was measured using the Roche Elecsys 2010 analyser. The assay uses two monoclonal antibodies directed against two epitopes on the central portion of human cardiac troponin T. One of the antibodies is attached to ruthenium and the other is bound to biotin. After the initial incubation streptavidin micro-particles are added to form a solid complex. The complexes are separated from the solution using a magnetised electrode. A voltage is passed to the electrode to cause electroluminescence that is measured to calculate the concentration of troponin T.

The range that can be measured using this assay is 3 to 10,000 ng/l. The lowest concentration with a CV less than 10% was 13 pg/l. The manufacturers found no interference on testing with 52 commonly used pharmaceuticals. Falsely low readings can occur in the presence of haemolysis when the haemoglobin level is greater than 0.1g/dl.

The high sensitivity troponin test was performed on the batch of frozen samples after the recruitment phase of this study was concluded.

### **7.2.10 The Reference Standard**

Two independent cardiologists reviewed all of the relevant historical data, clinical findings and investigations including chest x-rays and the recorded echocardiography to provide the reference standard for this study. A study number identified the patients. The ECGs were provided as photocopies of the original ECGs with identifying data obscured. A summary of the attending physician's interpretation of the ECG was also provided. The chest x-rays were provided as high-definition jpeg files along with a simplified reporting of the emergency physician's interpretation and the radiologist report was also

included where this was available. The cardiologists remained blind to the BNP, NT-proBNP and HS-TnT results. The cardiologists decided if the patient had definite, possible or no heart failure.

In the event of disagreement, a third senior physician was available in order to allow a majority decision.

## 7.3 Statistical Methods

### 7.3.1 Sample Size

In creating a diagnostic model, all variables that have been hypothesised on theoretical grounds or have been shown in previous research to be confounders of the relationship being studied should be considered. It has been recommended that researchers should err on the side of including potentially important variables but should not include extraneous ones. Where there is significant overlap in the diagnostic information provided by two variables it is recommended that only one should be selected for inclusion in the model.

Given the limitations of working with a finite number of outcomes within the population studied, some limitation in the number of considered variables is necessary. If the number of variables included in the model synthesis is too large relative to the number of outcomes of interest within the study population then there is a high probability of associations arising by chance; these events would lead to false conclusions and may obscure the true value of other variables. There has to be some selection of variables on the basis of clinical sensibility and demonstrable utility in order to enable the production of a useful model (Concato et al., 1993).

Cepeda (2003) used Monte Carlo simulations on a fabricated study to compare logistic regression with propensity scores. This study looked at comparing logistic regression and propensity scores in terms of bias, precision, empirical coverage probability, empirical power, and robustness

when the numbers of events are low relative to the numbers of confounders. In logistic regression, as the number of events per confounder increased, the magnitude of the bias decreased. Four or less events per confounder gave large bias; seven or more events per confounder gave a level of bias close to zero. The strength of the association of exposure with outcome did not have a clear effect on the amount of bias. Other authors have recommended a ratio of around ten to one with a minimum sample size of one hundred (Peng & al, 2002).

### **7.3.2 Estimation of Sample Size**

Wasson et al. (1985) stated in their landmark paper on the applications and methodological standards of clinical prediction rules, that there should be at least five patients for each predictive finding in the rule. Other authors have used mathematical modelling to suggest that at least seven patients should be used for each decision node in the rule (Laupacis et al., 1997). A decision rule incorporating five predictive variables would therefore require twenty-five to thirty-five patients with the target condition. Assuming a prevalence of heart failure of at least 35% in the recruited population, based on the studies from the systematic review, meant that a minimum sample size of one hundred patients was required.

## **7.4 Selection of Variables**

In order to create a clinical decision rule the first stage was to consider which variables demonstrated diagnostic utility.

The decision of how many variables to include depends on the intended outcome of the research. If the purpose of the research is to explore explanatory roles then all potential variables should be included. As the main reason for the creation of this model is for diagnostic purposes, then only

those variables that enable accurate predictions should be used. In addition to this, a smaller number of included variables makes the decision rule more practical to use and easier to remember.

Although statistical software, such as SPSS ©, has optional automatic variable selection settings there are limitations associated with their application. Methods include forward stepwise selection, backwards deletion, and best subset. For stepwise selection, the variable with the strongest association with the out-come is selected first, followed by the next strongest and so on until all the variables related to the outcome, at a significance level set by the investigator, are entered into the model. Any variable entered in the model that does significantly add to the outcome decision is deleted. For backward deletion, the reverse process occurs and the variable with the weakest association is removed first and so on.

Setting a fixed level of significance for individual variables to allow entry into the model introduces a risk of eliminating variables that are weak when tested individually but significant in combination. Another problem is that if two variables are similar, the model will exclude the one with the poorer statistical characteristics although this variable may have more clinical significance. A third issue is that some investigations, for example chest x-ray, will be performed in almost all patients with this presentation regardless of the diagnostic utility for the condition of heart failure, as findings may be pathognomonic of other conditions; if there is some diagnostic utility for heart failure, even if other variables are stronger, then this should not be disregarded.

It is for these reasons that automated methods were not used in the variable selection and for each of the considered variables the decision was based on performance in the univariate analysis, the clinical relevance and practicality of the model and in consideration of its relationships with the other variables.

Three separate methods were used to shortlist the most useful candidates for a clinical decision rule. The first method was simply selecting the variables

that had proved most useful in the systematic review of the relevant literature and considering the meta-analysis that had been performed. The second method was to use Random Forest® software to produce an ordinal list of diagnostic variables.

The third method was by using univariate logistic regression analysis for each of the potential variables; the Wald test significance was used to select the variables for further analysis. A value of 0.1 was chosen rather than the conventional 0.05 in order to minimise the risk of missing variables with useful diagnostic roles that were not apparent in isolation.

### **7.4.1 Random Forests®**

Random Forests® is software created by Salford Systems. This software provides a robust method for selecting the optimal predictors from a selection of variables for a given, binary target outcome. The software is based on the CART decision tree algorithms and Breiman's 'Bagger' bootstrap aggregation (Breiman et al., 1995). The software works by randomly selecting around two-thirds of the potential cases and then randomly selecting a chosen number of variables. The researcher selects the number of variables to be chosen at random but it is recommended that the initial number should be approximately the square root of the total number of variables. The optimal decision tree using the selected variables and cases is then produced. The intention is to provide a weakly predictive tree. The tree is then assessed to see which of the variables have the greatest predictive value within that tree. This process is then repeated numerous times and the findings regarding which variables have had the greatest predictive value are combined for all of the trees. Due to the double randomisation, all of the trees produced are unique and will tend to be quite different. The combination of information from a large number of different trees has been shown to be more useful than combining data from software systems designed to produce the ultimate unique tree and the problems of over-fitting of the data are avoided. The results have been verified by predicting data from samples from very large

data bases and this method was found to be more reliable than neural networks, solo decision trees, logistic regression or support vector machines.

For nodes where there is a selection of highly significant splitter variables there will be a significant improvement in the model. For nodes where only weak variables are available the best selected will add little to the overall value of the tree. Although the same splitter may be chosen as potential splitters for variable nodes this is unlikely to happen; for example if you have 100 potential variables and the software randomly chooses 10 of these then the chances of the same variable being chosen again is  $.1 * .1 = 0.01$  or one in a hundred.

If a variable with a strong predictive value is selected then it is likely that this will be chosen as the most powerful selector compared with the other nine variables. This should become apparent over multiple repetitions of the analysis model.

Each tree produced will be the optimal tree given the selected variables so no tree will actually produce misleading results, but any of these trees would be expected to perform less well than the optimum tree which would combine the best performing variables from the entire data set.

The different trees are combined through a voting or averaging process. The double random choices in the process mean that the trees will be substantially different from each other. This, paradoxically, actually improves the results of combining the trees. Other software which tries to predict the ultimate tree based on the available data tends to produce very similar trees and so the combination of these trees provides little new information and does not avoid the risks of over-fitting. Combining weaker models actually provides a stronger combined model. As each tree is independent then there is no over-fitting of data. Each tree provides a small amount of useful information towards the complete model. Genuine patterns will be discovered on repeated occasions by various trees while accidental patterns will be 'washed out' by the averaging out of the results (Friedman & Popescu, 2003).

## 7.4.2 Logistic Regression

Logistic regression is a useful way of describing the relationship between one or more independent variables (e.g. age, sex etc.) and a binary response variable expressed as probability; for example, the probability of a patient having heart failure (Peng & al, 2002).

Logistic regression allows the prediction of a discrete outcome from a set of variables that can be continuous, discrete, dichotomous, or a mix of any of these types. Discriminant analysis can only be used with continuous variables. No assumption is made about the distribution of the variables used in the logistic regression model (Chappell & Fine, 2010).

The goal of logistic regression is to correctly predict the category of outcome for the individual case using the most parsimonious model. The system can also be used to explore the strength of relationships of the included variables.

The basic assumption is that for every increase in a unit of the predictor variable the odds of the outcome are multiplied by a factor (the odds ratio of the predictor) and that the effect of several variables is the multiplicative product of their individual effects. This should be confirmed by plotting a graph of the logit of the probability in order to confirm a linear relationship (Anderson, 2005). If the relationship is not close to linear it may be more appropriate to transform the data before applying the logistic regression.

If the binary outcome of an event is plotted against a continuous variable that is associated with the outcome then the chart appears as follows:

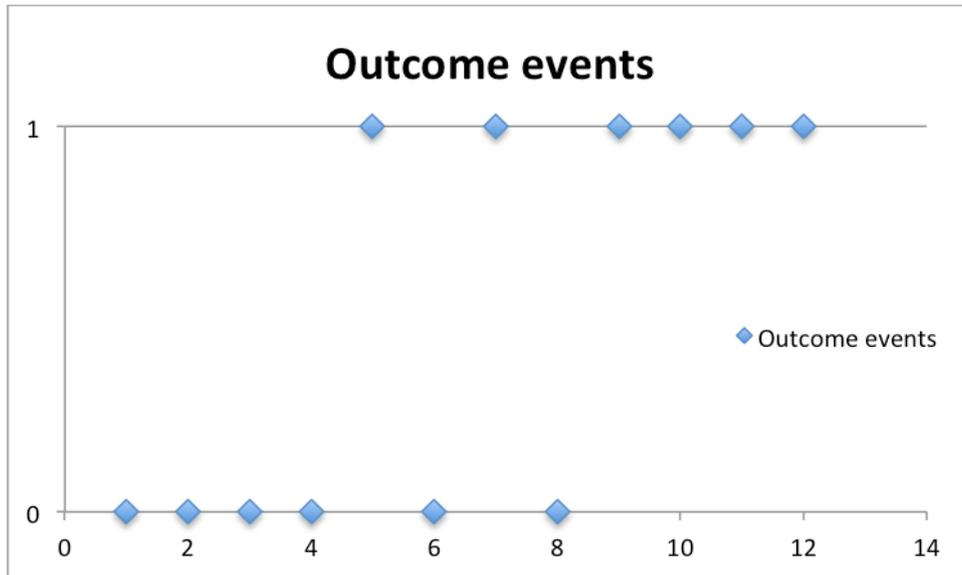


Figure 7-1 Binary outcome associated with a continuous variable

Mapping the probability or odds of a dichotomous outcome against a continuous event produces a 'S' shaped curve.

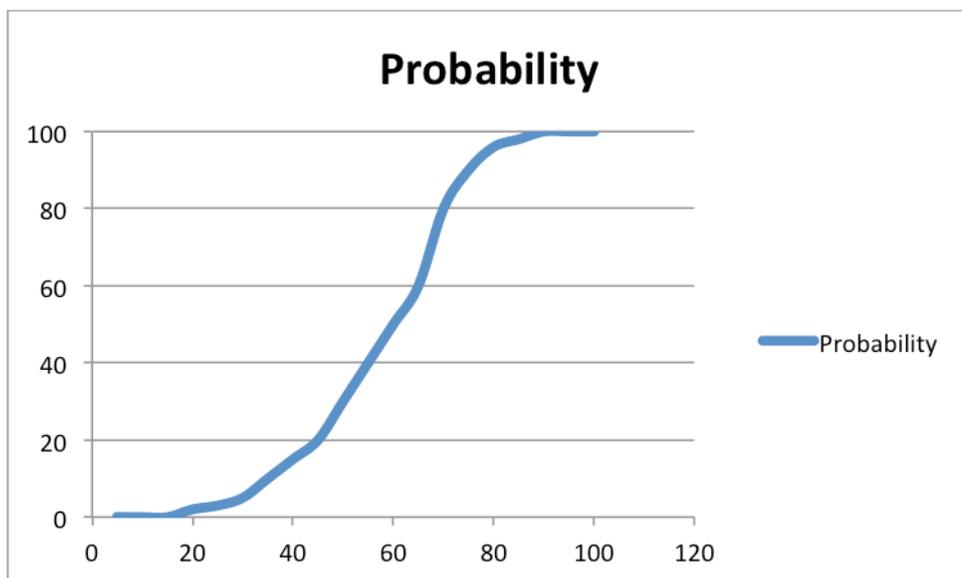


Figure 7-2 Probability of a binary outcome associated with a continuous variable

In order to perform logistic regression, the probabilities are converted to odds, and then the natural logarithm of each value is used to transform the data to resemble a straight line. Obtaining a straight line allows the relationship between the variables to be plotted and the relationship can be described using the simple equation:

$$\ln(b) = ax + c.$$

This process resembles simple linear regression but unlike linear regression the least sum of squares cannot be used to calculate the best fitting line. Instead, for each chosen value of the variable, the probability of the event is calculated. Then the maximal probability value is found given the measured outcome. There is an assumption of binomial distribution of the probability value. The position of the line is calculated using the values of the probabilities of the outcomes that are closest to the maximum probability value. The logit is then converted back into odds by using the exponential function (Bewick et al., 2004).

All of the variables were examined using univariate logistic regression to test for significance and all variables that had a probability value of less than 0.1 were considered further. Continuous variables were examined for correlation, and for pairs of variables, which demonstrated very significant correlation, only one of the variables was chosen. This is because the logistic regression model is constructed with the assumption that the variables are independent.

### **7.4.3 Wald Statistic**

The Wald statistic is a test of the significance of each coefficient in the model. Dividing the coefficient by the standard error of the coefficient and squaring the result calculates the Wald statistic. This statistic has a chi-squared distribution. The significance of the test is calculated by comparing the effect against the null hypothesis. The reliability of the Wald statistic is questionable, especially in smaller samples. If the significance of the Wald statistic is significant i.e. less than 0.05, then the predictor is accepted as being useful and included in the model (Bewick et al., 2004).

## **7.5 Creation of the Decision Rule**

Following the selection of the relevant components of the decision rule, the intention was to combine them in the optimum manner. The following methods were used to create decision rules: multivariable logistic regression; classification and regression tree software (CART); manual decision tree creation; scoring systems and consideration of the individual variables that provided the maximal diagnostic utility in this population.

### **7.5.1 Multivariable logistic regression analysis**

Multivariable analysis is used to determine the unique contributions of various factors to a single event or outcome. Multivariable analysis is necessary for diagnoses such as heart failure where there are numerous risk factors, aetiologies and signs and symptoms associated with the pathological condition. In these circumstances it is impossible to isolate individual variables from surrounding confounding variables in vivo. While it is relatively simple to conduct a trial where patients are randomised to exposure or avoidance of a single pharmaceutical agent, the same process cannot be applied to genetic factors, environmental exposure or lifestyle choices.

While univariate analysis can suggest an association, the relationship between this factor and the disease process can be difficult to tease out in the presence of confounding factors. An example provided by (Katz, 2003) is that males in low-income groups are more likely to smoke and more likely to have cardiovascular disease. The cause and effect relationship is much more difficult to prove than the association. In fact, both poverty and smoking are independent risk factors for cardiovascular disease but are confounding factors for each other.

Although it is relatively easy to stratify an analysis to take account of a small number of variables, for example sex or age, attempts to control for many variables leads to dividing the data set into smaller and smaller subgroups. The use of multiple, small subgroups increase the likelihood of introducing random findings that can become amplified by the analysis.

Multivariable analysis allows for adjustment for confounding factors in order to demonstrate an independent relationship. When using multivariate analysis or stratified analysis is used it is important to remember that adjustment can only be performed for measured variables; results may still be confounded by other known and unknown factors that have not been measured.

### **7.5.2 Limitations of Logistic Regression**

One limitation is that certain variables may be excluded despite a strong predictive value as another variable has a greater predictive power e.g. NT-proBNP may have an inferior predictive power in the presence of BNP. Excluding the inferior variable leads to a less adaptable model in the event that the superior variable result was not available (Steyerberg et al., 2000).

It is also difficult to recognise interactions between variables; for example, there may be a synergistic effect between two variables where the combination of both variables has a much greater effect on the out-come than would be predicted by a simple summation.

It may also be difficult to demonstrate non-conformity to a linear gradient; although the association is recognised, the true relationship between the variable and the outcome may be over simplified by the logistic regression equation.

Logistic regression does not allow for the same outcome to occur to the same person, more than once during the follow up period (Peng & al, 2002). It is possible to use generalised estimating equations to adjust for this but this has not been performed in this study; patients were not included in this study a second time even if they re-attended with the same symptoms. The patients were followed up and re-attendances to this hospital were recorded.

### **7.5.3 Assessing the logistic regression model**

Goodness of fit of the logistic regression model can be calculated using the technique described by (Lemeshow & Hosmer, 1982). The statistic is produced by creating ten ordered groups of subjects and compares the number that are actually in each group, the observed number, with the number predicted by the logistic regression model, the expected number. The test statistic is a chi-square statistic and a good model will be non-significant ( $p > 0.05$ ) as there will be no significant difference between the observed and predicted number values. The ordered groups are selected on the basis of the predicted probability for the selected patients so that the first group will have a predicted probability of the studied outcome of less than ten per cent, and the next group would have a predicted probability of between ten and twenty per cent and so on (Hosmer et al., 1991).

#### **7.5.4 Summary of CART analysis**

The purpose of classification trees is to discriminate between whether or not a categorical outcome is true, based on various parameters. The analyses are performed via tree-building algorithms that determine a set of if-then logical conditions allowing accurate classification of the included cases. The predictor variable data can be nominal, ordinal or scale. A growing method must be selected (Lewis & Stevens, 1991).

CART analysis divides the entire database into two using one variable. The tree examines every eligible predictor for its splitting power (measured by one of several mathematical criteria) and, for non-categorical data, examines each potential value of the variable, to create a database partition using the best performing predictor. The best splitter for a node is found by an exhaustive examination of every eligible splitter at that node and every node is searched in the same way. Splitting is carried out until it is no longer possible due to lack of data or a pre-specified stopping point.

The basic process of CART can be described as four steps:

- Specifying the criteria for predictive accuracy
- Selecting splits
- Determining when to stop splitting
- Selecting the 'right-size' tree

CRT uses surrogates for missing values of independent variables.

Essentially, if a particular variable is missing for a case then the next best predictor is used instead in its place for that particular case. A list of the most commonly used substitute predictors is provided with the production of the tree.

Advantages of CART analysis include the fact that the interpretation of the results is usually simple. This makes it simple to apply and to explain. There is no implicit assumption that the underlying relationships between the predictor variables and the dependent variable are linear or even that they are monotonic. The methods allow for many potential predictors and interactions.

### **7.5.5 Growing Methods**

The production of the decision tree results from the process of recursive partitioning. There are a variety of methods of performing this process; the commonest ones are CHAID, a method used to calculate the optimal predictors in a top-down approach, and the use of the Gini coefficient in calculating the 'purest' nodes (Kass, 1980).

CHAID - Chi-square Automatic Interaction Detection. At each step the independent variable with the strongest interaction with the dependent variable is chosen. Categories of each predictor are merged if they are not significantly different with respect to the dependent variable. The exhaustive form examines all possible splits for each independent variable.

The Gini coefficient is used to measure impurity of the nodes. A terminal node in which all the cases have the same value for the dependent variable is a homogenous 'pure' node. One advantage of this growing method is that the

nodes are split in a binary fashion while the CHAID method may split variables at multiple levels.

#### *7.5.5.1 Classification and Regression Trees*

The same approach can be applied to regression problems where optimal cut-off levels are determined from continuous variables in order to predict outcomes most accurately.

A major issue when applying CART analysis to real data is deciding when to stop splitting. With background random noise due to variation within the sample, classifying every individual in a study will result in over-fitting and the resulting tree is unlikely to be applicable to another sample of the population. The basic principle is to cease splitting when the splits do not add significant improvement to the prediction of the target outcome.

The optimal method is to derive the tree from one set of observations and then check it against a second, independent group of observations. There are techniques to ensure that the created tree is more robust if a validation set of data is not available.

The intention of CART algorithms is to achieve the best possible predictive accuracy. Operationally the most accurate prediction is defined as the prediction with the minimum costs. The notion of costs was developed as a method of generalising the idea that the best prediction has the lowest misclassification rate. In most applications the cost is measured in terms of proportion of misclassified cases or variance. The need for minimising costs, rather than just the proportion of misclassified cases, arises when some predictions that fail are more catastrophic than others. For example, it would be much worse to miss an early diagnosis of lung cancer than over diagnose the incidence of ankle sprains.

### 7.5.5.2 *Selecting Splits*

The second step consists of selecting the splits on the predictor variables that are used to predict membership in classes of the categorical dependent variables or to predict values of the continuous dependent variable. In general, the split at each node will be found that will generate the greatest improvement in predictive accuracy. This is usually measured with some type of node impurity measure that provides an indication of the relative homogeneity of cases in the terminal nodes. If all the cases in each terminal node show identical values then the node impurity is minimal, homogeneity is maximal and prediction is perfect (for the provided sample that has been used in this analysis, but perhaps not for a further independent sample).

There are several impurity measures: The Gini index, Chi-square or G-square. The Gini index of node impurity is the measure most commonly chosen for classification-type problems. As an impurity measure it reaches a value of zero when only one class is present at a node. With priors estimated from class sizes and equal misclassification costs, the Gini measure is computed as the sum of products of all pairs of class proportions for classes present at the node and it reaches maximum value when class sizes at the node are equal (Mingers 1988}. The Chi-square measure is similar to the standard Chi-square value computed for the expected and observed classifications (with priors adjusted for the misclassification cost) and the G-square measure is similar to the maximum-likelihood Chi-square. For regression-type problems, a least-squares deviation criterion is automatically used.

### 7.5.5.3 *Stopping splitting*

Two options are available, a minimum number or a fraction of objects. One method to control splitting is to allow splitting to continue until all terminal nodes are pure or contain no more than a specified number of cases. The other option is to allow splitting to continue until all terminal nodes are pure or

contains no more cases than a specified minimum fraction of the sizes of one or more classes.

## 7.5.6 Naïve Bayes Analysis

### 7.5.6.1 Bayes' Theorem

Bayes' theorem relates to the probability of a given event, and how this is affected by the results of other variables (Price, 1763a). The probability of an event, A, can be estimated how many times this event occurs when a trial is carried out N times. That is:

$$\mathit{Prob} (A) = \frac{NA}{N}$$

For a certain event, probability would equal one, for an impossible event, the probability would be zero. For all other events, as N approaches infinity, the P should approach the true probability of the event result.

If B is a variable that affects the probability of event A occurring then it is possible to study the probability of event A when B is also true.

$$\mathit{Prob} (A|B) = \frac{N(A \cap B)}{NB}$$

Where ' $A \cap B$ ' represents the intersection of the two events.

By dividing both the numerator and denominator by N, the equation includes the probability of each component.

$$\mathit{Prob} (A|B) = \frac{P(A \cap B)}{P(B)}$$

The probability of the intersection, where A and B are both true, can be seen to be equivalent to the probability of A if B is true, multiplied by the probability of B. This is also true for the probability of event B occurring when A is true.

$$P(A|B)P(B) = P(A \cap B) = P(B|A)P(A)$$

This can be rearranged to create the formula derived from Bayes' theorem:

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$$

This can be applied to diagnostic queries on the basis of clinical findings. For example if it is known that the 30% of patients presenting to an Emergency Department with shortness of breath have heart failure, that 20% of patients having a chest x-ray in the department have pulmonary oedema present and that 50% of patients with known heart failure have pulmonary oedema changes on their chest x-ray then it is possible to predict the probability that a patient with pulmonary oedema on chest x-ray will have a final diagnosis of heart failure.

P(A) = probability of pulmonary oedema

P(B) = probability of acute heart failure

P(A|B) = probability of pulmonary oedema in patients with acute heart failure

P(B|A) = probability of heart failure in patient with pulmonary oedema

$$P(B|A) = \frac{0.5 * 0.3}{0.2}$$

Thus, in a patient with acute breathlessness, with pulmonary oedema present on the chest x-ray, the probability of a final diagnosis of heart failure would be 0.75.

Once the data has been collected on a patient population with a given condition it is usually relatively simple to retrospectively calculate the

proportion of patients who had a given symptom or sign at the time of attendance. By calculating how the probability of the target condition is affected by each independent variable and multiplying these factors together it is possible to combine this information to calculate the probability of the target result from the available data. If the data is not available for a particular variable for a patient then the result can still be calculated with the available data.

This process assumes that the included variables are independent, for example, in determining whether or not a fruit was an apple, questions may be asked about the colour and the diameter of the object. In the consideration of the clinical features of a medical diagnosis, the relevant variables are rarely independent; hence the term 'Naïve' or 'Simple' Bayes analysis. However, despite the theoretical limitations of this technique, the results are far more accurate than would be predicted, especially in the derivation of decision rules where there are relatively few cases and many potential variables (Domingos & Pazzani, 1996).

Laplacian correction, named after the French mathematician Pierre-Simon Laplace, who developed Bayes' theory, simply means that 1 is added to each variable (Deakin, 1981). This overcomes the problem that may arise if there were zero incidences of a result. This would provide a probability of zero, which when multiplied with the other probabilities would result in an overall value of zero no matter how likely the other variables were. By adding one to each variable the effect on the likelihood will not be affected by much but the risk of the zero probability will not arise.

Using Naïve Bayes analysis has advantages over decision tree analysis as all of the included variables are used to inform the final diagnosis. The full spectrum of the variable can also be used in the decision process rather than having to choose a cut-off value to dichotomise the results into positive or negative. Also, the decision process works well with missing variables; the same decision tool could also contain alternative variables such as BNP and

NT-proBNP. The decision tool also allows a probability or likelihood ratio to be calculated rather than a simple Boolean result.

Možina et al. (2004) have developed a nomogram for use with Bayes' method. This allows visualisation of how the included variables contribute towards the end diagnosis. It also allows the application of simple decision tools to be used without the need for a calculator or computer. Applying a straight edge to the variable on the nomogram gives a value from zero to one hundred, adding the results for each available variable gives a total that has a corresponding probability value. An example is provided in Figure 7-3.

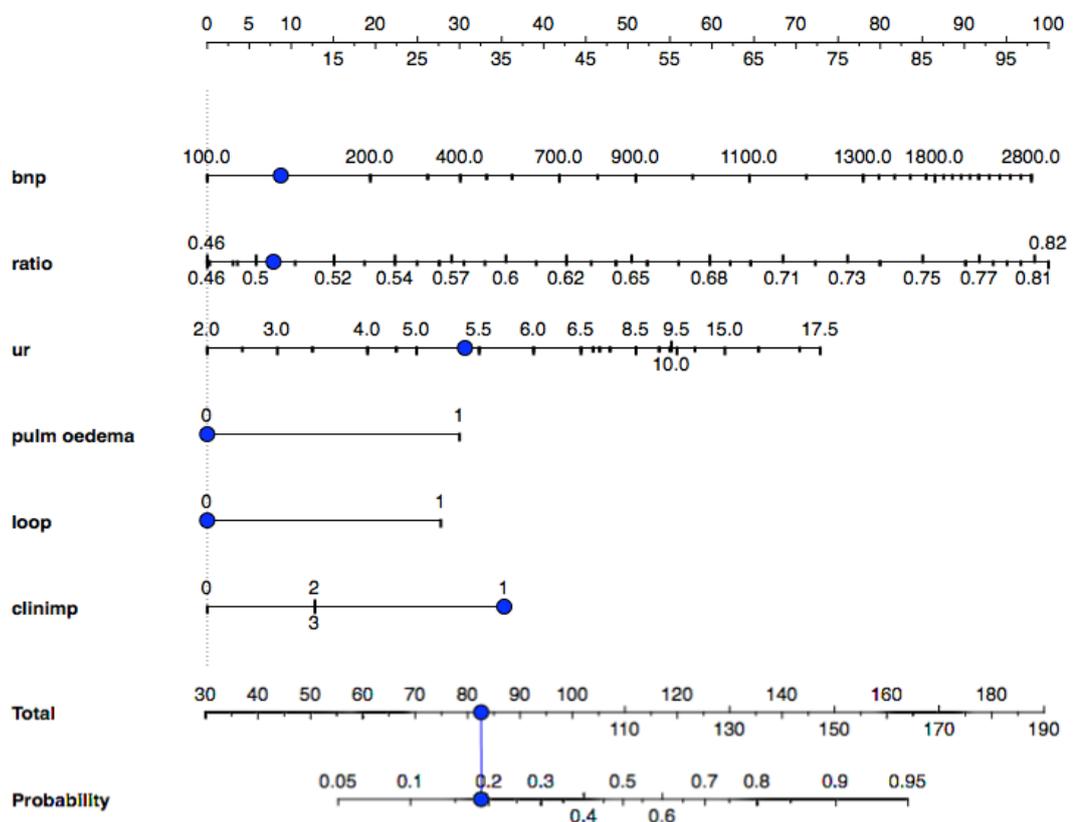


Figure 7-3 Example Using a Nomogram for Naive Bayes' Analysis

#### 7.5.6.2 The Standard Error

The standard error (SE) is a measure of how close the sample mean is likely to be to the true population mean while the standard deviation is a description of how widely scattered value in a sample are. For larger samples the SE can be used to estimate the 95% confidence intervals, as the SE will approximately equal the standard deviation (SD). The sampling distribution of the mean has the SD which is the square root of the SD (squared) /n where n is the number in the sample. A 95% confidence interval for a large sample can be calculated as the mean  $\pm 1.96*SE$

### 7.5.6.3 The Chi-Square Test

The Chi-squared statistic is calculated using a two-by-two table. The observed outcomes are compared with expected outcomes. For example:

	Death	No death	Totals
Treatment	5	45	50
No treatment	20	30	50

While the expected table, if the chosen treatment had no effect, would look like this:

	Death	No death	Total
Treat	12.5	37.5	50
No treat	12.5	37.5	50

The value of the expected result is subtracted from the observed result, the difference is squared and then this result is divided by the expected value. The squaring of the results means that positive and negative differences are considered equivalently and also that larger differences are amplified while small differences are less significant. When this is performed for all the cells in a table a Chi-square value is obtained. This can be read from a table of distribution (or software used) to give an estimate of the probability of this event depending on the number of degrees of freedom present.

If  $Z$  is a standard Normal variable with a mean of 0 and a standard deviation and variance of 1 then the value formed by  $Z^2$  follows the Chi-squared distribution with 1 degree of freedom. As the number of variables increases the distribution tends to Normal. The expected value of  $Z$  squared is the variance of  $Z$ . The expected value of  $Z$  is zero so  $E(Z^2)=1$ . The expected value of chi squared with  $n$  degrees of freedom is thus  $n$ . The variance is  $VAR(\text{chi-squared})=2n$ . The square root of the chi-square is the square root of  $n-0.5$  and has variance 0.5. If the data is from a Normal sample then the sample mean will follow a normal distribution and the sample variance will be from a chi-squared distribution times the sigma squared over  $(n-1)$ . If the value of  $n$  is greater than twenty then the distribution of the sample standard deviation is approximately normal (Bland, 2000).

The strength of an association between two qualitative variables can be difficult to measure, but it is easy to test the null hypothesis, i.e. that there is no relationship between them. This is done using the Chi-square test for larger samples. The chi-square test does not test the strength of an association, it merely indicates if the probability that the null hypothesis is true (Bland, 2000).

#### 7.5.6.4 Variance

The variance is the average squared difference from the mean.

E.g. For the number of heads after two tossed coins:

0 is 1 unit from the mean and happens a quarter of the time.

1 is 0 units from the mean and happens half the time.

2 is 1 unit from the mean and happens a quarter of the time.

$$\begin{aligned}\text{Variance} &= (0-1)(0-1)*1/4 + (1-1)(1-1)*1/2 + (2-1)(2-1)*1/4 \\ &= 1/4 + 1/4 \\ &= 0.5\end{aligned}$$

The square root of the variance is the standard deviation (sigma).

#### *7.5.6.5 Normal Distribution*

Standard Normal distribution has a mean of 0 and a standard deviation of 1. Tables provide values for the area under the curve. For systems where a random variable is derived by a factorial equation then the distribution tends not to be Normal. This is often the case for biological systems where the variable is dependent on several steps e.g. for production of a substance via several enzyme steps. Usually in these situations log transformation will provide Normal distribution. The random variable is then described as having Lognormal Distribution. The data can be transformed using logarithms to perform calculations and then transformed back.

There are a few ways of checking if a variable has normal distribution. One of them is to use a statistics program to construct a 'q-q' plot. This involves placing all of the variable results in order of size and then using approximation to calculate the probability that a value will be less than z for the ith observation (of n) z.

If Standard Normal points are plotted against the observed values there should be a straight line if the variable has Normal distribution, otherwise will obtain a curve.

#### *7.5.6.6 Recursive Partitioning*

In recursive partitioning, a question is asked of a data set which divides the data into two groups (partitioned) and then each of the 'children' groups is asked a new question and so on (recursive). The questions are decided by calculating what would optimise the purity of the groups. There also needs to be some stopping criteria. This may be a specified branch level, purity of the nodes or the minimum number of cases in the terminal nodes.

#### *7.5.6.7 Sampling distributions*

If there is a population with  $n$  of around 100 then it is possible to estimate values by taking ten samples of 4 from the 100, at random. This will give an approximate mean but the standard deviation of the estimate is small compared with the actual population standard deviation. The standard deviation of the sample is called the Standard Error (SE) of the estimate. The true mean is likely to be within one SE of the sample mean and is unlikely to be more than two SE away from it. With large samples the estimates are likely to be accurate predictions of the population values.

### **7.5.7 Summary**

This chapter has introduced the methods used for the prospective diagnostic study with a focus on the novel investigations used. The methods used in the analysis of the results and the intended techniques for derivation of the clinical decision rule have also been explored.

## Chapter 8 Results and Analysis

### 8.1 Patient Characteristics and Classification

One hundred and five patients with heart failure were recruited, as a convenience sample, between the start of March 2008 and the end of June 2009. All of the patients presented to a single Emergency Department with acute shortness of breath as their main presenting complaint. Forty-three patients (41%) were male and sixty-two were (59%) female.

The lead researcher recruited 48 of the patients, 2 were recruited by Emergency Department (ED) consultants, 37 by ED registrars, 9 by ED senior house officers and 9 by the receiving medical officers.

The mean age of the patients was 72 years, the age range of the patients varied from 52 years to 93 years. Three patients declined to participate in the study and no further data was recorded regarding these patients.

The mean duration of symptoms prior to attendance was 6.25 days; the median duration was 3.12 days. The range of this period extended from one hour to six weeks. Only nineteen patients (18.1%) were aware of having a prior diagnosis of heart failure although forty-five patients (42.9%) were receiving regular diuretics. The patient characteristics and final diagnostic classification as defined by the study reference standard are provided in Table 8-3.

Table 8-1 - Patient Characteristics and Classification

	All Patients	Definite Heart Failure	Possible Heart Failure	Not Heart Failure
Total Number	105	31	28	46
Male (%)	43 (41%)	13 (41.9%)	14 (48.3%)	16 (35.6%)
Mean Age	71.8 years	74.3 years	73.5 years	69.0 years
Median Age	72 years	76 years	72 years	69 years
Standard Deviation	10.5	9.0	8.9	11.8
Range	52 – 93 years	56-88 years	56-89 years	52 – 93 years
Admitted	98 (93.3%)	30 (96.8%)	26 (89.7%)	42 (93.3% )
Length of stay (LOS)				
Mean LOS	12.4 days	11.6 days	16.3 days	10.3 days
Median LOS	8 days	11 days	8 days	8 days
Standard Deviation LOS	14.8	8.6	21.1	13.1
Range LOS	0 – 76 days	0 – 35 days	0 – 76 days	0 – 59 days
Died during admission	6	4	1	1
Died within 30 days of admission	6	4	0	2
Died within 1 year of admission	11	6	2	3
Re-attended within 30 days	25	7	6	12
Previous diagnosis of heart failure (%)	19 (18.0%)	13 (41.9%)	4 (13.7%)	2 (4.4%)

Figure 8-1 shows the breakdown of the patient population by reference standard provided by the reviewing cardiologists with the discharge diagnoses provided to the patient's general practitioner. This diagnosis was taken from the discharge letter written at the time of the patient's discharge from hospital. Normal practice in this institution is for the discharge letter to be written by a junior doctor. In twenty-four cases, either no discharge letter was written or no specific discharge diagnosis was made within the letter.

Attendance to hospital with acute dyspnoea was associated with high admission rates and significant lengths of stay. There was also a significant mortality rate, most marked in the patients who were considered to have definite heart failure. Almost a quarter of the included patients re-attended this hospital within thirty days of their initial attendance.

All patients had a primary complaint of acute dyspnoea in order to be included in the study but other respiratory symptoms were recorded. These symptoms are provided in Table 8-2 with a breakdown by reference standard diagnosis.

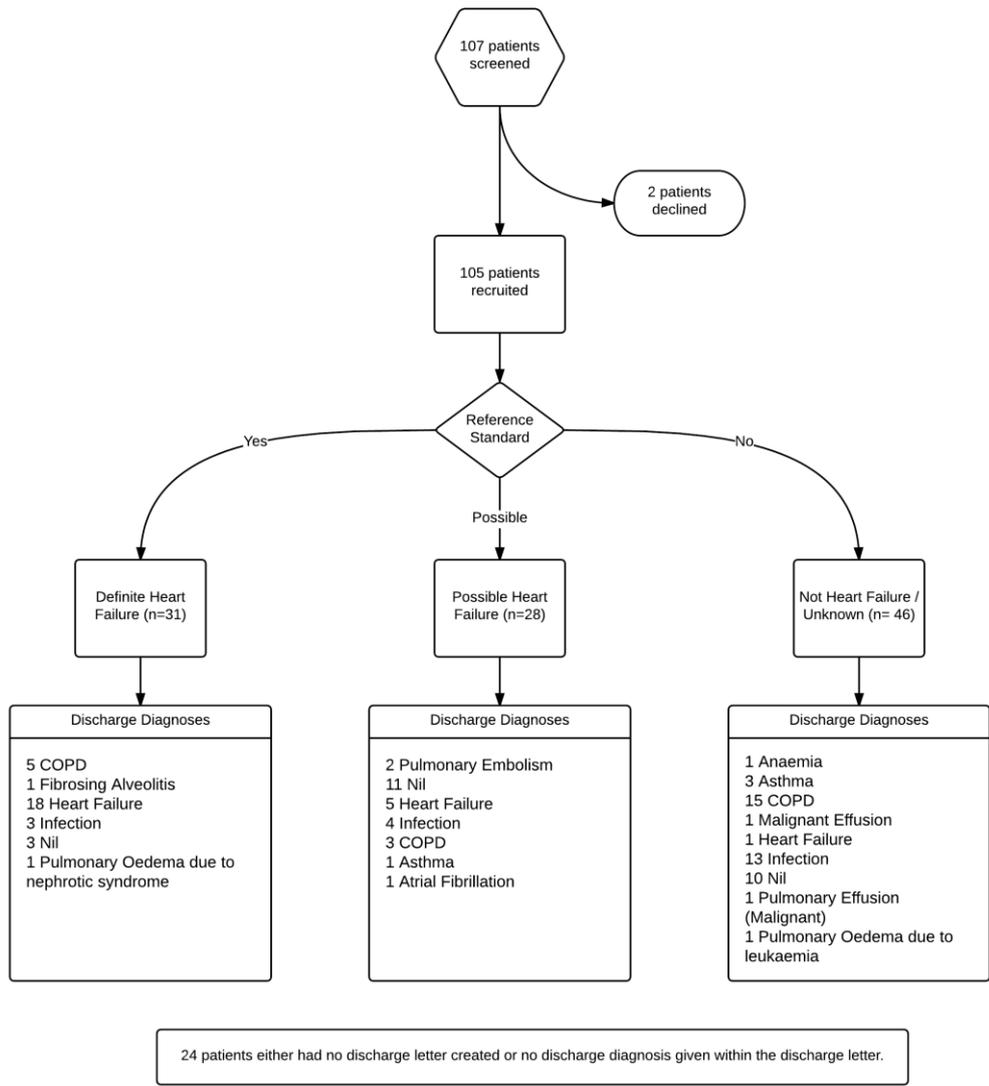


Figure 8-1 Reference Standard and Discharge Diagnoses

Table 8-2 Reported Symptoms

Symptoms	All Patients	Definite Heart Failure	Possible Heart Failure	Not Heart Failure
Chest pain	36 (34.3%)	12 (38.7%)	5 (17.2%)	18 (40.0%)
Cough	61 (58.1%)	11 (35.5%)	18 (62.1%)	32 (71.1%)
Orthopnoea	52 (49.5%)	17 (54.8%)	15 (51.7%)	20 (44.4%)
PND	41 (39%)	17 (54.8%)	13 (44.8%)	11 (24.4%)
Wheeze	61 (58.5%)	12 (38.7%)	21 (72.4%)	28 (62.2%)
Productive cough	26 (24.7%)	3 (9.7%)	6 (20.7%)	17 (37.8%)

The historical features of the patient were those provided to the physician at the time of attendance rather than from other sources such as medical records. These are recorded in Table 8-3.

Table 8-3 – Patient Historical Features

Historical features	All Patients	Definite Heart Failure	Possible Heart Failure	Not Heart Failure
Previous heart failure	19 (18.1%)	13 (41.9%)	4 (13.8%)	2 (4.4%)
Angina	20 (19%)	5 (16.1%)	6 (20.7%)	9 (20.0%)
Myocardial infarction	17 (16.2%)	8 (25.8%)	4 (13.8%)	5 (11.1%)
Diabetes	25 (23.8%)	6 (19.4%)	10 (34.5%)	9 (20.0%)
Hypertension	28 (26.7%)	12 (38.7%)	8 (27.6%)	8 (17.8%)
Smoker (current or ex)	70 (66.7%)	18 (58.1%)	21 (72.4%)	31 (68.9%)
Current Smoker	30 (28.6%)	8 (25.8%)	8 (27.6%)	14 (31.1%)
Chronic obstructive pulmonary disease	44 (41.9%)	8 (25.8%)	9 (31.0%)	27 (60.0%)
Cancer	17 (16.2%)	5 (16.1%)	5 (17.2%)	7 (15.6%)
Asthma	23 (21.9)	5 (16.1%)	8 (27.6%)	10 (22.2%)
Raised cholesterol	71 (67.6%)	18 (58.1%)	19 (65.5%)	34 (75.6%)
Previous stroke	9 (8.6%)	2 (6.5%)	4 (13.8%)	3 (6.7%)
Pacemaker fitted	5 (4.8%)	4 (12.9%)	1 (3.4%)	0

Cardiac Resynchroniser Therapy	5 (4.8%)	1 (3.2%)	1 (3.4%)	3 (6.7%)
ICD fitted	0	0	0	0
CABG	12 (11.4%)	6 (19.4%)	4 (13.8%)	2 (4.4%)

All of the patient's current medications were recorded at the time of recruitment and any medications that were considered to be relevant to respiratory or cardiac disease were included as diagnostic variables. These medications are listed in Table 8-4.

Table 8-4 Current Medications

Medications	All Patients	Definite Heart Failure	Possible Heart Failure	Not Heart Failure
Home oxygen	8 (7.6%)	3 (9.7%)	2 (6.8%)	3 (6.7%)
Loop Diuretics	37 (35.2%)	18 (58.1%)	11 (37.9%)	8 (17.8%)
Inhalers	58 (46.4%)	15 (48.4%)	15 (51.7%)	28 (62.2%)
Aspirin	25 (23.8%)	7 (22.6)	9 (31.0%)	9 (20.0%)
Clopidogrel	8 (7.6%)	3 (9.7%)	2 (6.9%)	3 (6.7%)
Beta-blockers	22 (21.0%)	11 (35.5%)	7 (24.1%)	4 (8.9%)
ARB	14 (13.3%)	7 (22.6%)	5 (17.2%)	2 (4.4%)
ACE Inhibitors	25 (23.8%)	10 (32.3%)	5 (17.2%)	10 (22.2%)
Calcium antagonists	22 (21.0%)	6 (19.4%)	7 (24.1%)	9 (20.0%)
Nicorandil	3 (2.9%)	1 (3.2%)	1 (3.4%)	1 (2.2%)
Digoxin	5 (4.8%)	3 (9.7%)	1 (3.4%)	1 (2.2%)
Warfarin	15 (14.3%)	9 (29.0%)	2 (6.9%)	4 (8.9%)
Nitrates	15 (14.3%)	5 (16.1%)	5 (17.2%)	5 (11.1%)
Proton-pump inhibitor	32 (30.5%)	12 (38.7%)	9 (31.0%)	11 (24.4%)
Statin	46 (43.8%)	18 (58.0%)	9 (31.0%)	19 (42.2%)

The patients' observations as recorded on arrival to the department and clinical signs picked up during the clinical examination are recorded in Table 8-5.

Table 8-5 Clinical Signs and Observations

Signs	All Patients	Definite Heart Failure	Possible Heart Failure	Not Heart Failure
Raised JVP	14 (13.3%)	9 (29.0%)	2 (6.9%)	3 (6.7%)
Peripheral Oedema	15 (14.3%)	9 (29.0%)	5 (17.2%)	1 (2.2%)
Temperature (Mean)	36.7 (Range 33.8 - 39.1, SD .92)	36.7 (33.8-38.7, SD $\pm$ 1.0)	36.4 (35.3 – 38.6, sd .64)	36.9 (34.0 -39.1, SD .94)
Normal heart sounds	90 (85.7%)	23 (74.2%)	25 (86.2%)	42 (93.3%)
Systolic Blood pressure (Mean)	138.8 (n=102, 80 – 204, SD 26.0)	144.3 (92-204, SD 30.1)	142.3 (96 – 201, sd 25.9)	133.0 (80 – 183, SD 22.3)
Cyanosis	3 (2.9%)	1 (3.2%)	1 (3.4%)	1
Respiratory rate (Mean)	23.0 (Range 12 – 40, SD 6.47)	24.6 (14-40, SD 7)	22.6 (16 – 40, sd 5.85)	22 (12 – 36, SD 6.34)
Crepitations	53 (50.4%)	19 (61.3%)	15 (51.7%)	19 (42.2%)
Heart Rate (Mean)	99.2 (28 – 197, sd 26.1)	98.5 (55-152, SD 27.5)	96.5 (56 – 140, sd 20.8)	101.5 (28 – 197, 28.3)

All of the relevant investigations for each patient were recorded. The results are presented as total number with a positive result for dichotomous variables

or as the mean result for the group for continuous variables. The results are presented in Table 8-6.

Table 8-6 Results of Investigations

Investigations	All Patients	Definite Heart Failure	Possible Heart Failure	Not Heart Failure
BNP (Mean)	386.591 (5.0 – 3450, SD 628.2)	843.8(Range 14.2 – 3450, SD 853.5)	311.1 (6.3 – 2790, SD 529.7)	115.1 (5.0 – 721, SD 156.4)
NT-proBNP (Mean)	464.3 (Range 1.1 – 5811, SD 993.2)	1145.9 (Range 11 – 5811, SD 1581.3)	276.9 (5.1 – 1292, SD 373.8)	107.6 (Range 1.19 -1061, SD 196.7)
Cardio-thoracic Ratio (Mean)	0.574 (Min 0.36, max 0.94, SD 0.1)	.6328 (Min .37, max .82, SD .0959)	.601 (min .41 - .94, SD .101)	.520 ( .355 - .693, SD .077)
Pulmonary oedema	14 (13.3%)	10 (32.3%)	3 (10.3%)	1 (2.2%)
Upper lobe diversion present	21 (20%)	13 (41.9%)	5 (17.2%)	3 (6.7%)
HS-Troponin	55.30 (3.0 - 757.4,SD 109.8)	90.8 (10.5- 757.4, SD 153.8)	50.8 (3.0 – 469.1, SD 91.1)	33.3 (3 - 395.7, SD 74.4)
Pulmonary oedema present on CXR	14 (13.3%)	7 (22.6%)	5 (17.2%)	2 (4.4%)
D-dimer (Mean)	665.2 (n=25, Range 5- 5740, SD 1281.8)	320.0 (n=7, Range 185 – 760, SD 200.5)	952.5 (n=8, 5 – 5470, SD 1942.1)	677 (n=10, range 115 – 3820, SD 1120.3)
Lactate (Mean)	1.71 (n=55, Range .6 -9.3, SD 1.58)	2.18 (n=14, range .7 - 9.3, SD 2.2)	1.44 (n=15, .7 - 4.7 SD 1.0)	1.62 (n=26, .6 – 7.5, 1.45)

C-Reactive protein (Mean)	47.9 (n=95, Range 3 – 365, SD 74.49)	44.1 (n=26, range 3 – 365, SD± 83.3)	37.8 (n=26, 3 – 248, SD 63.5)	55.8 (n=43, 3 – 324, SD 75.4)
Urea (Mean)	7.67 (n=105, range 1.9 – 25.9, SD 3.78)	9.8 (5.3 – 25.9, SD±4.9)	7.53 (3.1 – 15.9, SD 2.91)	6.3 (1.9 – 13.3, SD 2.65)
WCC (Mean)	10.9 (n=103, range 1.1 – 28.3, SD 4.85)	10.66 (4.5 – 23.6, SD±5.06)	10.4 (3.8 – 28.3, SD 4.93)	11.4 (1.1 – 23.7, SD 4.71)
Est. GFR (Mean)	64.3 (n=103, range 15 – 91, SD 21.33)	50.87 (15 – 90, SD± 19.21)	66.5 (19 – 90, SD 21.9)	72.25 (18 – 91 SD 18.05)
Creatinine (Mean)	98.5 (n=103, range 39 -351, SD 51.0)	122.71 (63 - 351, SD 61.393)	96.8 (39 – 226, SD 45.0)	82.7 (39 – 297, SD 39.9)

Table 8-7 compares the opinion of the attending clinician with the reference standard provided by the cardiologists.

Table 8-7 Clinical Opinion

	All Patients	Definite Heart Failure	Possible Heart Failure	Not heart failure
Clinician Impression Definite HF	24 (22.9%)	16 (51.6%)	7 (24.1%)	1 (2.2%)
Clinician Impression Possible HF	33 (31.4%)	10 (32.3%)	11 (37.9%)	12 (26.7%)

### 8.1.1 Echocardiograms

As discussed in Chapter 2, there were limitations in the success of the application of this investigation in the Emergency Department environment. Despite these constraints, some idea of whether left ventricular function was completely normal or was impaired to some degree, was obtained in the majority of patients. In some cases the results were not interpretable due to the mentioned factors.

In this study the echocardiograms were electronically stored and then reviewed by two cardiologists. The quality of the echocardiograms were considered good in 28 patients (26.6%), adequate in 31 patients (29.5%), suboptimal in 34 patients (32.3%) and missing or interpretable in 12 patients (11.4%). Even in the echocardiograms where the view was considered suboptimal, it was possible for the cardiologists to comment on their impression of global left ventricular function

## 8.2 Examination of Data

### 8.2.1 Random Forest Variables

#### *8.2.1.1 Random Forests Results*

This was performed using a selection of 9, 5 and 18 variables from a potential selection of 89 variables.

The output from the software shows a running relative error rate for the model in Figure 8-2. When the class weights are balanced the overall error rate is a simple average of the error rates in each class. The parallel coordinate graph shows the twenty-five records with the highest likelihood of a positive result and the twenty-five with the highest probability of a negative result. Eligible predictors can be selected or excluded to see how the model is affected.

The error rate is initially high then gradually improves as more and more models are added to the forest. Eventually the error rate plateaus as the model cannot be improved any further.

Gini is the measure of how much the split has improved the purity of the two resulting classes. The scores for all the variables are cumulated for all the generated trees and then standardised by giving the best variable a score of 100; the actual numbers attributed only provide information about that model and so cannot be used to compare different models.

The Standard method of validating the results involves 'scrambling' a variable cut-off value and then seeing how this affects the model. If a model is not affected much then the particular variable is obviously not that important. This is, again, recoded to have a value between zero and one hundred.

#### *8.2.1.2 Cross Validation*

The sample is divided into numerous subsamples or folds. Tree models are generated excluding the data from each subsample in turn. The first tree is based on all the cases except those in the first sample fold, the second tree is based on all the cases except those in the second sample fold and so on. For each tree the misclassification risk is estimated by applying the tree to the subsample excluded in its generation. A positive number, valued between 2 to 25 is used and the higher the number, the fewer cases are excluded for each tree model. Cross validation produces a single final tree model. The cross-validated risk estimate for the final tree is calculated as the average of the risks for all the trees.

Random Forest analysis was performed using the Salford Systems CART analysis software (Breiman et al., 1995). All potential variables were entered into the equation and the variables were ranked in order of importance by Gini analysis. Continuous variables tend to be ranked more highly than categorical variables, as there is the potential to use different cut-offs in order to provide the optimum split of the selected group data.

One thousand RandomForest trees were created using nine potential predictor variables at each node and with a minimum of two cases per parent node. The following variables were ranked in terms of importance as shown in Table 8-8.

Table 8-8 Variables Ranked by RandomForest

Variable	Score
Cardiac Size	100
Cardio-thoracic Ratio	59.80
Urea	49.36
NT-proBNP	40.87
BNP	36.54
Clinical Heart Failure	16.56
Creatinine	13.62
Troponin	11.83
Age	11.58
Diastolic Blood Pressure	9.13
Estimated GFR	7.91
Platelet Count	7.75
Potassium	7.02
Ankle Oedema	6.87
Loop Diuretic	6.73
Dyspnoea Score	5.74
Upper Lobe Diversion	5.55
Productive Cough	3.39
CRP	3.90
Cough	3.39
Pulmonary Oedema	3.25
History of Heart Failure	2.71
Respiratory Rate	2.60
pH	2.35

Systolic Blood Pressure	2.18
-------------------------	------

The selected variables have considerable overlap with the variables that demonstrated significance on univariate analysis. This analysis identifies the variables that tend to have the greatest significance in the majority of the decision trees created but does not compile a single summary decision tree and so the variables selected cannot be assumed to be independent of each other.

The model provides a graphic representation of the overall error rate as shown in Figure 8-2. The summary model had an overall sensitivity of 95% with a specificity of 48%.

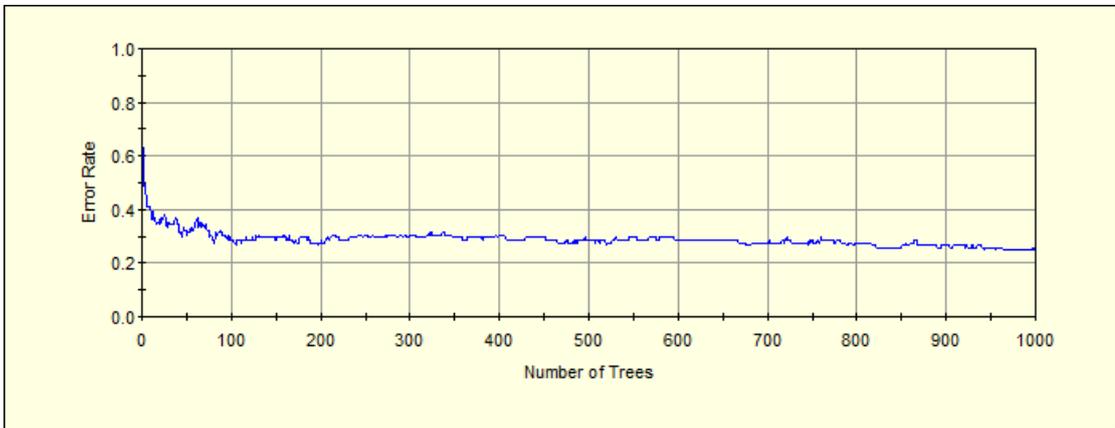


Figure 8-2 - Overall Error Rate for Random Forest Model

### 8.2.2 Univariate Analysis

The patient data was entered into the SPSS software to allow further analysis (SPSS, 2008). The diagnostic importance of each of the variables was examined using univariate analysis against the provided clinical reference of possible or definite heart failure. All of the data with a Wald significance value of less than or equal to 0.1 were examined further, these variables are highlighted in the table for clarity. The results are displayed in Table 8-9.

Table 8-9 - Results of Univariate Analysis

Variable	B	S.E.	Wald	Sig.	Exp(B)	95% CI for Exp(B)	
						Lower	Upper
Sex	.394	.405	.945	.331	1.483	.670	3.282
Cardiac size (xr)	.008	.004	3.571	.059	1.008	1.000	1.016
Cardio-thoracic ratio	13.356	3.185	17.581	.000	631534.500	1227.583	3.249E8
Age (years)	.047	.020	5.570	.018	1.048	1.008	1.090
Duration symptoms	.000	.000	.069	.793	1.000	1.000	1.000
Chest pain	-.030	.265	.013	.910	.970	.578	1.630
Orthopnoea	.278	.199	1.947	.163	1.320	.894	1.951
Hx diabetes	.165	.233	.500	.479	1.179	.747	1.861
Home oxygen	.107	.357	.089	.766	1.112	.552	2.240
Loop diuretic	1.504	.471	10.202	.001	4.5	1.788	11.325
On diuretics	.941	.356	7.007	.008	2.563	1.277	5.144
On inhalers	NS						
PND	1.128	0.432	6.810	.009	3.091	1.324	7.214
Wheeze	-.298	.402	.550	.458	.742	.337	1.633
Hx of HF	2.140	.778	7.566	.006	8.500	1.850	39.005
Hx of MI	.693	.574	1.460	.227	2.00	.650	6.157
Hx of Angina	.108	.500	.046	.830	1.114	.418	2.968
Hx of CABG	1.459	.802	3.307	.069	4.300	.893	20.709
Hx of COPD	-1.333	.418	10.180	.001	.264	.116	8.607
Hx CVA	.442	.736	.360	.549	1.556	.367	6.588
Hx of asthma	-.032	.476	.005	.946	.968	.381	2.462
Hx HTN	.838	.476	3.096	.079	2.312	.909	5.884
Clopidogrel	.248	.760	.107	.744	1.282	.289	5.686
Aspirin	.417	.476	.768	.381	1.518	.597	3.858
Beta-blocker	1.529	.597	6.567	.010	4.615	1.433	14.867

AR Blockers	1.699	.794	4.581	.032	5.467	1.154	25.901
Ca Antagonist	.139	.489	.081	.776	1.149	.440	2.997
Digoxin	21.036	1797 4.847	.000	.999	1.367E9	.000	.
Nitrates	.237	.362	.427	.513	1.267	.623	2.575
PPI	.497	.446	1.244	.265	1.644	.686	3.937
Warfarin	.872	.623	1.961	.161	2.391	.706	8.102
ACE inhib	.044	.134	.109	.741	1.045	.803	1.360
Statin	.257	.314	.666	.414	1.293	.698	2.394
Inhaler	-.190	.158	1.440	.230	.827	.607	1.128
Smoker	-.192	.171	1.271	.259	.825	.591	1.153
Alcohol intake	-.108	.290	.140	.709	.897	.508	1.585
Temperature	-.417	.230	3.293	.070	.659	.420	1.034
O2 Sats in air	-.077	.053	2.062	.151	.926	.834	1.028
Pulse	-.006	.008	.604	.437	.994	.979	1.009
Systolic BP	.016	.008	3.804	.051	1.016	1.000	1.033
Diastolic BP	.028	.012	5.287	.021	1.029	1.004	1.054
Fine Creps	.981	.456	4.618	.032	2.667	1.090	6.524
Raised JVP	1.145	.684	2.799	.094	3.143	.822	12.020
Cyanosis	.417	1.241	.133	.737	1.517	.133	17.273
Ankle oedema	2.595	1.056	6.033	.014	13.391	1.689	106.169
Respiratory rate	.040	.032	1.576	.209	1.041	.978	1.108
Q waves ECG	.537	.412	1.694	.193	1.710	.762	3.838
ECG rhythm	.065	.083	.623	.430	1.068	.907	1.256
Dyspnoea score	-.019	.097	.039	.844	.981	.811	1.187
Sodium	-.002	.041	.003	.958	.998	.921	1.081
Potassium	.714	.375	3.621	.057	2.043	.979	4.265
Urea	.257	.084	9.482	.002	1.294	1.098	1.524
Creatinine	.016	.006	6.478	.011	1.016	1.004	1.029
Est. GFR	-.034	.011	9.799	.002	.967	.946	.987
Hb	.005	.094	.003	.954	1.005	.836	1.209
WCC	-.038	.042	.852	.356	.962	.887	1.044

Platelets	-.004	.002	4.394	.036	.996	.992	1.000
Glucose	-.010	.043	.057	.811	.990	.910	1.077
D-dimer	.000	.000	.001	.969	1.000	.999	1.001
CRP	-.003	.003	.870	.351	.997	.992	1.003
PO2	-.024	.028	.697	.404	.977	.924	1.032
pCO2*	.338	.196	2.983	.084	1.402	.955	2.057
Bicarbonate	-0.29	.077	.147	.701	.971	.835	1.129
pH	-8.727	4.281	4.154	.042	.000	.000	.715
Base XS	.004	.061	.004	.948	1.004	.891	1.132
Lactate	.076	.179	.181	.670	1.079	.760	1.532
BNP	.004	.001	11.068	.001	1.004	1.002	1.006
HS-TnT	.005	.003	2.422	.120	1.005	.999	1.011
NT-proBNP	.003	.001	9.417	.002	1.003	1.001	1.006
Age (years)	.047	.020	5.570	.018	1.048	1.008	1.090
Prod cough	1.236	.475	6.778	.009	3.440	1.357	8.722
Pulm oedema*	2.499	1.059	5.571	.018	12.170	1.528	96.939
Upper lobe diversion	1.792	.661	7.355	.007	6.000	1.644	21.904
Auscultation	Only fine crepitations achieved significance						
ECG axis	.577	.387	2.220	.136	1.780	.834	3.802
ECG ST segment	-.239	.516	.214	.643	.788	.287	2.164
ECG conduct	Only LBBB achieved significance						
ECG T waves	1.109	.480	5.329	.021	3.031	1.182	7.771
ECG LVH	.577	.387	2.220	.136	1.780	.834	3.802
Clinical Impress	3.309	1.046	10.014	.002	27.351	3.524	212.313
ECG Conduction	.681	.304	5.027	.025	1.976	1.089	3.582

### 8.2.3 Correlation Analysis

As the first step for the multivariate analysis a correlation table was compiled of all of the continuous variables that reached significant levels according to

Wald's test in the univariate analyses. Pearson's test was used to look for significant correlation as shown in Table 8-10.

Table 8-10 - Correlations of Selected Continuous Variables

Correlations

		CT ratio	BNP	NT-proBNP	Cardiac diameter	Urea	Cr	EGFR	Potassium	Age	pH	pCO <sub>2</sub>	Plts	Syst. BP	Diast. BP	Temp.
CT ratio	Pearson Correlation	1	.264**	.188	.719**	.305**	.179	-.287**	.157	.261**	-.193	.338*	-.135	.214*	.239*	-.206*
	Sig. (2-tailed)		.009	.063	.000	.002	.080	.005	.146	.009	.180	.016	.188	.036	.019	.042
	N	99	98	98	99	99	97	96	87	99	50	50	97	96	96	98
BNP	Pearson Correlation	.264**	1	.877**	.288**	.350**	.332**	-.314**	.275**	.203*	-.109	-.065	-.155	-.036	.091	-.159
	Sig. (2-tailed)	.009		.000	.004	.000	.001	.001	.008	.039	.435	.638	.121	.719	.367	.109
	N	98	104	103	98	104	102	101	91	104	54	54	102	101	101	103

NT- proBNP	Pearson Correlation	.188	.877**	1	.235*	.393**	.385**	-.315**	.237*	.218*	-.059	-.181	-.265**	-.150	.021	-.126
	Sig. (2-tailed)	.063	.000		.020	.000	.000	.001	.024	.026	.670	.189	.007	.133	.838	.206
	N	98	103	104	98	104	102	101	91	104	54	54	102	101	101	103
Widespread cardiac diameter	Pearson Correlation	.719**	.288**	.235*	1	.315**	.281**	-.250*	.167	.160	-.289*	.175	-.292**	.188	.292**	-.104
	Sig. (2-tailed)	.000	.004	.020		.001	.005	.014	.121	.113	.042	.225	.004	.067	.004	.309
	N	99	98	98	99	99	97	96	87	99	50	50	97	96	96	98
Urea	Pearson Correlation	.305**	.350**	.393**	.315**	1	.748**	-.683**	.201	.276**	-.238	-.162	-.216*	-.116	-.104	-.073
	Sig. (2-tailed)	.002	.000	.000	.001		.000	.000	.055	.004	.080	.236	.028	.245	.296	.461
	N	99	104	104	99	105	103	102	92	105	55	55	103	102	102	104

Creatinine	Pearson Correlation	.179	.332**	.385**	.281**	.748**	1	-.847**	.290**	.214*	-.156	-.236	-.299**	-.008	.025	-.057
	Sig. (2-tailed)	.080	.001	.000	.005	.000		.000	.005	.030	.260	.086	.002	.934	.807	.572
	N	97	102	102	97	103	103	102	91	103	54	54	101	100	100	102
Estimated GFR	Pearson Correlation	-.287**	-.314**	-.315**	-.250*	-.683**	1	-.847**	-.221*	-.279**	.146	.301*	.275**	-.025	.034	.014
	Sig. (2-tailed)	.005	.001	.001	.014	.000		.000	.036	.004	.297	.028	.006	.803	.737	.886
	N	96	101	101	96	102	102	102	90	102	53	53	100	99	99	101
Potassium	Pearson Correlation	.157	.275**	.237*	.167	.201	.290**	-.221*	1	.003	-.274*	.206	.106	.085	.141	-.109
	Sig. (2-tailed)	.146	.008	.024	.121	.055	.005	.036		.976	.049	.143	.318	.429	.186	.305
	N	87	91	91	87	92	91	90	92	92	52	52	91	89	89	91

Age in years	Pearson Correlation	.261**	.203*	.218*	.160	.276**	.214*	-.279**	.003	1	-.078	-.239	.015	-.067	-.058	-.093
	Sig. (2-tailed)	.009	.039	.026	.113	.004	.030	.004	.976		.571	.079	.883	.502	.559	.349
	N	99	104	104	99	105	103	102	92	105	55	55	103	102	102	104
pH	Pearson Correlation	-.193	-.109	-.059	-.289*	-.238	-.156	.146	-.274*	-.078	1	-.123	.028	-.247	-.192	.093
	Sig. (2-tailed)	.180	.435	.670	.042	.080	.260	.297	.049	.571		.370	.843	.071	.163	.504
	N	50	54	54	50	55	54	53	52	55	55	55	54	54	54	54
pCO2	Pearson Correlation	.338*	-.065	-.181	.175	-.162	-.236	.301*	.206	-.239	-.123	1	.142	.298*	.077	-.207
	Sig. (2-tailed)	.016	.638	.189	.225	.236	.086	.028	.143	.079	.370		.306	.029	.582	.133
	N	50	54	54	50	55	54	53	52	55	55	55	54	54	54	54

platelets	Pearson Correlation	-.135	-.155	-.265**	-.292**	-.216*	-.299**	.275**	.106	.015	.028	.142	1	.038	-.089	-.031
	Sig. (2-tailed)	.188	.121	.007	.004	.028	.002	.006	.318	.883	.843	.306		.706	.376	.758
	N	97	102	102	97	103	101	100	91	103	54	54	103	100	100	102
Systolic BP	Pearson Correlation	.214*	-.036	-.150	.188	-.116	-.008	-.025	.085	-.067	-.247	.298*	.038	1	.598**	-.057
	Sig. (2-tailed)	.036	.719	.133	.067	.245	.934	.803	.429	.502	.071	.029	.706		.000	.569
	N	96	101	101	96	102	100	99	89	102	54	54	100	102	102	101
Diastolic BP	Pearson Correlation	.239*	.091	.021	.292**	-.104	.025	.034	.141	-.058	-.192	.077	-.089	.598**	1	-.220*
	Sig. (2-tailed)	.019	.367	.838	.004	.296	.807	.737	.186	.559	.163	.582	.376	.000		.027
	N	96	101	101	96	102	100	99	89	102	54	54	100	102	102	101

Temp eratur e	Pearson	-.206*	-.159	-.126	-.104	-.073	-.057	.014	-.109	-.093	.093	-.207	-.031	-.057	-.220*	1
	Correlatio n															
	Sig. (2-tailed)	.042	.109	.206	.309	.461	.572	.886	.305	.349	.504	.133	.758	.569	.027	
	N	98	103	103	98	104	102	101	91	104	54	54	102	101	101	104

\*\* . Correlation is significant at the 0.01 level (2-tailed).

\* . Correlation is significant at the 0.05 level (2-tailed).

In order to allow the decision pathway to include only variables that display independent diagnostic utility the data was examined to identify the most obvious dependent factors. Variables with significant correlation were examined in terms of their significance on univariate analysis and their contribution to the model in combined logistic regression. This led to the elimination of Cardiac Size in favour of Cardio-thoracic ratio (CTR), NT-proBNP in favour of BNP, estimated Glomerular Filtration (GFR) rate in favour of Urea and Systolic Blood pressure (SBP) in favour of Diastolic blood pressure (DBP).

The remaining continuous variables (CTR, BNP, DBP, Urea, Potassium, Age, pH, Platelet count and Temperature) were added to a logistic regression equation using the Enter Method in SPSS. Age, Platelet count and Temperature were then eliminated from the model, as they did not add any significant improvement. Although pH had been a significant diagnostic variable on univariate analysis it was excluded at this juncture, as it had only been performed on only a few of the patients. There were two reasons for this exclusion; the small number samples means that the analysis is less reliable and the investigation is likely to have been applied selectively in patients who are considered more unwell.

Performing the logistic regression analysis again with the reduced number of variables led to the removal of the potassium level and the diastolic blood pressure, as they did not add any significant value to the model. This left urea, BNP and the cardiothoracic ratio as the continuous variables in the equation.

The categorical data that had been significant on univariate analysis were then added to the logistic regression equation. Most of the categorical data was in the form of a simple dichotomy that could be positive or negative, for example a history of heart failure; other questions had multiple potential results, for example the diuretic used category was divided into the various types of diuretics that patients were taking. In the

later case the individual components were examined to see if they achieved significance, for example, the use of loop diuretics was positively associated with an end-diagnosis of heart failure while the use of thiazide diuretics did not achieve significance for this diagnosis.

The list consisted of: Loop Diuretic use, Paroxysmal Nocturnal Dyspnoea (PND), History of Heart Failure (HxHF), History of Coronary Artery Bypass Graft (HxCABG), History of Chronic Obstructive Pulmonary Disease (COPD), History of Hypertension (HxHBP), Use of Beta Blocker, Use of Angiotensin Receptor Blocker (ARB), presence of fine crepitations, Raised Jugular Venous Pulse (JVP), Ankle Oedema, Presence of Productive Cough, Pulmonary Oedema, Electrocardiogram (ECG) Conduction abnormalities, Presence of T wave abnormalities on ECG and Presence of Upper Lobe Diversion (ULD).

The binary logistic regression equation was then run repeatedly using the Enter method in SPSS. Following each calculation, the categorical variables that added the least to the model were removed and the process repeated. The presence of upper lobe diversion or cephalisation of vessels on the chest x-ray did not add anything significant to the model in the presence of the pulmonary oedema variable and so were removed.

The final model contained the following variables: Urea, Cardio-thoracic ratio, Brain Natriuretic Peptide, Loop Diuretic, Presence of Pulmonary Oedema and Clinical Impression of Heart Failure. The presence of pulmonary oedema was the least significant factor in the equation but was included as a positive finding of pulmonary oedema is highly predictive of an end-diagnosis of heart failure.

Although the presence of pulmonary oedema was considered as a binary variable for the purposes of this decision pathway it is recognised that the condition exists as a spectrum ranging from normal to grossly oedematous. It is reasonable to assume that severe pulmonary oedema

is likely to be less subjected to inter-observer disagreement and more pathognomonic for heart failure.

Table 8-11 - The Logistic Regression Model

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
							Lower	Upper
Step 1 <sup>a</sup>	9.314	3.664	6.463	1	.011	11095.450	8.443	14580842.456
ratio								
bnp	.002	.001	2.651	1	.103	1.002	1.000	1.004
ur	.170	.117	2.128	1	.145	1.186	.943	1.490
loopdiur(1)	1.492	.631	5.591	1	.018	4.446	1.291	15.316
clinhf(1)	1.581	1.180	1.795	1	.180	4.862	.481	49.149
pulmoed(1)	1.407	1.567	.807	1	.369	4.084	.189	88.017
Constant	-7.362	2.180	11.405	1	.001	.001		

Variable(s) entered on step 1: ratio, bnp, ur, loopdiur, clinhf, pulmoed.

S.E. = Standard Error, df = degrees of freedom, Sig. = Significance, Exp(B) = Exponential function of B

It can be seen on examining the data in Table 8-11, that only the Cardiothoracic ratio and the history of use of a loop diuretic achieve significance in terms of the Wald statistic. This is due to the chosen variables still having some degree of overlap in their association with the diagnosis of heart failure despite each being highly significant on univariate analysis. This is predictable as all of these variables are likely to be present in the patient with classical heart failure and the variables are related to each other in causing this condition. There is also the risk of the inclusion bias, as, with the exception of the BNP level, the cardiologists are likely to have

considered these variables in providing the diagnosis for use as the reference standard. Given the significance of each of these factors on univariate analysis, it would seem appropriate to include all of them in the logistic regression equation. This also adds robustness to the model although a slightly different equation would be required if any of the variables are missing; for example, a patient with a missing BNP value would not have the same risk of heart failure as a patient with a BNP value of zero.

The Hosmer-Lemeshow Test demonstrated a good fit of the model. The test statistic was 0.595 with 8 degrees of freedom indicating that there was no significant difference between the probability results predicted by the model and the actual results recorded.

To apply the logistic regression equation in practice means using the following equation:

$$\text{Log } [p/(1+p)] = -7.362 + [9.314 \cdot \text{CTR}] + [.002 \cdot \text{BNP}] + [.170 \cdot \text{Urea}] + [1.492 \cdot \text{Loop Diuretic}] + [1.581 \cdot \text{Clin HF}] + [1.407 \cdot \text{pulm oedema}]$$

Where  $p$  is the probability of heart failure. The data from the first five patients has been placed in Table 8-12 to demonstrate the calculated probabilities and how this compares with the reference standard applied. Figure 8-3 shows the ROC curve for the application of this logistic regression equation on this data set. The area under the curve for this equation was 0.817 (95% confidence 0.713 to 0.902). Table 8-13 shows the values of sensitivity and (1- specificity) that result from accepting different values of probability as cut-off values for defining whether a patient is considered to have heart failure. These values are the coordinates that are used to create the ROC curve in Figure 8-3.

Table 8-12 - Probability of Heart Failure for Five Patients Calculated Using the Logistic Regression Equation

Patient ID	CTR	Urea	Pulm. Oedema	Clinical HF	BNP	Loop Diuretics	Pred. Prob.	Ref. Stand.
1	.50	5.4	0	0	146.0	0	18.2%	Not
2	.61	7.1	0	0	40.1	0	39.3%	Not
3	.59	6.4	0	0	74.6	1	69.2%	Poss.
4	.54	4.2	0	0	78	0	18.3%	Not
5	.74	9.4	0	1	639	1	99.5%	Definite

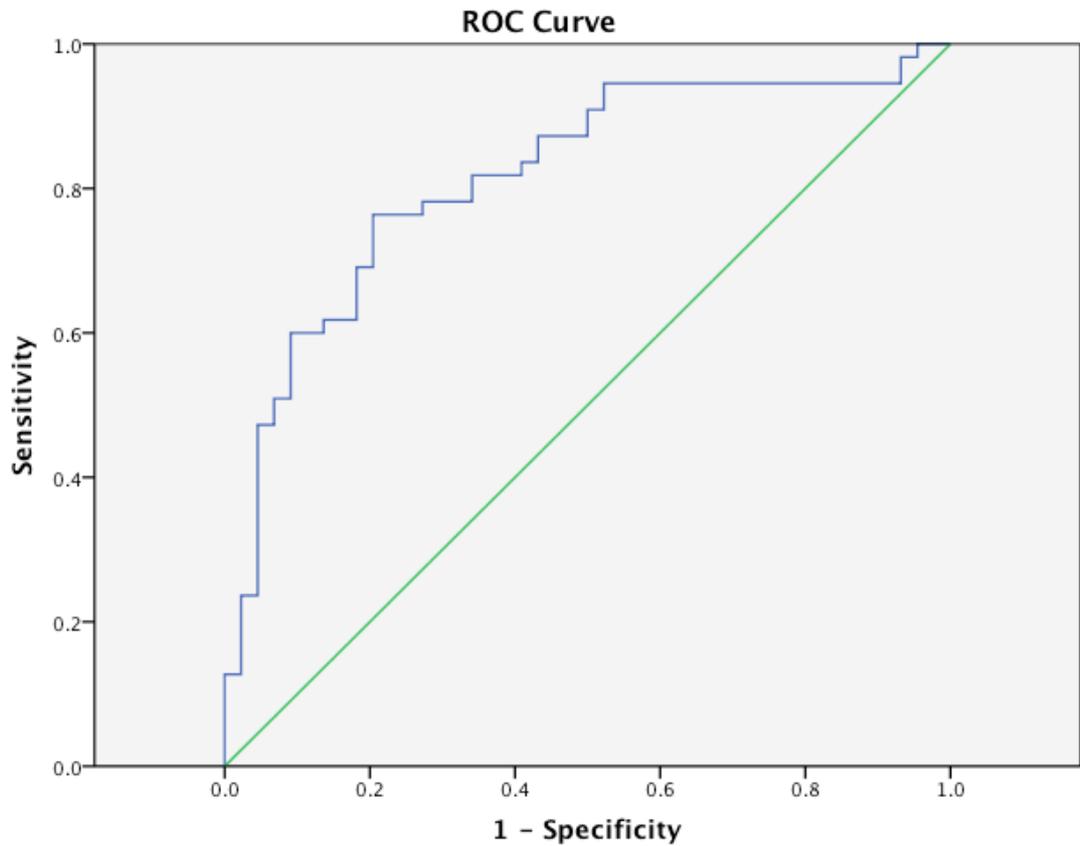


Figure 8-3 - ROC Curve Logistic Regression Equation

### Area Under the Curve

Test Result Variable(s): Predicted probability

Area	Std. Error <sup>a</sup>	Asymptotic Sig. <sup>b</sup>	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.817	.043	.000	.731	.902

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

Std. Error = Standard Error, Asymptotic Sig.= Asymptotic Significance.

Figure 8-4 - AUC Logistic Regression

Table 8-13 Coordinates of ROC Curve for LR Equation

Positive if $\geq$	Sensitivity	1 - Specificity
0.025	1.00	0.95
0.179	0.945	0.52
0.944	0.473	0.04
1.000	0.218	0.02

Testing of the logistic regression method was carried out using the leave-one-out method. In this method the equation is generated using all of the patients except one and then the equation is applied to the remaining patient. This process is repeated for every patient included in the study and the overall performance of the generated equations assessed. This resulted in an area under the curve of 0.8181 with a sensitivity of 0.7833, specificity of 0.6889 and a Brier value of 0.3517.

### 8.2.4 Scoring System Derived from the Logistic Regression Equation

Moon et al. (2002) have suggested a method of deriving a simple scoring system from logistic regression in order to make the application of the

equation easier without electronic support. The authors suggest rounding the value of the regression coefficient and multiplying by ten to provide a score for each variable in the equation.

$$\text{Log } [p/(1+p)] = -7.362 + [9.314 \cdot \text{CTR}] + [.002 \cdot \text{BNP}] + [.170 \cdot \text{Urea}] + [1.492 \cdot \text{Loop Diuretic}] + [1.581 \cdot \text{Clin HF}] + [1.407 \cdot \text{pulm oedema}]$$

Multiplying the B values for each factor by 10 and then rounding gives the following values:

- Loop diuretic = 15
- Clinical HF = 16
- Pulm oedema = 14
- CTR = 10 for every .1 (100 for 1)
- Urea = 2 for every mg
- BNP = 2 for every 100

Applying this to the first five patients from my data set would give the following result:

Table 8-14 - Scoring System Derived from Logistic Regression Equation

Pt ID	CTR	Urea	Pulm Oed	Clin HF	BNP	Loop Diuretic	Score	Ref. Stand.
1	.50	5.4	0	0	146.0	0	64	Not
2	.61	7.1	0	0	40.1	0	75	Not
3	.59	6.4	0	0	74.6	1	88	Possible
4	.54	4.2	0	0	78	0	63	Not
5	.74	9.4	0	1	639	1	122	Definite

Even with rounding up values and using approximations this equation is still complicated to use and the result is more difficult to interpret than a simple percentage probability. Although the score results show correlation with the probabilities calculated by the logistic regression

equation, as shown in Figure 8-5 this method does not appear to confer any major advantage compared with using the original logistic regression equation.

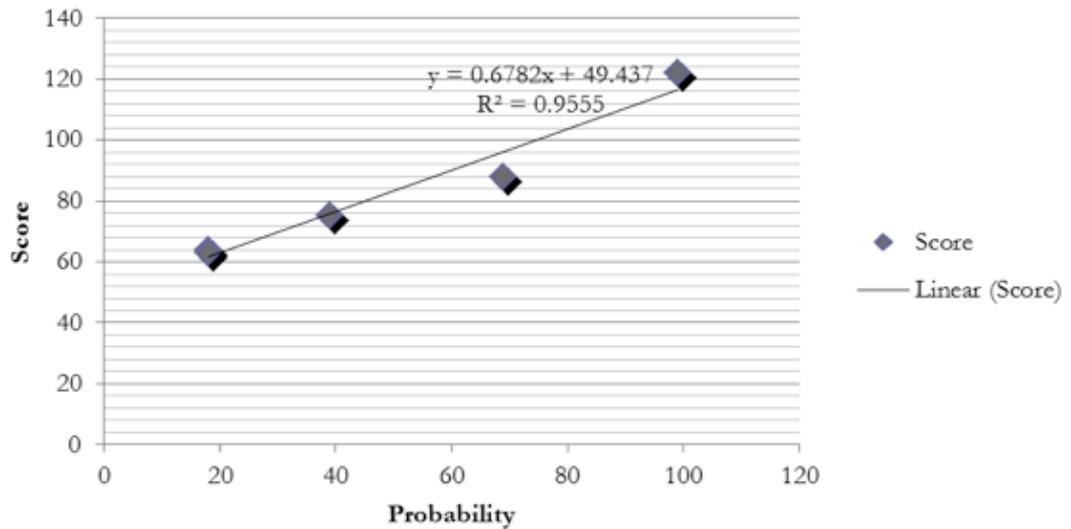


Figure 8-5 LR Model Compared Against Scoring System

### 8.2.5 Classification and Regression Tree Analysis

The method described by Breiman et al. (1995) was used in this analysis. The results were of limited value due in part to the relatively small study sample population and the nature of the clinical application.

Trees were created using all of the variables, the variables that were significant on univariate analysis, the variables selected for the logistic regression model, the variables selected by the RandomForest software and the variables selected from the systematic review and meta-analysis.

There were several problems with the automated tree production; selection of the variables, the order that the variables were utilised, multiple cut-off values for the same variable and limitation of some variables to distal portions of the tree.

In clinical practice it makes sense to apply the broadest and least invasive investigative tools earlier and apply tests with more specific results only in selected groups of patients. For example, all patients presenting to an Emergency Department with acute breathlessness will have chest auscultation performed, whereas only a selection of these patients, in whom the attending physician considers at risk of heart failure, may benefit from a BNP test.

The software is applied only to compare the diagnosis of heart failure with the diagnosis of 'not heart failure' but this second group is extremely heterogeneous. Diagnostic tests such as ECG or chest x-ray may produce findings that are pathognomonic for other conditions and therefore rule-out heart failure by being specific for the other cause.

Another issue was that some aspects of the produced trees did not make clinical sense. For example, patients with a certain level of troponin or higher were deemed to be in the high-risk group for heart failure. While patients in the lower group were dichotomised again, further down the tree, on the basis of their troponin level but in this instance the patients with the lower level were considered to be the high-risk group for heart failure.

The depth of the tree could be controlled either by the number of branches, the number of patients in the end-nodes or the purity of the end-nodes but no optimal level was obtained. This may result from the high number of potential variables and the relatively low number of patients, but had the consequence that the predictive value of the model was directly related to the complexity of the model. Trees that had a very good predictive model were very large, cumbersome, limited in sound theory and over-fitted the data. When the trees were limited by controlling the number of variables or nodes, the simpler trees did not demonstrate diagnostic utility greater than that of the individual components. This may also be related to the lack of independence between the end variables.

The fourth issue was the need to split continuous variables to define high or low risk groups. This essentially discarded potentially useful data. For example, while a patient with a BNP value of greater than 150 pg/ml is more likely to have heart failure than a patient with a value less than this, a patient with a BNP value of over 1500 pg/ml is much more likely to have this condition than a patient with a value of 151 pg/ml. This was partially overcome where the same variable was used at different nodes with different thresholds but this made of a cumbersome and complicated tree.

Another issue in applying the trees is a result of the dichotomous nature of the nodes; for example, a patient could be classified as lower risk on the basis of the initial variable, e.g. BNP and no other variable would then be considered. On some trees it was possible for a patient to have a BNP level of 300pg/ml and be classified as low risk on the first node despite a previous history of heart failure, cardiomegaly and pulmonary oedema on the chest x-ray, on-going diuretic use and an abnormal ECG.

Despite trying many different variations of included variables and tree-growing methods, it was not possible to synthesise a clinically useful decision tree from the available data.

### **8.2.6 Manual Classification Tree**

Attempts were made to create a decision tree manually using Orange software, open-source data-mining software (Demšar et al., 2013). This software can be used to synthesize a decision tree automatically or can allow the user to choose the variables, select the cut-off values, control the order of the decision nodes and decide the depth of the tree or any combination of these methods.

Various efforts were made to create decision trees varying from completely automated to completely manual. Various data sets were used; the complete data set, variables that had been selected by the automatic tree synthesis, variables with significant univariate analysis

results, variables selected by the systematic review and meta-analysis, variables selected by the Random Forest software and the variables selected from the logistic regression model.

Despite experimenting with every possible setting the same issues arose as when the CART analysis was performed: trees that provided excellent levels of sensitivity, specificity, or both were huge, cumbersome and over-fitted; while honing down the trees to contain only the variables selected as having the greatest utility produced trees with diagnostic utility no better than their individual components.

An example of a manual decision tree is shown in Figure 8-6. The pie charts represent the proportion of patients with an end-diagnosis of heart failure in red, and the proportion of patients with an end-diagnosis of not heart failure in blue. This particular model had a sensitivity of 82.1% and a specificity of 75.0% as shown in Table 8-15 and as such added little to the diagnostic performance of the individual variables.

Table 8-15 - Diagnostic Performance of Manual Decision Tree

	HF	Not HF	Totals
Test +ve	46	10	56
Test -ve	10	30	40
Totals	56	40	96

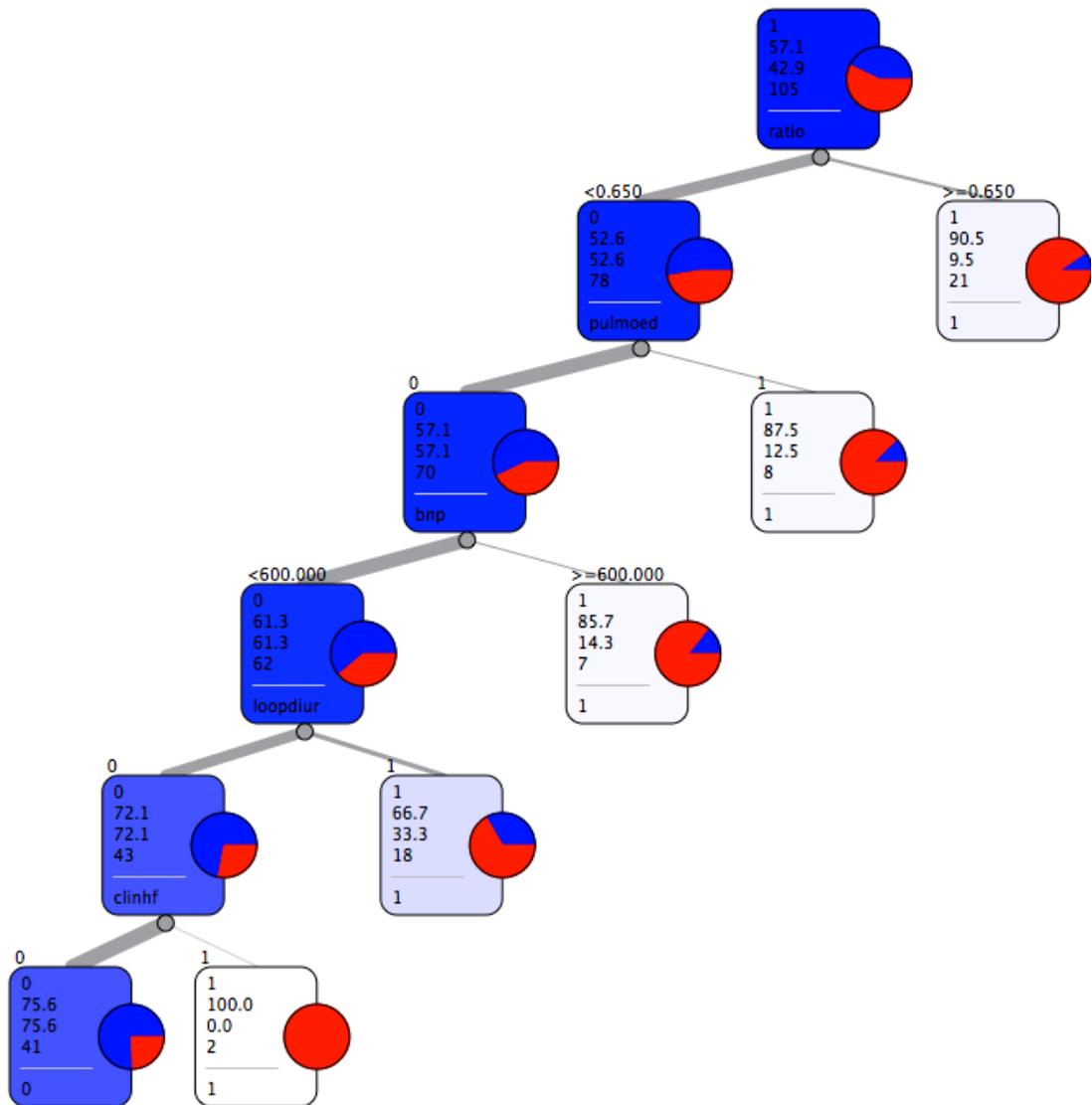


Figure 8-6 Example of Manual Decision Tree Created Using Orange Software (Demšar et al., 2013)

### 8.2.7 Naïve Bayes Analysis

Naïve Bayes analysis was performed using Orange software (Demšar et al., 2013). The first analysis was performed with all the variables that achieved significance on univariate analysis and demonstrated independence on correlation analysis.

The included continuous variables were: cardiothoracic ratio, BNP, urea, age, temperature, potassium, diastolic blood pressure, pH and platelets. The discrete variables included were: use of loop diuretics, paroxysmal nocturnal dyspnoea, a history of heart failure, a history of CABG, a history of COPD, a history of hypertension, use of a Beta blocker, use of an ARB, presence of fine crepitations, presence of a raised JVP, a conduction abnormality on ECG, T wave inversion on ECG and the presence of upper lobe diversion.

When this system is used to its full potential it classifies all of the included patients correctly. Where the system dichotomises each of the variables, selecting the optimum cut-off point for continuous variables to have a positive or negative influence, the overall sensitivity is 0.7333, specificity is 0.844 and the area under a ROC curve is 0.8359 with a Brier value of 0.4147. This was estimated using the leave-one-out analysis.

	Heart Failure	Not Heart Failure	
Test +ve	44	16	60
Test -ve	7	38	45
	51	54	105

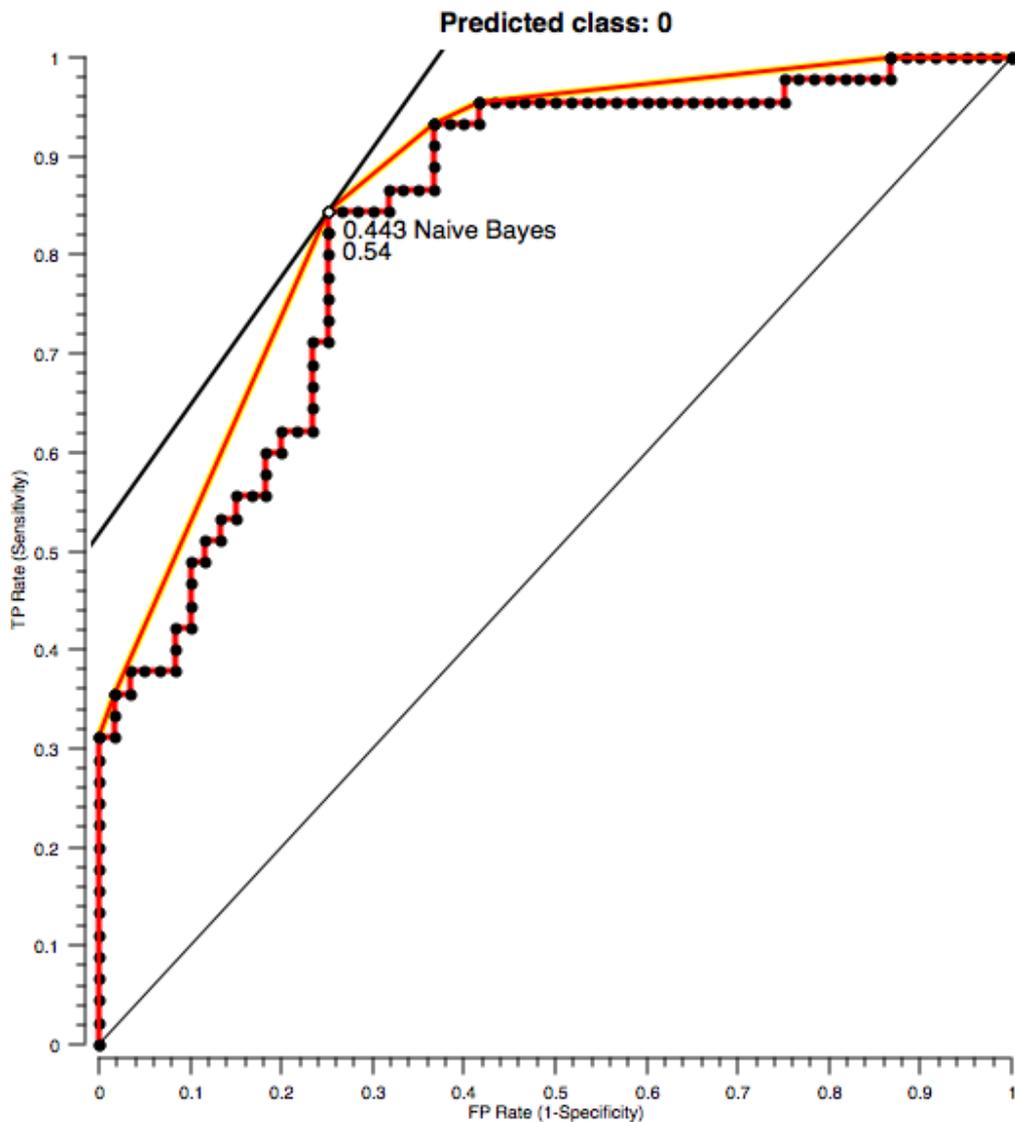


Figure 8-7 ROC Curve for Naive Bayes Analysis of Variables Significant on Univariate Analysis

The analysis was repeated using the first ten variables ranked in the RandomForest analysis. The Naïve Bayes analysis produced a decision tool with a sensitivity of 0.750, a specificity of 0.711, an AUC of 0.851 and a Brier value of 0.3851. The variables used were: cardiac size, cardiothoracic ratio, urea, NT-proBNP, BNP, clinical heart failure, creatinine, troponin, age and diastolic blood pressure.

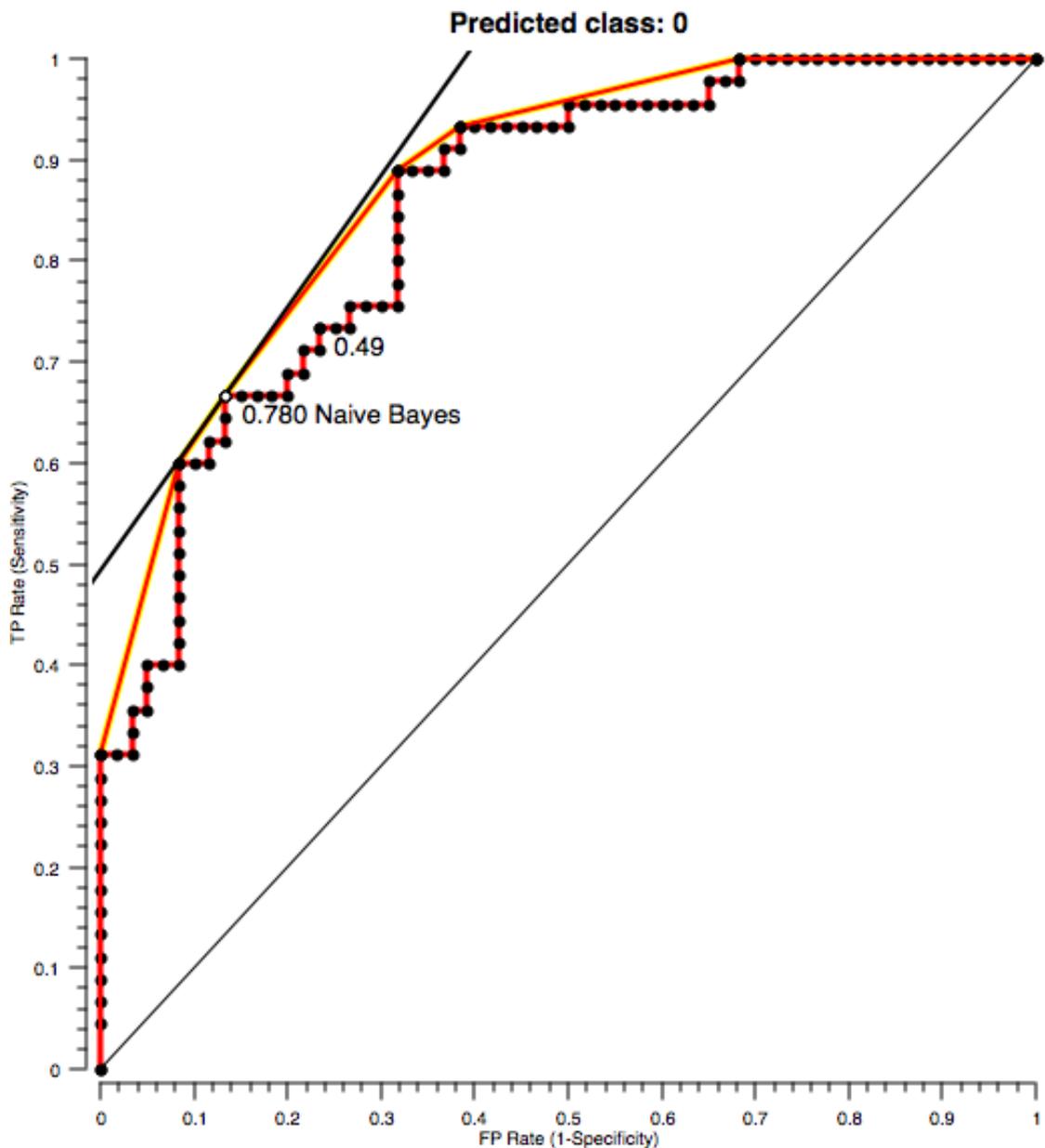


Figure 8-8 ROC Curve for Bayes Analysis using First Ten Variables from RandomForest Analysis

The Naïve Bayes Analysis was also performed for the six variables that were found to perform best in the logistic regression equation. With a limited number of variables considered it is feasible to use the full volume of information provided by them rather than simple dichotomise each variable into the best-performing positive or negative result. This can be performed using a computer but can also be described using a nomogram. Drawing a straight vertical line from the result of each

available variable provides a score. Totalling the scores for all of the available variables provides a probability value. The results of this decision tool have been completed for the first patient in the trial as an illustration in Figure 8-9. Performing this provides 100% sensitivity and specificity for the patients included.

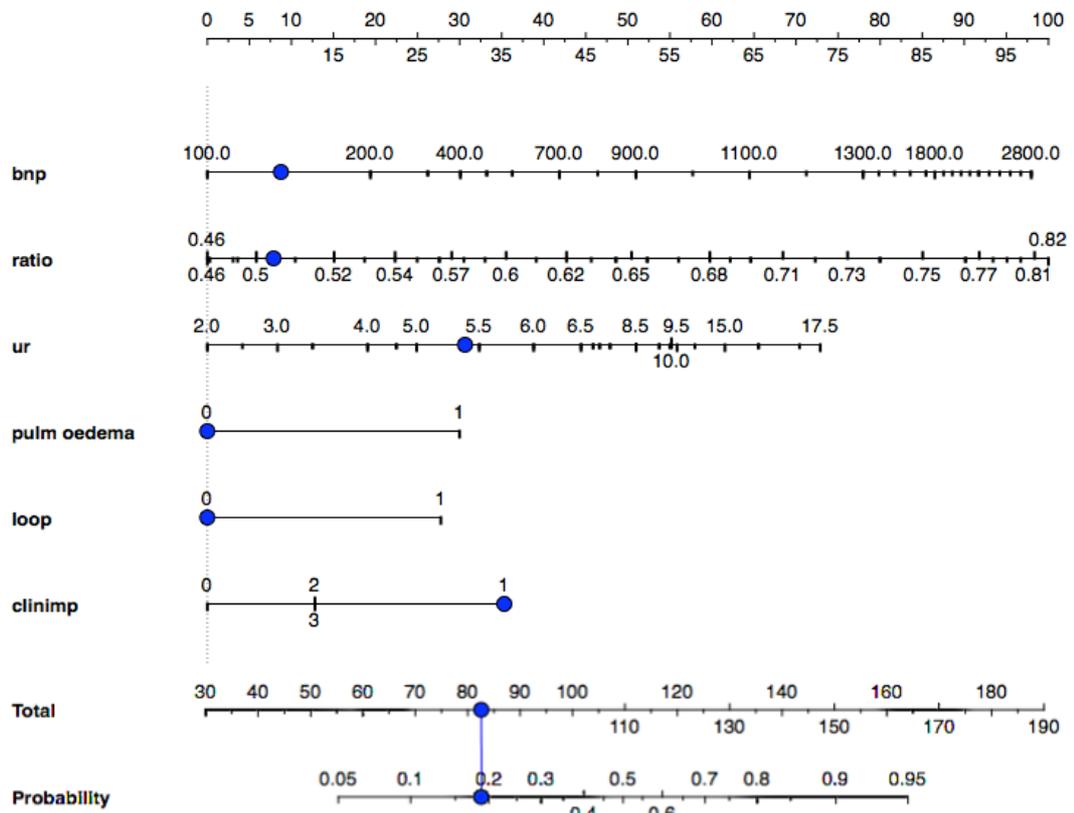


Figure 8-9 Naive Bayes Analysis Nomogram Applied to Patient 1 Results

Performing leave-one-out analysis demonstrates a more realistic estimation of the performance. In this process, the optimal scoring values for each of the variables are calculated based the results of all of the patients bar one, the rule is then applied to the single remaining patient and it is recorded whether or not this is correct. This process is then applied repeatedly missing out each patient in turn until the analysis has been performed missing out each patient in turn.

This resulted in sensitivity of 0.816, specificity of 0.667, AUC of 0.8274 and a Brier score of 0.3907 as shown in Figure 8-10.

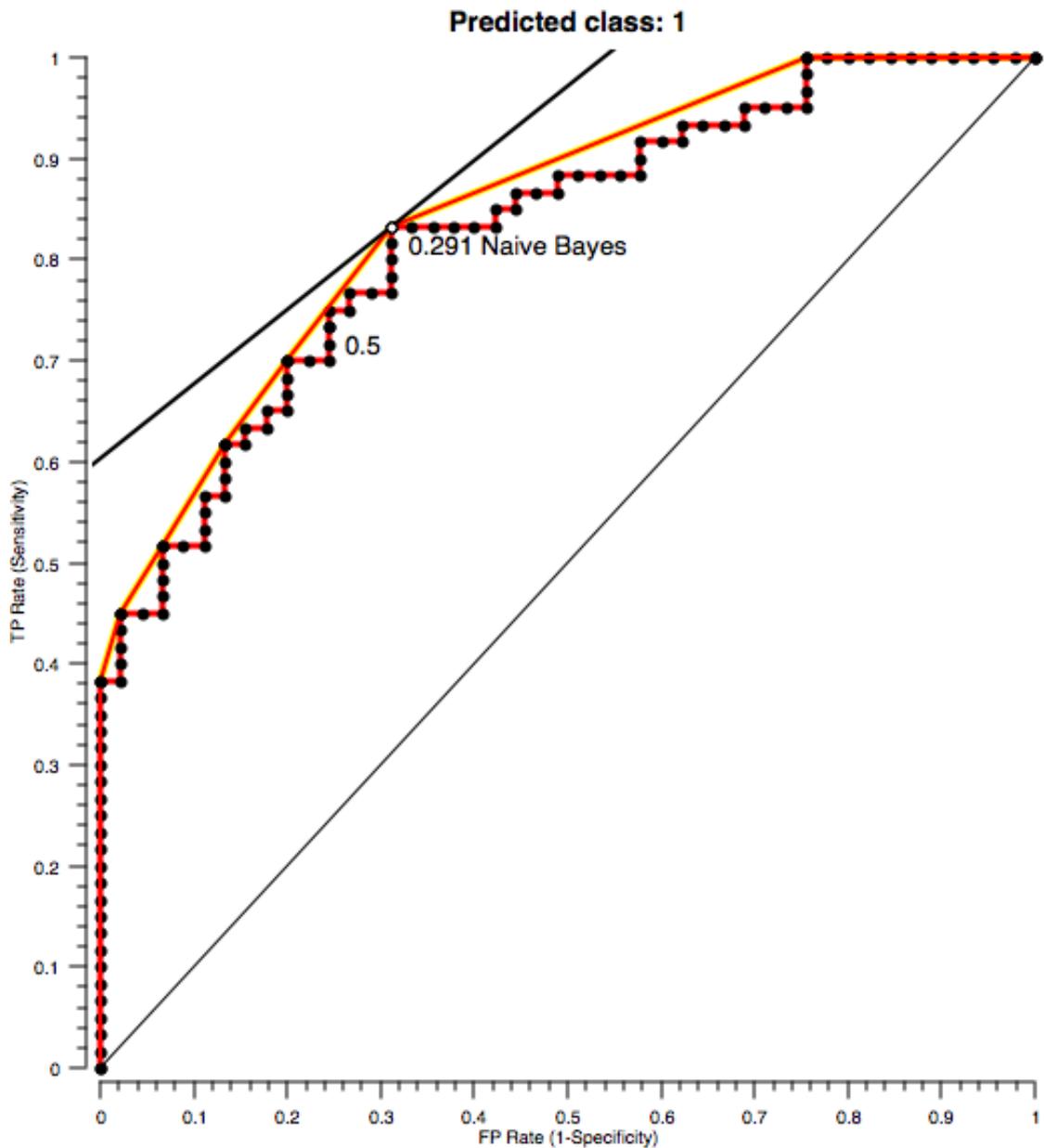


Figure 8-10 ROC Curve for Variables Selected by Logistic Regression when Naive Bayes Analysis performed using 'Miss-One-Out' analysis

In order to obtain a more realistic idea of how this decision rule would function in real life a random selection of the patients were chosen to derive the rule. The patients were divided into a derivation group (70%)

and a validation group (30%). The rule derived again had 100% sensitivity and specificity for the derivation group if the rule was completely applied. For the patients within the derivation group using the 'leave-one-out' analysis the sensitivity was 0.833, the specificity 0.7500, the AUC 0.8802 and the Brier statistic 0.3150.

When this rule was applied to the validation group the sensitivity result decreased to 0.66 and the specificity was 0.69 as shown in Table 6.10.

Table 8-16 Results of Decision Tool on Validation Group

Pt ID	CTR	Oed	Loop	BNP	Urea	Clin Imp	Final Diag	NB	Prob
4	0.533142	0	1	295	9	1	0	1	0.99
11	0.524334	0	1	23.8	8.2	0	1	0	0.28
15	0.411386	0	1	10.9	7.2	0	1	0	0
20	0.600845	1	1	15.8	3.3	1	1	1	1
21	0.489349	1	1	411	7.5	1	0	1	1
22	0.73969	1	0	549	7.6	1	1	1	1
25	0.3704	0	0	162	7.5	0	1	0	0
30	0.62047		0	354	5.9	0	1	0	0.45
34	0.355315	0	0	259	8.7	1	0	0	0.01
35	0.556744	1	0	166	6.9	1	0	1	1
36	0.683598	1	0	1280	8.5	1	1	1	1
37	0.507344	0	0	15.9	4.1	0	0	0	0.01
40	0.622122	1	1	370	9	1	1	1	1
47			0	2680	5.3	1	1	1	0.98
49	0.685531	1		193	5.7	1	1	1	1
53	0.437447		0	146	3.4	0	0	0	0
57	0.456707		0	12.7	2.2	0	0	0	0
59	0.611252		1	90.4	1.9	0	0	0	0.13
62	0.418732		0	25	9.4	0	1	0	0.01

66	0.462505		0	6.1	5.2	0	0	0	0
67	0.619351		1	1680	12	1	1	1	1
68		0	0	27.3	5.7	0	1	0	0.05
71	0.607582		1	118	7.7	1	1	1	0.9
81	0.5		0	1290	6.6	1	1	1	0.91
85	0.447697		0	143	6.5	0	0	0	0
91	0.578213		1	106	4.9	1	1	1	0.55
92			0	28.9	7.7	0	0	0	0.14
94	0.601588		0	30	7.9	0	1	0	0.16
97	0.454202		0	85.9	3.5	0	0	0	0
99	0.640744		1	3450	7.2	1	1	1	1
100	0.692517		0	83.7	8	0	0	1	0.66

### 8.2.8 Application of the Framingham Criteria

The Framingham criteria for heart failure have been derived from the Framingham study to aid in the diagnosis of heart failure (McKee et al., 1971). The criteria are listed in Table 8-17. The majority of these criteria are available from the data recorded from the patients included in this trial. Central venous pressure, hepato-jugular reflux, weight loss over five days and vital capacity were not recorded.

Table 8-17 Framingham Criteria for Heart Failure

□□□□□□□□□□□□□□□□
<ul style="list-style-type: none"> <li>• Paroxysmal Nocturnal Dyspnoea</li> </ul>
<ul style="list-style-type: none"> <li>• Neck vein distension</li> </ul>
<ul style="list-style-type: none"> <li>• Rales</li> </ul>
<ul style="list-style-type: none"> <li>• Radiographic cardiomegaly</li> </ul>
<ul style="list-style-type: none"> <li>• Acute pulmonary oedema</li> </ul>
<ul style="list-style-type: none"> <li>• S3 gallop rhythm</li> </ul>
<ul style="list-style-type: none"> <li>• Increased central venous pressure</li> <li>• (&gt;16cm H<sub>2</sub>O at right atrium)</li> </ul>
<ul style="list-style-type: none"> <li>• Hepatojugular reflux</li> </ul>
<ul style="list-style-type: none"> <li>• Weight loss &gt;4.5kg in 5 days response to treatment</li> </ul>
Minor criteria
<ul style="list-style-type: none"> <li>• Bilateral ankle oedema</li> </ul>
<ul style="list-style-type: none"> <li>• Nocturnal cough</li> </ul>
<ul style="list-style-type: none"> <li>• Dyspnoea on ordinary exertion</li> </ul>
<ul style="list-style-type: none"> <li>• Hepatomegaly</li> </ul>
<ul style="list-style-type: none"> <li>• Pleural effusion</li> </ul>
<ul style="list-style-type: none"> <li>• Decrease in vital capacity by 1/3 from maximum</li> <li>• recorded</li> </ul>
<ul style="list-style-type: none"> <li>• Tachycardia (&gt;120bpm)</li> </ul>

The application of the available criteria from the Framingham study resulted in the data shown in Table 8-18. This results in a sensitivity value of 79.7% and a specificity of 60.9%.

Table 8-18 Application of Framingham Criteria to Included Patients

	Ref. Stand. +ve	Ref. Stand. -ve	
Framingham score +ve	47	18	65
Framingham score -ve	12	28	40
Totals	59	46	105

Ref. Stand = Reference Standard

Applying the Boston scoring systems, as shown in Table 8-19, gives the results shown in Table 8-20.

Table 8-19 Boston Criteria for Heart Failure

Criterion	Point value
Category 1: history	
Rest dyspnoea	4
Orthopnoea	4
Paroxysmal Nocturnal Dyspnoea	3
Dyspnoea walking on flat	2
Dyspnoea while climbing	1
Category II: physical examination	
Heart rate abnormality (1pt 91-110bpm, 2 pts >110bpm)	1 or 2
JVP (2 pts if >6cmH <sub>2</sub> O, 3 pts >6cm H <sub>2</sub> O and hepatomegaly)	2 or 3
Lung crackles (1pt basilar, 2pts more than basilar)	1 or 2
Wheezing	3
Third heart sound	3
Category III: chest radiography	
Alveolar pulmonary oedema	4
Interstitial pulmonary oedema	3
Bilateral pleural effusion	3
Cardiothoracic ratio > 0.50	3
Upper zone flow redistribution	2

Pts = Points

Table 8-20 Application of Boston Criteria to Study Patients

Boston Criteria	Not HF	Possible HF	Definite HF
Low	3	1	0
Mod	9	3	5
High	34	25	25

This translates into a sensitivity of 98.3% but a specificity of 6.7% if moderate to high probability Boston score is compared with possible or definite heart failure on clinical opinion.

### 8.2.9 Performance of Individual Variables

Certain variables were consistently identified by the different selection methods to have reasonable diagnostic utility when applied as a single diagnostic test. By altering the cut-off threshold for the continuous variables it is possible to manipulate the test to provide the required level of sensitivity or specificity.

#### 8.2.9.1 Cardiothoracic Ratio

The cardiothoracic ratio (CTR) could not be calculated in six patients. This was due to lung pathology such as pulmonary effusions, which obscured the cardiac border although this also demonstrated the likeliest aetiology of the patients' symptoms and so provided a diagnosis. A ROC curve of the CTR against the reference standard of possible or definite heart failure is shown in Figure 8-11. This provided an area under the curve of .792 (95% confidence interval .703 -.881). Table 8-21 shows various cut-off values and the corresponding sensitivity and specificity results illustrated in the ROC curve.

It can be seen from this table that using a cut-off of  $\geq .49$  for the CTR would provide a sensitivity of 94.6% and a specificity of 37.2% (Positive Likelihood Ratio (PLR) 1.51 and Negative Likelihood Ratio (NLR) .14). Alternatively, this measurement could be used as a specific diagnostic test by using a cut-off value of .65 resulting in a sensitivity of 33.9% with a specificity of 95.3% (PLR of 7.29, NLR .69).

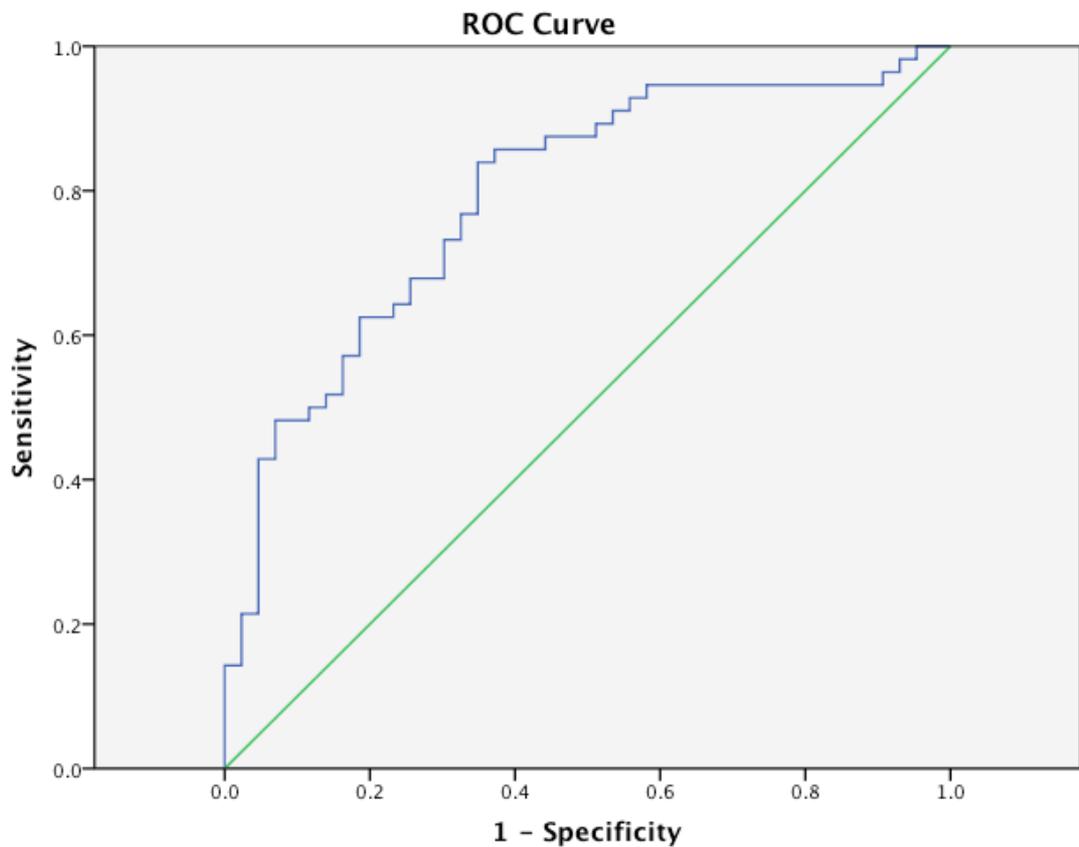


Figure 8-11 Receiver Operating Characteristic (ROC) Curve for cardiothoracic ratio for Possible/Definite Heart Failure

Area Under the Curve

Test Result Variable(s): Cardiothoracic ratio

			Asymptotic 95% Confidence Interval	
Area	Std. Error <sup>a</sup>	Asymptotic Sig. <sup>b</sup>	Lower Bound	Upper Bound
.792	.046	.000	.703	.881

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

- Std. Error = Standard Error, Asymptotic Sig. = Asymptotic Significance

Figure 8-12 AUC for CT ratio

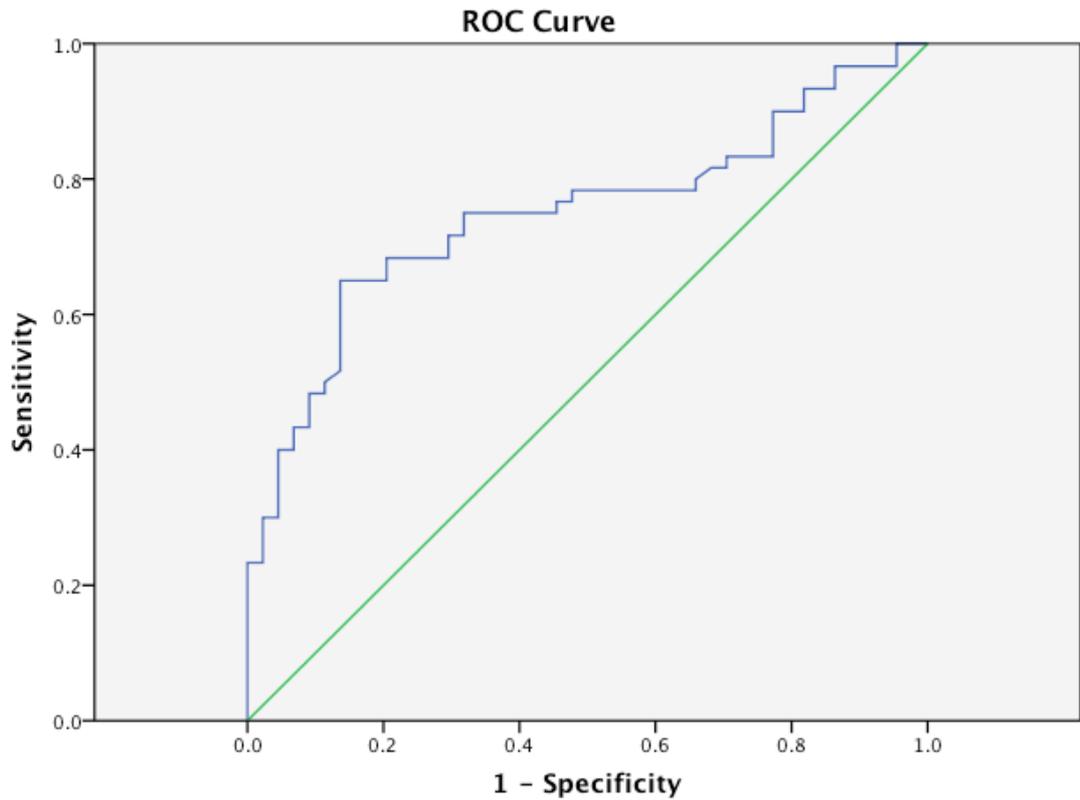
Table 8-21 Coordinates for Cardiothoracic Ratio ROC Curve

Positive if $\geq$	Sensitivity	1 - Specificity
0.3629	1.00	0.953
0.4995	0.946	0.581
0.6249	0.429	0.047
0.6975	0.143	0.000

8.2.9.2 Brain Natriuretic Peptide

BNP performed well as a diagnostic test over a range of values as can be seen on the ROC Curve in Figure 8-13. The area under the curve is .749 (95% confidence interval .655 to .843). Lowering the threshold did not provide sensitivity sufficient to rule-out the condition unless it was reduced to such a point that only a few patients were included in the analysis. Raising the cut-off to 400 pg/ml provided a sensitivity of 43.3% and a specificity of 90.9% (PLR 4.77, NLR.62). Raising the threshold further to 600 pg/ml provided an improved rule-in performance with a

sensitivity of 30.0% but a specificity of 97.7% (PLR 13.2, NLR .95). The sensitivity and specificity values associated with the various potential cut-off values are listed in Table 8-22.



Diagonal segments are produced by ties.

Figure 8-13 Receiver Operating Characteristic Curve (ROC) BNP for Definite / Probable Heart Failure

Area Under the Curve

Test Result Variable(s): Brain Natriuretic Peptide

Area	Std. Error <sup>a</sup>	Asymptotic Sig. <sup>b</sup>	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.749	.048	.000	.655	.843

The test result variable(s): BNP has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

Std. Error = Standard Error, Asymptotic Sig. = Asymptotic Significance

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

Table 8-22 - Diagnostic Performance of BNP

Cut-off value (pg/ml)	Sensitivity (%)	Specificity (%)	Positive Likelihood Ratio	Negative Likelihood Ratio
100	73.3	68.2	2.3	.76
400	43.3	90.9	4.77	.62
600	30.0	97.7	13.2	.95

Table 8-23 Coordinates for ROC Curve for BNP

Positive if greater than:	Sensitivity	1 - Specificity
6.2	1.000	0.955
11.250	0.950	0.864
482.00	0.400	0.045

906.5	0.217	0.000
-------	-------	-------

### 8.2.9.3 Urea

Urea performed less well than CTR or BNP with a smaller AUC of .712 (95% Confidence Interval .612 to .813) as shown in Figure 8-14. Although cut-offs could be chosen that have sufficient sensitivity or specificity to rule-out or rule-in the condition of heart failure they are at the extremes of the range of distribution of this variable and so are only applicable to a small proportion of presenting patients.

Table 8-24 - Diagnostic Utility of Urea as a Single Variable

Cut-off Value	Sensitivity	Specificity	PLR	NLR
4.5 mg/l	95.0	24.4	1.26	.20
12.8 mg/l	11.7	97.8	5.25	.90

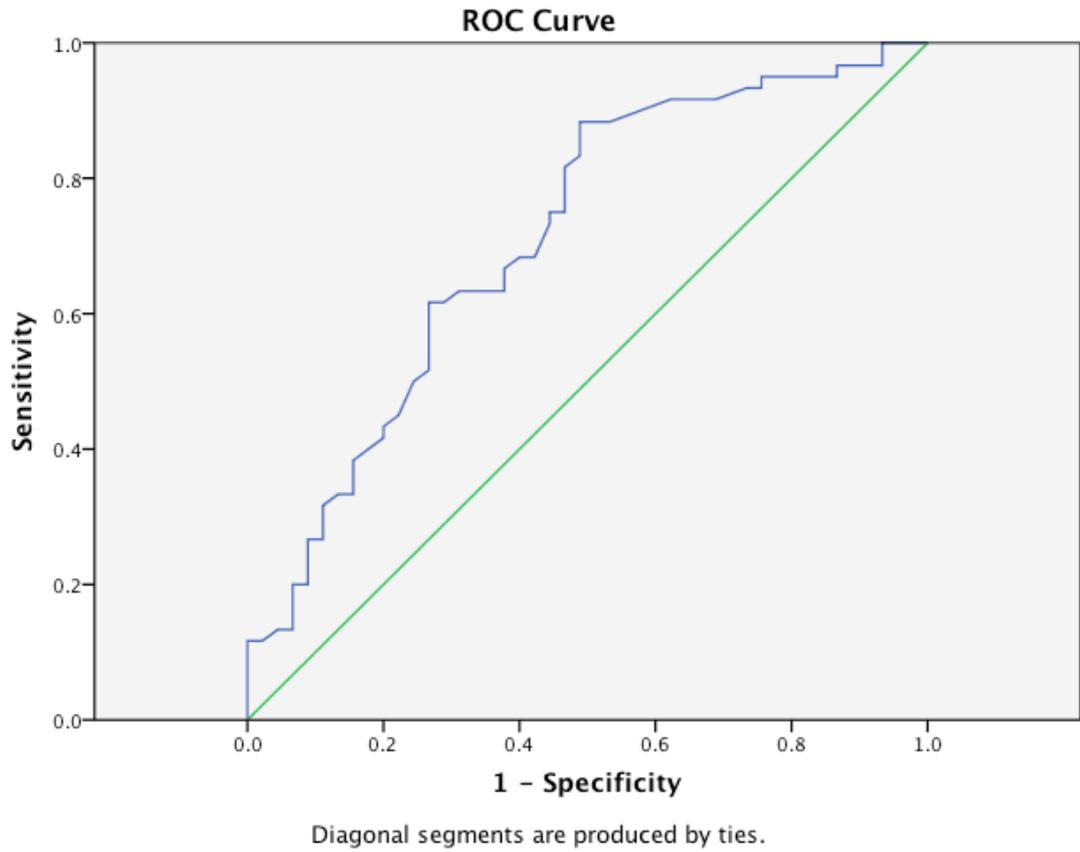


Figure 8-14 – Receiver Operating Characteristic Curve for Urea for Diagnosis of HF

### Area Under the Curve

Test Result Variable(s):Urea

Area	Std. Error <sup>a</sup>	Asymptotic Sig. <sup>b</sup>	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.712	.051	.000	.612	.813

The test result variable(s): Urea has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

Std. Error = Standard Error, Asymptotic Sig. = Asymptotic Significance

Figure 8-15 Area Under the Curve for Urea

Table 8-25 Coordinates for ROC Curve for Urea

Positive if greater than:	Sensitivity	1 – Specificity
2.650	1.00	0.933
4.550	0.950	0.756
12.750	0.133	0.044
13.350	0.117	0.000

#### 8.2.9.4 Clinical Impression of Heart Failure

The overall clinical impression of heart failure performed well in this study in terms of the specificity of this test. The sensitivity was 30.0%, the

specificity was 97.7%, the positive likelihood ratio was 13.2 and the negative likelihood ratio was 0.72.

#### *8.2.9.5 Presence of Pulmonary Oedema*

The presence of pulmonary oedema was also a specific finding for this data set. The sensitivity was 20.0%, the specificity of 95.5%, the positive likelihood ratio was 4.5 and the negative likelihood ratio was 0.84.

#### *8.2.9.6 Use of Loop Diuretic*

The use of a loop diuretic is likely to be a proxy marker for a previous diagnosis of heart failure. The other main indication for the prescription of this class of drugs is hypertension, itself a risk factor for developing heart failure. The sensitivity of the use of a loop diuretic as a predictor for heart failure was 38.3%, specificity was 81.8%, positive likelihood ratio was 2.75 and negative likelihood ratio 0.61.

#### *8.2.9.7 NT-proBNP*

NT-proBNP also performed well as a diagnostic test over a range of values. The ROC Curve is presented in Figure 109. The area under the curve was 0.756 (95% Confidence Interval 0.663 to 0.849). NT-proBNP performed in a very similar manner to BNP in that while it was possible to choose a cut-off value that was sensitive enough to rule-out the condition, the specificity at this level was so poor that there was little practical benefit. Choosing a cut-off value of 500pg/ ml, well below the recommended value of 1800 pg/ml, was sufficiently specific to rule the condition in as shown in Table 8-26

Table 8-26 - Diagnostic Value of NT-proBNP

Cut-off Value	Sensitivity	Specificity	PLR	NLR
7 pg/ml	93.5%	4.8%	1.35	.98
300 pg/ml	46.7%	88.6%	4.33	.60
500 pg/ml	31.2%	95.3%	6.70	.72
1800 pg/ml	13.7%	100%	Inf	.86

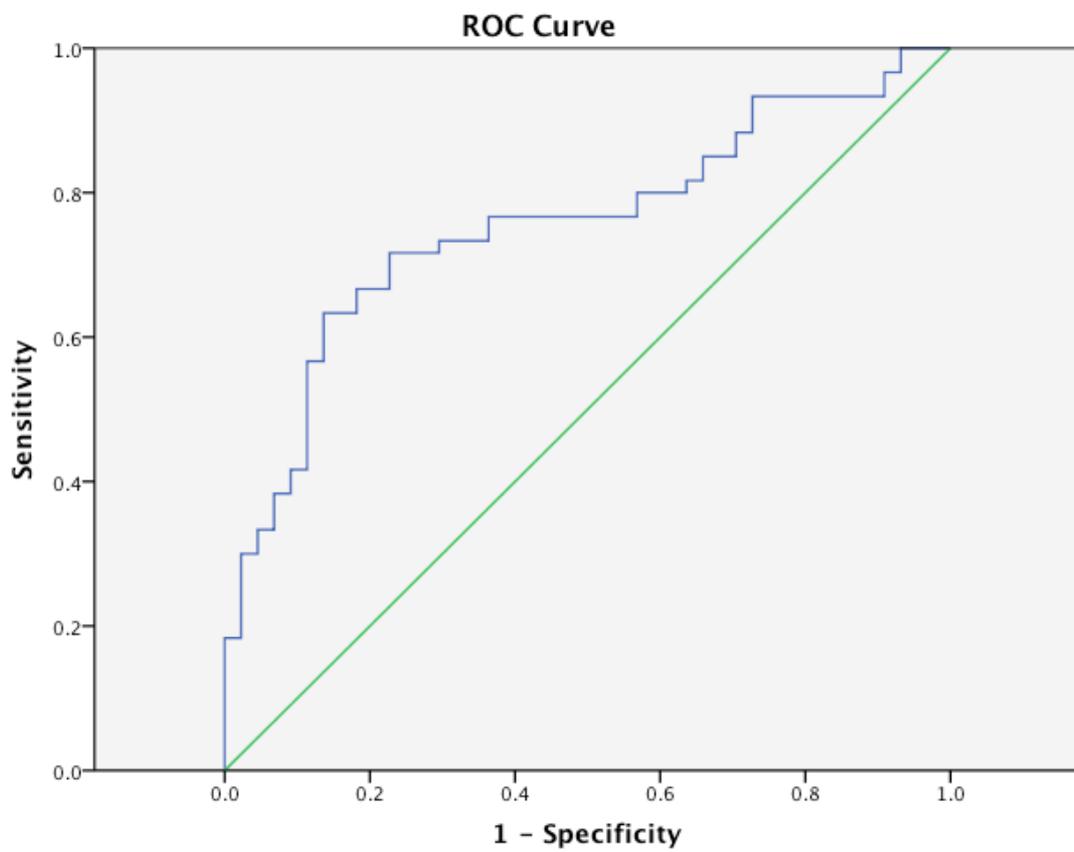


Figure 8-16- Receiver Operating Characteristic Curve for NT-proBNP

### Area Under the Curve

Test Result Variable(s): NT-proBNP

Area	Std. Error <sup>a</sup>	Asymptotic Sig. <sup>b</sup>	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.756	.048	.000	.663	.849

a. Under the nonparametric assumption

b. Null hypothesis: true area = 0.5

Std. Error = Standard Error, Asymptotic Sig. = Asymptotic Significance

Table 8-27 Coordinates for ROC Curve for NT-proBNP

Positive if greater than:	Sensitivity	1 - Specificity
4.505	1.00	0.932
7.825	0.950	0.909
452.45	0.333	0.045
1077.5	0.183	0.000

#### 8.2.9.8 Upper Lobe Diversion

Upper lobe diversion or cephalisation of the vessels on chest x-ray was found to be a useful diagnostic feature for heart failure but was outperformed by the presence of pulmonary oedema in this study population. The sensitivity of this test was 30.0%, the specificity was 93.3%, the positive likelihood ratio was 4.5 and the negative likelihood ratio was 0.75.

### 8.2.9.9 Combining the variables

CTR, BNP, NT-proBNP, presence of pulmonary oedema on chest x-ray and clinical impression all demonstrate sufficient specificity to be used to rule-in heart failure. The presence of any of these individual features has specificity greater than 95%.

Table 8-28 - Summary Table for Individual Variables With High Specificity for Heart Failure

Variable	Sensitivity (95% Confidence Interval)	Specificity (95% Confidence Interval)	Likelihood Ratio Positive	Likelihood Ratio Negative
Cardiothoracic Ratio $\geq$ 0.65	35.1% (23.0 – 49.0%)	95.4% (83.2 99.2%)	7.74	0.68
BNP $\geq$ 600 pg/ml	30.5% (19.5 – 44.0 %)	97.7% (86.7 99.8%)	13.7	0.71
NT-proBNP $\geq$ 500 pg/ml	31.1% (20.2 - 41.4%)	95.4% (83.2 – 99.2%)	6.85	0.72
Pulmonary Oedema	20.0% (11.2 – 32.9%)	95.5% (83.6 – 99.2%)	4.5	0.84
Clinical Impression	38.3% (26.3 – 51.8%)	97.8% (86.7 – 99.8%)	17.25	0.63
Urea $\geq$ 12.8 mg/l	13.3% (6.0 25.0%)	95.6% (85.6 – 99.2%)	3	0.91

These variables can be combined so that for any patients who have any two or more of these variables present, the specificity of this finding approaches 100%. This is illustrated in a STARD (STAndards for the Reporting of Diagnostic accuracy studies) diagram as recommended by the Cochrane group in Figure 8-17.

Table 8-29 Combining Specific Variables (Number of patients with both variables positive / Number of patients diagnosed as not having heart failure)

	CTR	BNP	NT-proBNP	Pulmonary Oedema	Clinical Impression	Urea
CTR		9 / 0	7 / 0	7 / 0	10 / 0	3 / 0
BNP			15 / 1	3 / 0	11 / 0	4 / 0
NT-proBNP				3 / 0	11 / 0	4 / 0
Pulmonary Oedema					5 / 0	0 / 0
Clinical Impression						7 / 0
Urea						

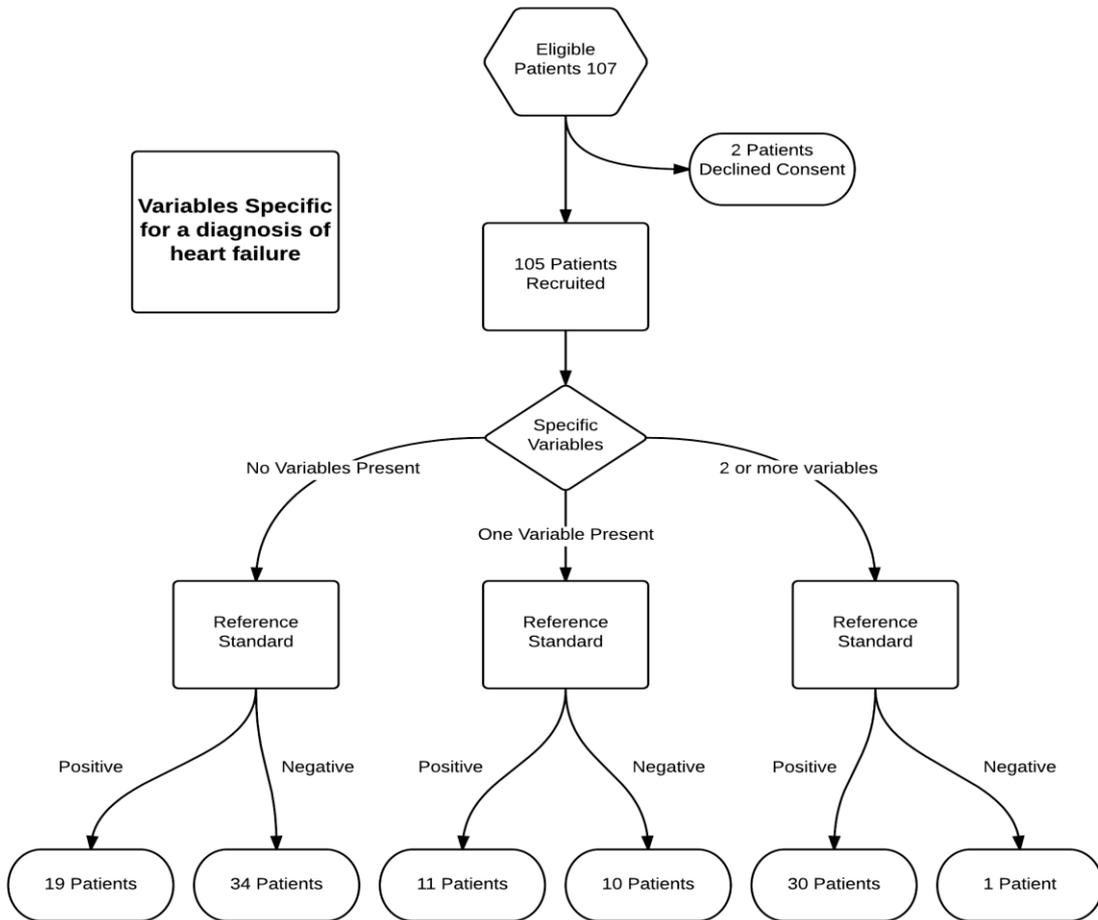


Figure 8-17 STARD Diagram For Patients with Variables Specific For the Diagnosis of Heart Failure

Table 8-30 Summary of Variables with High Sensitivity for the Absence of Heart Failure

Variable	Sensitivity (95% Confidence Interval)	Specificity (95% Confidence Interval)	Likelihood Ratio Positive	Likelihood Ratio Negative
CTR $\leq$ 0.49	37.2% (23.4 – 523.3%)	94.4% (83.6 – 98.6%)	6.7	0.67
Urea $\leq$ 4.5 mg/l	22.2% (11.8 – 37.5%)	95.0% (85.3 – 98.7%)	4.4	0.81
BNP $\leq$ 11.2 pg / ml	13.7 % (5.7 – 28.0%)	94.9% (84.9 – 98.7%)	2.7	0.9
NT-proBNP $\leq$ 20 pg/ml	34.1% (20.9 – 50.0%)	83.3% (71.0 – 91.2%)	2.0	0.8

By selecting low cut-offs for these variables, the specificity for the absence of heart failure approach 95%. Combining these variables meant that there improved this specificity but the numbers of patients with any two of these variable values was very small.

Table 8-31 Combining Variables Sensitive for Ruling Out Heart Failure  
 (Number of patients with both variables positive / Number of patients  
 diagnosed with heart failure)

	CTR ≤ 4.9	Urea ≤ 4.5	NT- proBNP ≤ 20 pg/ml	BNP ≤ 11.2 pg/ml
CTR ≤ 4.9		5 / 0	8 / 2	4 / 1
Urea ≤ 4.5 mg/l			4 / 2	0 / 0
NT-proBNP ≤ 20 pg/ml				6 / 3
BNP ≤ 11.2 pg/ml				

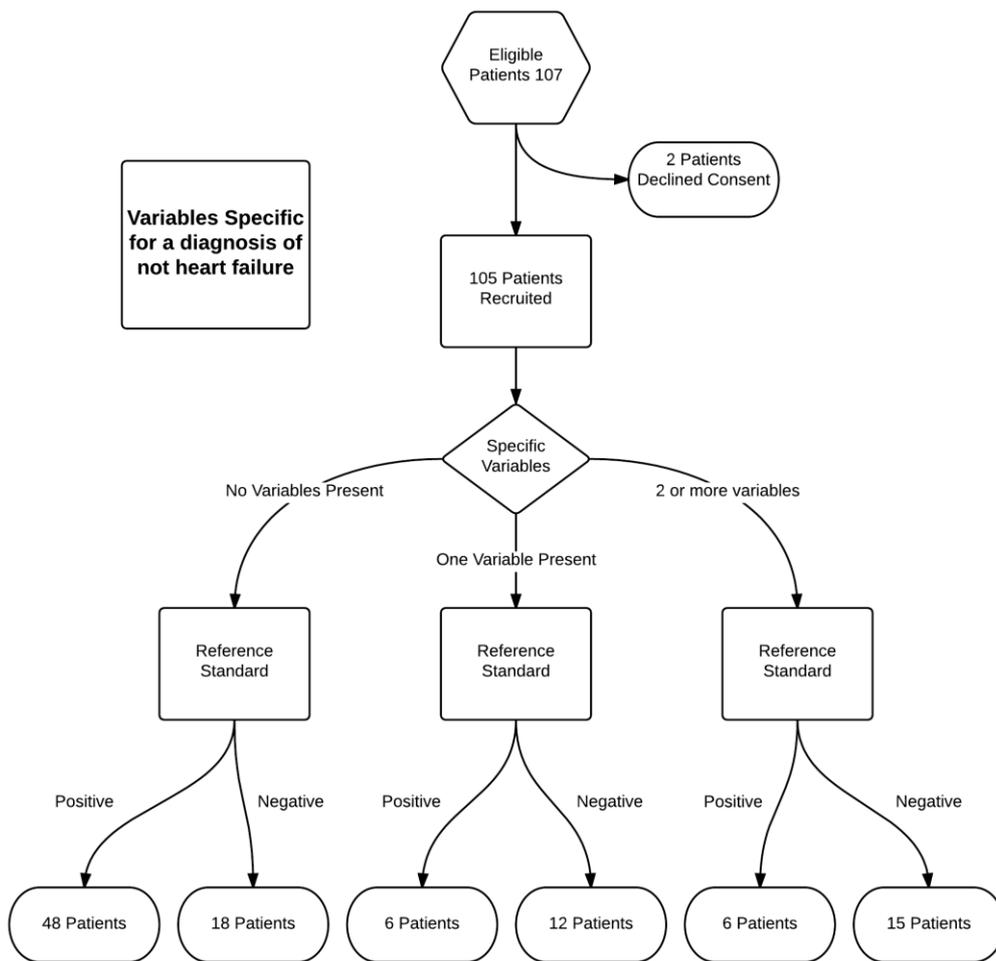


Figure 8-18 STARD Diagram for Patients With Variables Specific for the Diagnosis of Not Heart Failure

### 8.2.10 Summary

Certain variables were found to have the maximal utility whichever method was used to derive them. Other findings, such as the presence of a nocturnal cough, that have been traditionally considered useful in the diagnosis, were found to be of little value. Diagnostic tools using the selected variables performed as well as strategies that included all of the potential variables.

Despite this, it was not possible to derive a decision tree, using either automatic or manual methods that could usefully be derived and applied to this data set to confirm or refute the diagnosis of heart failure. The optimal tree structures synthesised did not adding any significant value to the initial clinical impression.

Other methods to derive a decision tool were used. Logistic regression utilises the full range of continuous variables without the need to dichotomise the results into positive or negative. The Naïve Bayes method shares these advantages and can still be used even if there are unknown variables for the included patients. Both of these methods allow all of the selected variables to inform the diagnosis; one disadvantage of decision trees is that a patient may be diagnosed on the basis of the first node, even if every other result suggests the opposite diagnosis.

Logistic regression resulted in an equation that delivered a good degree of accuracy though was complicated and cumbersome to use. Using the Naïve Bayes method resulted in a similar equation though this could be applied using a nomogram in order to simplify the technique without losing efficacy. Both of these methods were likely to be subject to overfitting as testing the equations using 'Leave-one-out' sampling or splitting the data into derivation and validation samples resulted in much less impressive figures.

Finally, by identifying two or more positive variables that were specific for the presence of heart failure, it was possible to derive a method that approached 100% specificity for this condition. Unfortunately, neither using the variables and cut-off values specific for the presence of heart failure, nor those specific for the absence of heart failure, provided methods sensitive enough to reliably rule out this condition although it could certainly be considered less likely.

## Chapter 9 Discussion And Conclusions

### 9.1.1 Introduction

In the introduction to this thesis, there was exploration of the issues around the diagnosis of acute decompensated heart failure. Evidence was provided to show the significant prevalence of this condition and the devastating impact on the lives of the individuals affected, their families and carers, and the population at large. There was an introduction to the difficulties associated in making this diagnosis; namely that this condition lacks a standardised definition, can present with a wide range of symptoms of variable severity and that there are several potential aetiologies.

In the following chapters, the strengths and limitations of the various modalities used to confirm or refute this diagnosis were addressed and the potential role for natriuretic peptides was presented. The limitations in current practice and potential benefits of an evidence-based, clinical decision rule were raised in Chapter 5.

The main ambition of this thesis was to derive a clinical decision rule that could be applied in the Emergency Department setting and assist clinicians in diagnosing patients with acute, decompensated heart failure. To this end a systematic review of all literature relevant to the diagnosis of heart failure in an Emergency setting was undertaken. Meta-analysis of the data derived from the systematic review was performed to rank and quantify the diagnostic utility of the potential predictor variables assessed.

The second part of this thesis consisted of an original diagnostic study carried out to obtain individual patient data to allow multivariable analysis to be performed. This was to assess the independence of potential diagnostic variables and to select the optimum combination and order of

these to synthesise the optimal clinical decision rule. Various techniques were explored for the selection of the variables, and several methods were used to derive a clinical decision rule.

### 9.1.2 Findings of the Thesis

The systematic review of the literature and the meta-analysis allowed ranking of the performance of all reported potential diagnostic variables in terms of their diagnostic utility. This brought to the fore the diagnostic elements with potential for inclusion in a CDR while allowing the variables with limited utility to be excluded from consideration.

These variables were found to have the greatest diagnostic utility:

- A history of heart failure
- Clinical opinion
- Presence of cardiomegaly on chest x-ray
- Presence of pulmonary oedema on a chest x-ray
- Cephalisation of the vessels on a chest x-ray
- BNP levels
- NT-proBNP levels

Variables with poor predictive value, or very variable reported results included features such as a history of orthopnoea, the presence of peripheral oedema, a history of COPD or a history of ischaemic heart disease.

From the original diagnostic study, analysis using RandomForest software produced a similar selection of variables with the greatest diagnostic value:

- Cardiomegaly on chest x-ray
- Renal function
- NT-proBNP
- BNP

- Clinical opinion
- History of heart failure

The diagnostic utility of the variables was also assessed using univariate logistic regression and correlation analysis that also produced a similar list of variables:

- Use of loop diuretics
- Clinical opinion
- Pulmonary oedema on chest x-ray
- Cardiomegaly on chest x-ray
- Renal function
- BNP

Attempts were made to construct a simple but robust CDR using the data from the original diagnostic study by a variety of methods:

- Multivariable logistic regression
- Scoring system derived from logistic regression
- CART analysis
- Manual generation of a classification tree
- Bayesian analysis
- Combination of individual variables

The above methods were used with all potential variables, the variables selected by the systematic review, the variables selected using logistic regression and the variables selected using the RandomForest analysis.

Decision pathways with excellent diagnostic utility that were synthesised were over-fitted, large, cumbersome and impractical to use. Simpler and more practical rules failed to improve diagnostic utility significantly over the performance of the original components.

The finding of two or more of the individual variables specific for heart failure approached 100% specificity for this condition. Although variables

were found that had high sensitivity for this condition, no combination of these could be produced that could be reliably used to rule this condition out.

### **9.1.3 The Implications of These Findings**

This thesis has demonstrated that certain findings are useful in the rapid assessment of patient with acute dyspnoea when determining if this patient has heart failure. It also highlights some variables that have limited utility despite traditional teaching that would suggest otherwise.

The patient who has acute dyspnoea with severe pulmonary oedema, a raised JVP, a third heart sound and a high BNP does not form a diagnostic quandary; the patient who has ambiguous results for one variable is more likely to have ambiguous findings for all of the variables. This reflects the spectrum of the condition.

This also raises the question of the value of adding natriuretic peptide levels into the diagnostic equation. As this is a new diagnostic test involving laboratory time and equipment there is a cost involved which is not necessarily justified by the marginal diagnostic benefits. Introduction of a novel test also needs explicit guidance to prevent the overuse of the test that can result in over-investigation and over-diagnosis if used inappropriately. And, unlike other investigations such as chest x-ray, the information that the natriuretic peptide levels produce is almost solely related to the presence or absence of cardiac strain.

On the other hand, point-of-care systems allow a rapid result to be available, which is objective and reliable. This makes it easy to incorporate into a diagnostic pathway without making assumptions about the skills or experience of the person interpreting it. Mueller et al. (2006) provided some evidence that the introduction of this test is cost-effective.

#### **9.1.4 Reliability and validity of the selected variables**

There is little published literature to assess the reliability of clinical signs, for cardiac disease, or for any other conditions. Studies that have assessed the presence of the third heart sound or the presence of a raised JVP have found a level of agreement that is as likely to have occurred by chance (Lok et al., 1998; Cowie et al., 1999).

The interpretation of chest x-rays fares little better with poor reliability score (Gatt, 2003). This remains the case even when performed by experienced radiologists (Robinson et al., 1999; Fox et al., 2001). This would suggest an advantage to an objective measurement such as the cardiothoracic ratio although there is the obvious disadvantage that this feature will be relatively fixed and will not vary in the presence of an acute exacerbation.

#### **9.1.5 Strengths and Weakness of the Systematic Review**

An extensive systematic review was performed with a deliberately open literature search. Two readers (CJF & KMJ) reviewed the literature search independently and selected which papers to obtain on the basis of the title and abstract. The selected papers were obtained in electronic or paper form and the quality and relevance assessed. CJF and a panel of seven other readers then assessed papers that were felt to be of sufficient quality, and provide relevant information. Each reader independently extracted the relevant data and made an assessment of the quality of the paper based on the QUADAS format.

Due to the period of time that had elapsed between this literature search and the writing up of this thesis the same search was repeated in 2011 to elicit any further evidence that had been published in the elapsed interval. Due to restrictions in time this evidence has been selected and the data

extracted by a single reader (CJF). Allowing for two independent readers to assess the evidence reduces the risk of selection or interpretation bias by the single reader. Although, new data was added to this thesis, this did not change any of the outcomes or conclusions.

### **9.1.6 Meta-analysis of the Selected Variables**

Due consideration was applied to the methods used in the meta-analysis. Although there is no universal agreement as to the optimum method for novel technique of diagnostic study meta-analysis, there was a significant amount of literature supporting the bivariate method. This method has also been used in other published diagnostic meta-analyses within medical literature (Vasan et al., 2002; Scal, 2005; Willis, 2012). The methods used allowed assessment of the available data and exploration of any heterogeneity found.

Use of bivariate analysis allowed the production of summary data with confidence intervals for the variables with diagnostic potential. The production of this allowed the distinction to be made for the diagnostic variables as to sensitive, specific, both or neither. The most relevant variables were highlighted and other variables without diagnostic value were removed from consideration.

The selection of the most important variables was useful as it confirmed the value of the variables in my original research by independently selecting the same variables and predicting similar diagnostic utility.

### **9.1.7 Strengths of the Original Diagnostic Study**

The original diagnostic study that was performed as part of this thesis did have some positive features. It was a prospective study with appropriate and transparent inclusion and exclusion criteria that were relevant and justified. The study had a reasonable reference standard consisting of the review of all available results by two cardiologists including the recorded

echocardiograms. It also afforded the blinded assessment of both point-of-care BNP, and laboratory-based NT-proBNP measurements in the UK, Emergency department setting.

A wide selection of potentially useful variables were prospectively collected and recorded on a specific form. Follow up data in terms of readmissions and mortality of the patients was also recorded.

### **9.1.8 Weaknesses of the Original Diagnostic Study**

There were also significant weaknesses inherent in this study. The original intention of the study was to be at least twice as large but this was not possible due to two reasons. There was a delay in commencing the study while the department procured a suitable echocardiogram machine. The second issue was due to a national restructuring of the training system for Emergency Medicine that meant the lead investigator had to take on full-time clinical commitments. The research position was then continued on a part-time basis around a busy full-time clinical job.

A major shortcoming of this study is that it was a single site study. This was necessary for pragmatic reasons given that the research project was dependent on a single investigator with clinical commitments. In the event of developing a useful clinical decision rule, validation should be performed at multiple centres separate to the one that was involved in the derivation study.

The recruitment of patients was done on a convenience basis. This was for pragmatic reasons as the study relied on the principal investigator being available and free of clinical duties to perform the echocardiography and to handle the preparation of the blood samples. This introduces the potential of selection bias although every effort to recruit every potential patient possible. The majority of patients were obtained during a two-month research secondment when the lead investigator was free of clinical duties.

Another opportunity to strengthen this study would have been to use a more experienced technician or cardiologist to record the echocardiograms. Again, a pragmatic reason was behind this although it was also useful to assess the value of training an emergency physician with this skill.

### **9.1.9 Development of the Clinical Decision Rule**

Attempts to derive a simple but robust clinical decision rule to aid the diagnosis of heart failure in the acutely breathless patient in the Emergency Department setting revealed that it was difficult to do. The condition is common, has serious consequences and the current diagnostic process is variable and based on little evidence, suggesting that a clinical decision rule would have been useful. However, the end diagnosis and most of the diagnostic tests involved in obtaining it incorporate some degree of subjectivity that made defining consistent rules challenging.

The only way to achieve very high sensitivity or specificity using a clinical decision rule was by deriving a large, cumbersome decision system that was demonstrably over-fitted to the available data. The alternative was to provide a selection of specific variables, the finding of two or more being associated with a very high specificity for this condition.

While there are certainly weaknesses with the study that was used to attempt to derive the rule, as have been discussed in the previous section, I do not think that these reasons are the primary reason behind the failure to derive the rule for the following reasons:

1. The variables with the greatest diagnostic utility derived from the original study were similar to those found from the systematic review of the medical literature.
2. The reference standard was reasonable.

3. The diagnostic performances of the selected variables were similar in both the original research and the results of the meta-analyses.
4. The number of patients with the target condition was adequate to derive a decision rule with a selected number of variables included in the model.
5. Application of the other published scoring systems for heart failure to the original data did not provide any better results than the diagnostic strategies synthesised within this thesis.

There are several factors with the condition of acute, decompensated heart failure that make the derivation of a clinical decision rule in the emergency setting difficult to the degree of unlikely:

- a. There is no agreed gold standard for this condition.
- b. The mechanism for decompensation of heart failure remains poorly understood and it is likely that there is more than one.
- c. The onset of heart failure tends to be insidious rather than a fixed time-event so dynamic changes in biological markers are less useful.
- d. The subjectivity in the diagnosis and the interpretation of many of the diagnostic findings.
- e. The association between heart failure and age raises the difficulty in distinguishing normal findings in an ageing human from a diseased state.

### **9.1.10 Recommendations for Further Research**

Although the concentration of natriuretic peptides was could be applied to cut-offs as a diagnostic tool, it did not significantly outperform other variables, including clinical opinion. However, there may still be clinical benefits if the use of natriuretic peptides improves the speed of diagnosis and therefore reduces the time to commencement of appropriate treatment.

Clinicians may also be positively guided in the interpretation of more subjective investigations, such as chest x-ray, in the light of known natriuretic peptide levels. This increased confidence may influence the speed and aggression with which clinicians treat this condition.

A prospective diagnostic trial in the UK setting where patients were randomised to either have, or not have, a natriuretic peptide level measured could provide answers to these queries. The cost-effectiveness of introducing this into clinical practice could also be assessed.

### **9.1.11 Conclusion**

In the diagnosis of acute decompensated heart failure, chest x-ray findings, natriuretic peptides and clinical opinion have been verified as providing the optimal diagnostic role in making this diagnosis. Other variables, some of which have been traditionally taught as being consistent with heart failure have been shown to have little diagnostic utility.

It has been shown that the combination of certain specific variables can be considered pathognomonic for heart failure though ruling the condition out appeared more difficult despite some of the variables demonstrating high sensitivity. Although natriuretic peptides demonstrated diagnostic worth, they did not improve the diagnosis making process beyond the combination of other variables. Further evidence is of improvement in clinical outcomes would be required before recommending the routine use of natriuretic peptides in the diagnosis of this condition.

# Chapter 10 Appendices

## 10.1 Appendix I - Literature Search

### 10.1.1 EMBASE Search

Database: EMBASE <1980 to 2009 Week 35>

Search Strategy:

```
-----  
-----  
1      medical history taking.mp. or exp anamnesis/ (106585)  
2      (clinical adj3 assessment).af. (33653)  
3      (signs adj3 symptoms).af. (29489)  
4      (clinical adj3 evaluation).af. (37241)  
5      (clinical adj3 examin$).af. (74107)  
6      (clinical adj3 finding$).af. (39521)  
7      (physical adj3 evaluat$).af. (2969)  
8      (physical adj3 finding$).af. (5062)  
9      (physical adj3 assess$).af. (4055)  
10     (aetiological adj4 features).af. (70)  
11     (etiological adj4 features).af. (131)  
12     xray.af. (213)  
13     x-ray.af. (144097)  
14     radiogra$.af. (186110)  
15     roentgenogra$.af. (8432)  
16     (medical adj3 history).af. (11760)  
17     (physical adj3 exam$).af. (83665)  
18     electrocardiogra$.af. (90138)  
19     ecg.af. (31088)  
20     ekg.af. (1834)  
21     (natriuretic adj peptide$).af. (16359)  
22     NTproBNP.af. (74)  
23     NT-proBNP.af. (1206)  
24     NTpro-BNP.af. (9)  
25     BNP.af. (4120)  
26     (biological adj3 marker).af. (36305)  
27     1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11  
or 12 or 13 or 14 or 15 or 16 or 17 or 18 or 19 or 20 or 21  
or 22 or 23 or 24 or 25 or 26 (773600)  
28     acute$.af. (576780)  
29     emergen$.af. (171867)  
30     urgent.af. (17360)  
31     decompensat$.af. (7586)  
32     sudden.af. (48954)  
33     or/28-32 (773709)  
34     (heart adj3 fail$).af. (111653)  
35     (heart adj3 dysfunction$).af. (1968)  
36     (heart adj3 insufficien$).af. (620)  
37     (cardiac adj3 insufficien$).af. (1541)  
38     (cardiac adj3 fail$).af. (10043)  
39     (cardiac adj3 dysfunction$).af. (5141)
```

40 (ventric\$ adj3 dysfunction\$).af. (12614)  
 41 (ventric\$ adj3 fail\$).af. (15799)  
 42 (ventric\$ adj3 insufficien\$).af. (288)  
 43 (systolic adj3 dysfunction).af. (4242)  
 44 (systolic adj3 fail\$).tw. (1286)  
 45 (diastolic adj3 fail\$).af. (1005)  
 46 (diastolic adj3 dysfunction).af. (3925)  
 47 hf.af. (11728)  
 48 lvf.af. (427)  
 49 adhf.af. (105)  
 50 adcf.af. (4)  
 51 ccf.af. (688)  
 52 chf.af. (7792)  
 53 (pulmonary adj3 edema).af. (7466)  
 54 (pulmonary adj3 oedema).af. (1912)  
 55 breathless\$.af. (2180)  
 56 dyspnoea.af. (4411)  
 57 dyspnea.af. (41738)  
 58 (short\$ adj3 breath).af. (2938)  
 59 sob.af. (270)  
 60 or/34-59 (184624)  
 61 27 and 60 and 33 (11802)  
 62 di.fs. (1433897)  
 63 predict\$.tw. (510452)  
 64 specificity.tw. (195785)  
 65 or/62-64 (1975218)  
 66 61 and 65 (7074)  
 67 limit 66 to human (6909)

## 10.1.2 Ovid MEDLINE Search

Database: Ovid MEDLINE(R) <1950 to August Week 4 2009>

Search Strategy:

---

1 medical history taking.mp. or exp Medical History Taking/ (15741)  
 2 (physical adj3 exam\$).af. (55686)  
 3 (clinical adj3 assessment).af. (18036)  
 4 (signs adj3 symptoms).af. (36461)  
 5 (clinical adj3 evaluation).af. (39223)  
 6 (clinical adj3 examination).af. (27277)  
 7 (clinical adj3 findings).af. (45839)  
 8 (physical adj3 evaluation).af. (1748)  
 9 (physical adj3 findings).af. (5685)  
 10 (physical adj3 assessment).af. (1910)  
 11 aetiological features.af. (26)  
 12 etiological features.af. (47)  
 13 xray.af. or exp X-Rays/ (13757)  
 14 radiogra\$.af. (201299)  
 15 roentgenogra\$.af. (16905)  
 16 electrocardiogra\$.af. (169539)  
 17 ecg.af. (37470)

18 ekg.af. (4219)  
19 natriuretic peptide.mp. or exp Natriuretic Peptides/ (21237)  
20 NTpro-bnp.af. (8)  
21 NT-proBNP.af. (1191)  
22 exp Natriuretic Peptide, Brain/ or BNP.af. or exp Biological Markers/  
(448302)  
23 or/1-22 (1057279)  
24 (heart adj3 fail\$.af. (101414)  
25 (heart adj3 dysfunction\$.af. (1964)  
26 (heart adj3 insufficien\$.af. (1251)  
27 (cardiac adj3 dysfunction\$.af. (5506)  
28 (cardiac adj3 insufficien\$.af. (4023)  
29 (cardiac adj3 fail\$.af. (13517)  
30 (ventric\$ adj3 dysfunction\$.af. (25378)  
31 (ventric\$ adj3 fail\$.af. (6272)  
32 (ventric\$ adj3 insufficien\$.af. (598)  
33 (systolic adj3 fail\$.af. (1415)  
34 (systolic adj3 dysfunction).af. (4079)  
35 (diastolic adj3 fail\$.af. (989)  
36 (diastolic adj3 dysfunction).af. (3861)  
37 hf.af. (11656)  
38 lvf.af. (494)  
39 adhf.af. (172)  
40 adcf.af. (2)  
41 ccf.af. (709)  
42 chf.af. (7880)  
43 (pulmonary adj3 oedema).af. (2436)  
44 (pulmonary adj3 edema).af. (17607)  
45 breathless\$.af. (2394)  
46 dyspnea.af. (24763)  
47 dyspnoea.af. (4926)  
48 (short\$ adj3 breath).af. (3253)  
49 sob.af. (446)  
50 or/24-49 (188573)  
51 (acute or emergen\$ or urgen\$ or decompensat\$ or sudden).af. (990218)  
52 50 and 51 and 23 (10634)  
53 limit 52 to "diagnosis (sensitivity)" (5633)

## 10.2 Appendix II - QUADAS questions

- 1 Was the spectrum of patients representative of the patients who will receive the test in practice? (spectrum)
  - 2 Were selection criteria clearly described? (selection criteria)
  - 3 Is the reference standard likely to correctly classify the target condition? (reference standard)\*
  - 4 Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests? (disease progression bias)\*
  - 5 Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis? (partial verification)
  - 6 Did patients receive the same reference standard regardless of the index test result? (differential verification)
  - 7 Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)? (incorporation bias)
  - 8 Was the execution of the index test described in sufficient detail to permit replication of the test? (index test execution)
  - 9 Was the execution of the reference standard described in sufficient detail to permit its replication? (reference standard execution)
  - 10 Were the index test results interpreted without knowledge of the results of the reference standard? (test review bias)
  - 11 Were the reference standard results interpreted without knowledge of the results of the index test? (reference standard review bias)
  - 12 Were the same clinical data available when test results were interpreted as would be available when the test is used in practice? (clinical review bias)\*
  - 13 Were uninterpretable/ intermediate test results reported? (uninterpretable test results)
  - 14 Were withdrawals from the study explained? (withdrawals)
- (Whiting et al., 2006)

## 10.3 Appendix III – Ethical Approval

Bolton Local Research Ethics Committee  
Room 181, Gateway House  
Piccadilly South  
Manchester M60 7LP

Telephone: 01612372585

Facsimile: 01612372383

12 September 2005

Private & Confidential

Dr C J Ferguson, Clinical Research Fellow  
Department of Emergency Medicine  
Manchester Royal Infirmary  
Oxford Road  
MANCHESTER  
M13 9WL

Dear Dr Ferguson

Full title of study: Creation of a Clinical Decision Rule for the Diagnosis  
and Management of Heart Failure within Emergency  
Medicine involving the use of B Natriuretic Peptide

REC reference number: 05/Q1409/53

Thank you for your response to the Committee's request for further  
information on the above research and for submitting revised  
documentation.

The further information has been considered on behalf of the Committee  
by the Vice Chair.

## Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion for the above research on the basis described in the application form, protocol and supporting documentation as revised.

However, the Committee has not yet been notified of the outcome of any site-specific assessment (SSA) for the research site taking part in this study. The favourable opinion does not therefore apply to any site at present. I will write to you again as soon as one Local Research Ethics Committee has notified the outcome of a SSA. In the meantime no study procedures should be initiated at sites requiring SSA.

## Conditions of approval

The favourable opinion is given provided that you comply with the conditions set out in the attached document. You are advised to study the conditions carefully.

## Approved documents

The final list of documents reviewed and approved by the Committee is as follows:

Document	Version	Date
Application	4.1	23 May 2005
Investigator CV K Mackway-Jones		20 May 2005
Investigator CV C J Ferguson		20 May 2005
Protocol		20 May 2005
Summary/Synopsis		20 May 2005
GP/Consultant Information Sheets		12 May 2005

Participant Information Sheet		02 September 2005
Participant Consent Form Gifting	3	02 September 2005
Participant Consent Form	3	

#### Management approval

The study should not commence at any NHS site until the local Principal Investigator has obtained final management approval from the R&D Department for the relevant NHS care organisation.

#### Membership of the Committee

The members of the Ethics Committee who were present at the meeting are listed on the attached sheet.

#### Notification of other bodies

The Committee Administrator will notify the research sponsor that the study has a favourable ethical opinion.

#### Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees (July 2001) and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

05/Q1409/53  
correspondence

Please quote this number on all

With the Committee's best wishes for the success of this project,

Yours sincerely

Chair

Email: [elaine.hutchings@gmsa.nhs.uk](mailto:elaine.hutchings@gmsa.nhs.uk)

## 10.4 Appendix IV – Consent Form

REC Ref: 05/Q14093/53 Version 3 2/9/5

Patient Identification Number for this trial:

### CONSENT FORM

**Title of Project: Creation of a CDR for the Diagnosis and Management of Heart Failure within Emergency Medicine involving the use of BNP**

Name of Researcher: Craig Ferguson

**Please initial box**

1. I agree that the blood sample taken from me and information gathered about me may be stored by the research team in an anonymous manner so that it cannot be traced back to me.

2. I understand that this blood sample taken from me may be kept and used in further research related to the diagnosis of heart failure

\_\_\_\_\_  
Name of Patient

\_\_\_\_\_  
Date

\_\_\_\_\_  
Signature

\_\_\_\_\_  
Name of Person taking consent  
(if different from researcher)

\_\_\_\_\_  
Date

\_\_\_\_\_  
Signature

\_\_\_\_\_  
Researcher

\_\_\_\_\_  
Date

\_\_\_\_\_  
Signature

1 for patient; 1 for researcher; 1 to be kept with hospital notes

Patient Identification Number for this trial:

## CONSENT FORM

**Title of Project: Creation of a CDR for the Diagnosis and Management of Heart Failure within Emergency Medicine involving the use of BNP**

Name of Researcher: Craig Ferguson

Please initial box

1. I confirm that I have read and understand the information sheet dated .....   
for the above study and have had the opportunity to ask questions.
2. I understand that my participation is voluntary and that I am free to withdraw at any time without giving any reason, without my medical care or legal rights being affected.
3. I understand that sections of any of my medical notes may be looked at by responsible individuals where it is relevant to my taking part in research. I give permission for these individuals to have access to my records.
4. I agree that the blood sample taken from me and the information gathered about me maybe stored by the research team in an anonymous manner so that it cannot be traced back to me. I understand that this sample may be used in the future for further research related to the diagnosis of heart failure
5. I agree to take part in the above study.

\_\_\_\_\_  
Name of Patient

\_\_\_\_\_  
Date

\_\_\_\_\_  
Signature

\_\_\_\_\_  
Name of Person taking consent  
(if different from researcher)

\_\_\_\_\_  
Date

\_\_\_\_\_  
Signature

\_\_\_\_\_  
Researcher

\_\_\_\_\_  
Date

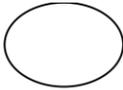
\_\_\_\_\_  
Signature

1 for patient; 1 for researcher; 1 to be kept with hospital notes

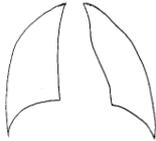
## 10.5 Appendix V – Clinical Details Form

# Breathless Patient Form

Attending Doctor: \_\_\_\_\_  
Date & Time: \_\_\_\_\_

  
Study Number

<b>Inclusion Criteria</b> <input type="checkbox"/> Main complaint is shortness of breath <input type="checkbox"/> No history of trauma <input type="checkbox"/> Not receiving dialysis <input type="checkbox"/> Able to consent		<b>Patient Details</b> Name: _____ Sex: M / F Date of Birth: _____ A&E Number: MAE																												
<b>History</b>          Duration of breathlessness: _____ Chest pain <input type="checkbox"/> Haemoptysis <input type="checkbox"/> Orthopnoea <input type="checkbox"/> PND <input type="checkbox"/> Cough <input type="checkbox"/> Sputum <input type="checkbox"/> Wheeze <input type="checkbox"/> Other symptoms: _____																														
<b>Past Medical History</b> <table border="0"> <tr> <td>Anaemia</td> <td>Heart failure</td> <td>Angina</td> </tr> <tr> <td>Heart valve prosthesis</td> <td>CVA</td> <td>Myocardial Infarction</td> </tr> <tr> <td>CABG</td> <td>Peripheral Vascular Disease</td> <td>Cancer</td> </tr> <tr> <td>Pacemaker</td> <td>COPD</td> <td>PTCA</td> </tr> <tr> <td>CRF</td> <td>Pulmonary Embolism</td> <td>Current smoker</td> </tr> <tr> <td>Seizure</td> <td>Diabetes</td> <td>Syncope</td> </tr> <tr> <td>DVT</td> <td>TIA</td> <td>Former smoker</td> </tr> <tr> <td>Other _____</td> <td></td> <td></td> </tr> <tr> <td>_____</td> <td></td> <td></td> </tr> </table>				Anaemia	Heart failure	Angina	Heart valve prosthesis	CVA	Myocardial Infarction	CABG	Peripheral Vascular Disease	Cancer	Pacemaker	COPD	PTCA	CRF	Pulmonary Embolism	Current smoker	Seizure	Diabetes	Syncope	DVT	TIA	Former smoker	Other _____			_____		
Anaemia	Heart failure	Angina																												
Heart valve prosthesis	CVA	Myocardial Infarction																												
CABG	Peripheral Vascular Disease	Cancer																												
Pacemaker	COPD	PTCA																												
CRF	Pulmonary Embolism	Current smoker																												
Seizure	Diabetes	Syncope																												
DVT	TIA	Former smoker																												
Other _____																														
_____																														
<b>Current Medication</b> <table border="0"> <tr> <td>ACE inhibitor _____</td> <td>Becotide _____</td> <td>Amiodarone _____</td> </tr> <tr> <td>Salbutamol _____</td> <td>Aminophylline _____</td> <td>Atrovent _____</td> </tr> <tr> <td>Aspirin _____</td> <td>Warfarin _____</td> <td>Beta-blocker _____</td> </tr> <tr> <td>NSAIDs _____</td> <td>Calcium antagonist _____</td> <td>Home oxygen _____</td> </tr> <tr> <td>Diuretic _____</td> <td>Digoxin _____</td> <td></td> </tr> <tr> <td>Others _____</td> <td></td> <td></td> </tr> </table>				ACE inhibitor _____	Becotide _____	Amiodarone _____	Salbutamol _____	Aminophylline _____	Atrovent _____	Aspirin _____	Warfarin _____	Beta-blocker _____	NSAIDs _____	Calcium antagonist _____	Home oxygen _____	Diuretic _____	Digoxin _____		Others _____											
ACE inhibitor _____	Becotide _____	Amiodarone _____																												
Salbutamol _____	Aminophylline _____	Atrovent _____																												
Aspirin _____	Warfarin _____	Beta-blocker _____																												
NSAIDs _____	Calcium antagonist _____	Home oxygen _____																												
Diuretic _____	Digoxin _____																													
Others _____																														
<b>Allergies:</b> _____ _____																														
<b>Social History</b> _____ _____		<b>Family History</b> _____ _____																												

<b>Examination Findings</b>		
Oxygen Saturation: ___ %	Respiratory Rate: ___ bpm	
Temperature: ___ C	Expansion= Auscultation:	
Pulse: ___ bpm		
Blood pressure: ___ / ___ mmHg		
JVP: HS=		
Presence of cyanosis: <input type="checkbox"/>		Abdominal examination
Presence of sweating: <input type="checkbox"/>		
Peripheral oedema:		

<b>Treatment Given</b>			
Oxygen <input type="checkbox"/>	Diuretics <input type="checkbox"/>	Buccal nitrates <input type="checkbox"/>	Intravenous nitrates <input type="checkbox"/>
Opiates <input type="checkbox"/>	Aminophylline <input type="checkbox"/>	Magnesium Sulphate <input type="checkbox"/>	Nebulisers <input type="checkbox"/>
Antibiotics <input type="checkbox"/>	Steroids <input type="checkbox"/>	CPAP <input type="checkbox"/>	BiPAP <input type="checkbox"/>
Others _____			

<b>Investigations</b>						
FBC <input type="checkbox"/>	U+Es <input type="checkbox"/>	CRP <input type="checkbox"/>	Glucose <input type="checkbox"/>	Blood cultures <input type="checkbox"/>	ABGs <input type="checkbox"/>	INR <input type="checkbox"/>
CXR <input type="checkbox"/>	ECG <input type="checkbox"/>	V/Q scan <input type="checkbox"/>	CT <input type="checkbox"/>	Peak Flow: _____		

If the patient is well enough, ask the patient to mark how breathless they feel on this line:

Not Breathless
→
Extremely Breathless

If this is not possible ask the how breathless the patient feels on a scale from 1 to 10 where 1 is not breathless at all and 10 is completely breathless.

1   2   3   4   5   6   7   8   9   10

<b>ECG Interpretation</b>	
The ECG is normal sinus rhythm <input type="checkbox"/>	The ECG has the following abnormalities:
AF <input type="checkbox"/>	ST elevation <input type="checkbox"/>
ST depression <input type="checkbox"/>	LBBB <input type="checkbox"/>
LVH <input type="checkbox"/>	T wave inversion <input type="checkbox"/>
Q waves <input type="checkbox"/>	
Other findings _____	

<b>CXR Interpretation</b>	
Normal <input type="checkbox"/>	Compatible with heart failure <input type="checkbox"/>
	Other lung pathology <input type="checkbox"/>

<b>Clinical Impression</b>
This patient definitely has heart failure / may have heart failure / does not have heart failure.

<b>Outcome</b>	
Medical referral	<input type="checkbox"/>
Other speciality referral : _____	<input type="checkbox"/>
Discharge with clinic appointment	<input type="checkbox"/>
Discharge with no follow up arranged	<input type="checkbox"/>
Other:	

## 10.6 Appendix VI – Characteristics of the Natriuretic Peptides

BNP and NT-proBNP have skewed distribution in common with many biological variables. The log transformation of these values does have normal distribution.

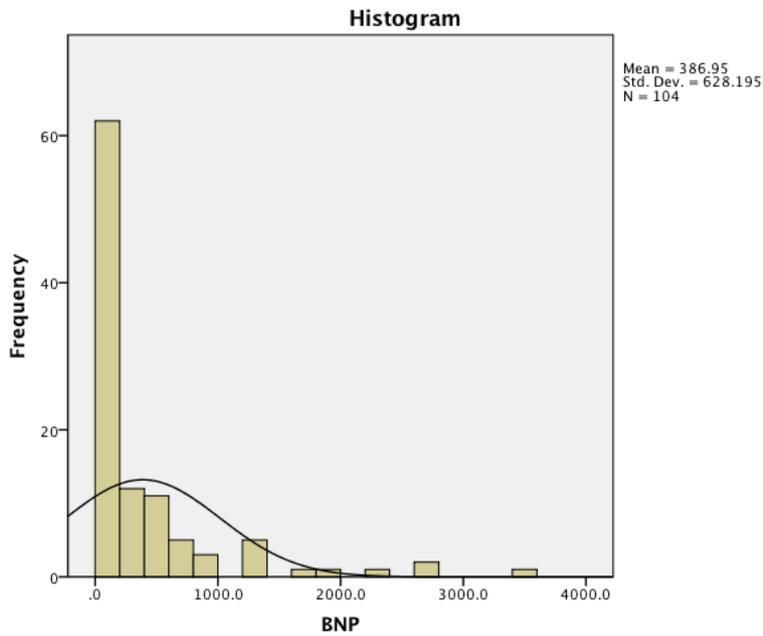


Figure 10-1 Frequency Histogram of BNP

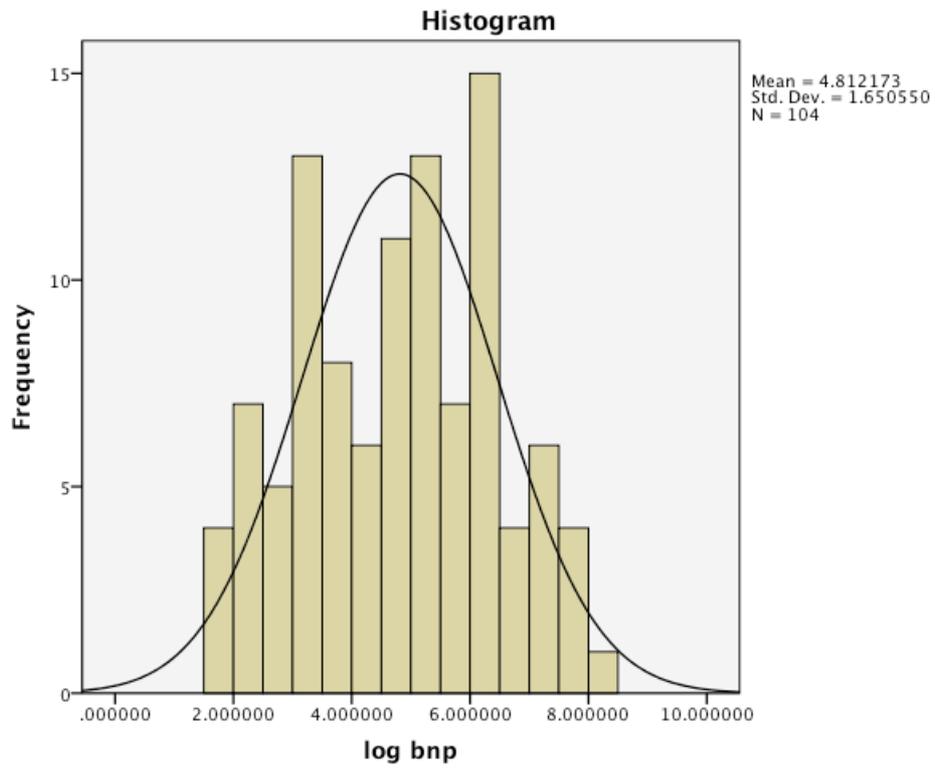
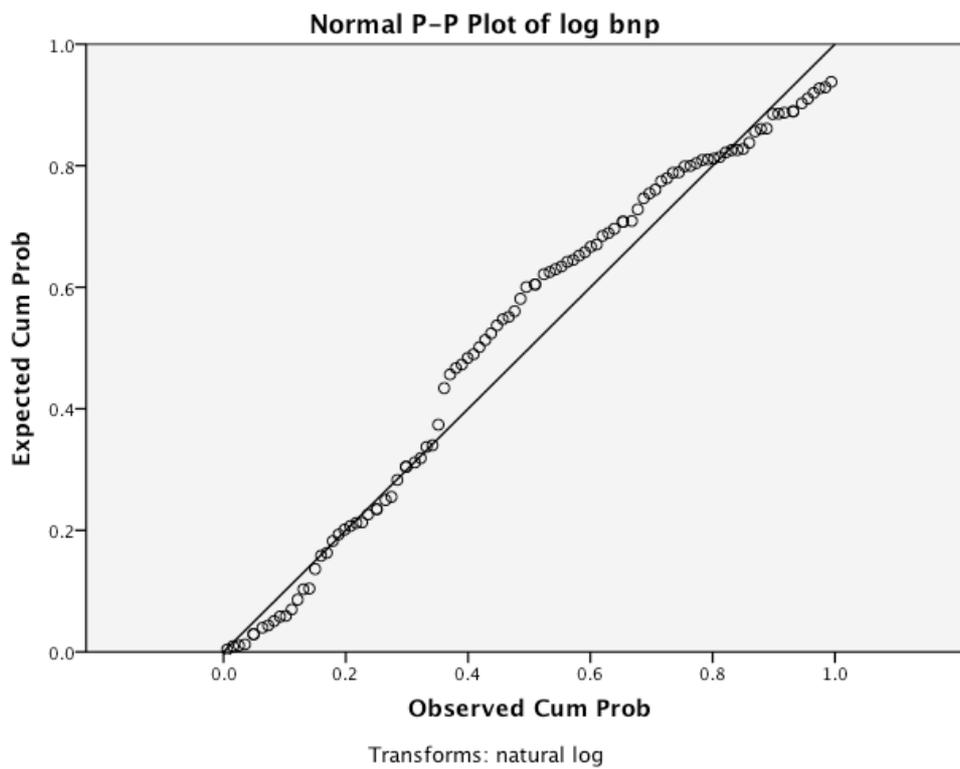


Figure 10-2 Frequency Histogram of Log BNP



### Figure 10-3 Normal P-P Plot for Log BNP

BNP and NT-proBNP demonstrate correlation with a Pearson Correlation value of 0.951 with a significance of 0.000.

BNP correlates with age, correlation coefficient of 0.424 with 2-tailed significance of .000, considered significant at the 0.01 level.

I did not find any significant difference in mean BNP values dependent on sex using the Mann-Whitney U test (Significance = 0.251). The same was true for NT-proBNP with a significance level of 0.151.

There was a significant correlation for BNP with age (Spearman's Rho Correlation Coefficient of 0.424 and a significance of 0.000. The same was true with NT-proBNP CE of 0.474 and significance of 0.000.

There was a negative correlation for NT-proBNP with estimated GFR with a CC of -0.429 and a significance of 0.000. For BNP this was -0.459 and 0.000.

The values of NT-proBNP were significantly higher in the patients who subsequently died using the Mann-Whitney U test (significance .006). The same was true of BNP with a significance level of 0.022.

For male patients the mean BNP level was 475.3 (95% CI 255 – 695) with a median value of 166.000. For female patients the mean level was 324.7 (95% CI 182 – 467) with a median of 95.2.

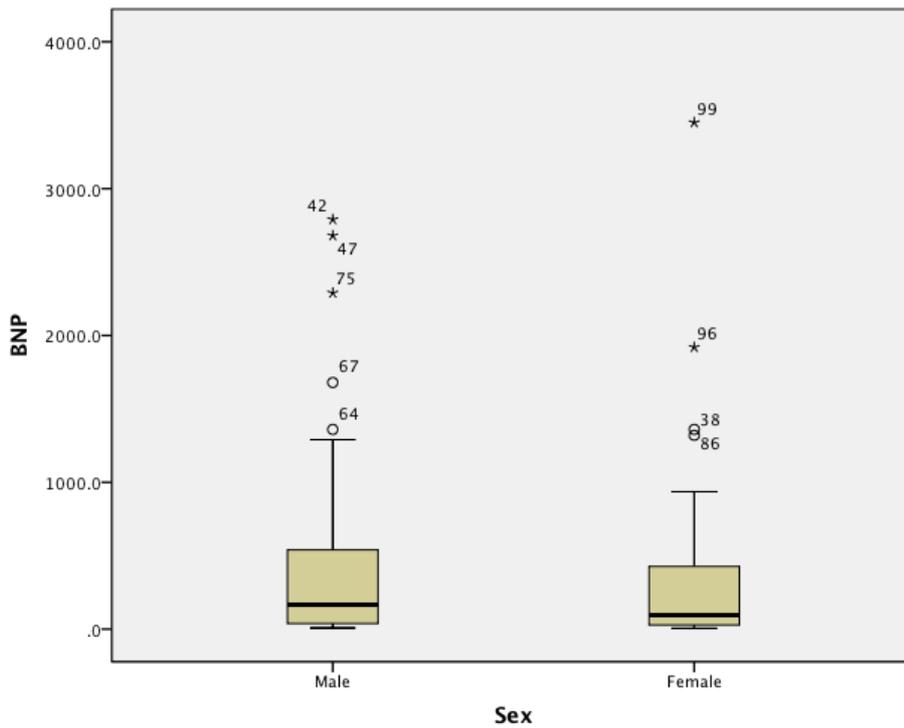


Figure 10-4 BNP Values by Sex of Patients

For male patients the mean NT-proBNP level was 616.2 (95% CI 246 – 986) with a median value of 124.7.

For female patients the mean NT-proBNP levels was 361.3 (95% CI 150 – 572) with a median value of 56.6.

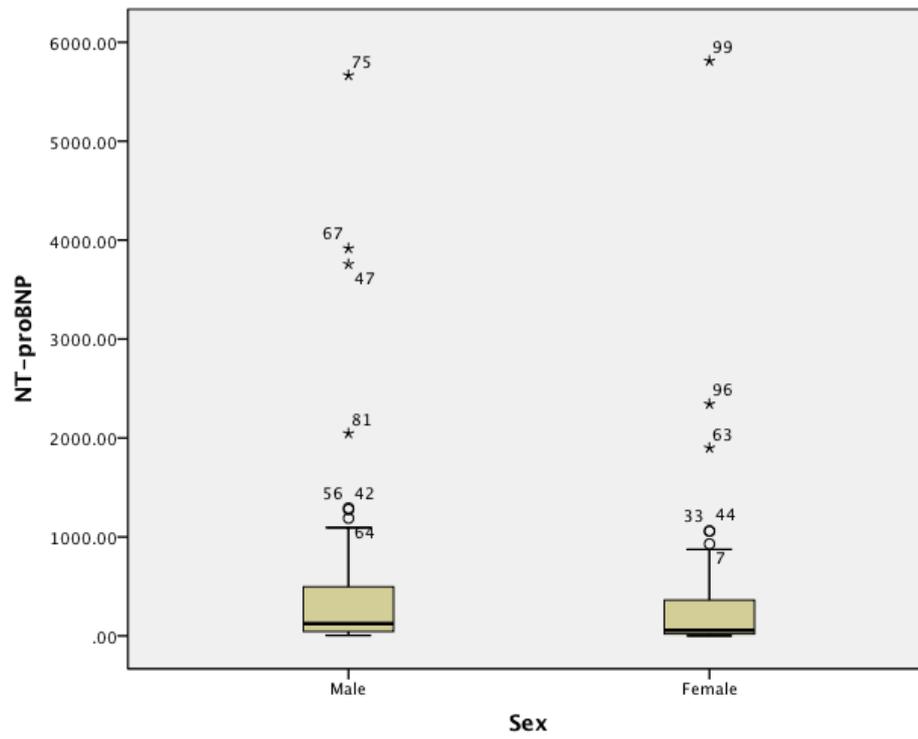


Figure 10-5 NT-proBNP Values by Patient Sex

## 10.7 Appendix VI – Summary of Included Studies

### Ababsa 2005

**Clinical Features** - Acute dyspnoea as the primary complaint.

**Setting** - A single emergency department in Pontoise, France. Time period not specified.

**Participants** - 86 adult patients with mean age of 66 years (Range 38 to 94 years). 36 female: 50 male patients. Excluded patients with severe renal impairment.

**Study Design** - Selection method for patients not clear.

**Target condition** - Heart failure.

**Reference standard** - The opinion of two cardiologists. Mentions that treating clinicians were unaware of BNP results, not clear if cardiologists blinded.

**Index and comparator tests** - BNP was measured using AxSYM BNP kit by Abbot laboratories. All specimens assayed immediately on arrival at laboratory. Used results to determine optimal levels to rule-in and rule-out heart failure.

**Follow up** - Not mentioned.

**Funding** - Abbot Laboratories provided BNP assay kits.

**Notes** - Data provided as sensitivity, specificity, PPV, NPV and +LR for certain cut-off values. I have calculated best fitting numbers but it does not match the figures exactly. Mentions that treating clinicians were unaware of BNP results - not clear if cardiologists blinded.

### Alibay 2005

**Clinical features** - Patients presenting with acute dyspnoea

**Setting** - A single emergency department in Paris, France between June 2003 and January 2004.

**Participants** - 160 patients. Mean age  $80.1 \pm 13.5$  years (range 16-98 years). 84 female: 76 male patients. 60 patients diagnosed with heart failure.

**Study design** - Convenience sample?

**Target condition** - Heart failure.

**Reference standard** - Two cardiologists using all data including ECG, echocardiography, CXR and response to treatment based on the European Cardiology Society (ESC) guidelines.

**Index and comparator tests** - BNP levels were measured. The coefficient variation was 10.3% at low end, 29.1 pg/ml and 15.2% at high end, 5,000pg/ml. Biosite Kit. NT-proBNP was also used. The coefficient variation was 5% at 380pg/ml, 4.4% at 8700 pg/ml and 5% at 13,000pg/ml. Roche Elecsys.

**Follow up** - Not clear.

**Funding** - Not stated.

**Notes** - Little demographic data provided.

## **Barcarse 2004**

**Clinical features** - Adult patients with acute dyspnoea. Excluded patients who clearly had another cause of dyspnoea.

**Settings** - Patients presenting to the Veterans Association Hospital, San Diego, USA between June 2001 and April 2002.

**Participants** - 98 male patients. Mean age  $64.6 \pm 1.2y$ .

**Study design** - Convenience sample. All patients had BNP test done followed by non-invasive haemodynamic monitoring by Impedance Cardiography. The attending physician was told if the BNP level was greater than 100pg/ml upon request. Biosite kit.

**Target condition** - Heart failure.

**Reference standard** - The reference standard was a single cardiologist who reviewed all the patient data apart from the ICG result and determined if the patient had heart failure.

**Index and comparator tests**- Impedance cardiography.

**Follow up** - Patients were followed up for 90 days with regard to cardiac deaths, readmissions, ED visits within 90 days in patients with BNP >100pg/ml.

**Disclosures** - Kits for BNP testing provided by Biosite Inc. BioZ device loaned by Cardiodynamics for ICG monitoring. Dr Maisel is a consultant for Biosite Inc.

**Notes** - Compared BNP and CI by using ROC curves.

## **Bayes-Genis 2004**

**Clinical features** - Adult patients presenting with shortness of breath as the most prominent complaint.

**Settings** - One emergency department at a hospital in Barcelona, Spain between April 2002 and February 2003.

**Participants** - 100 patients recruited, six died before diagnosis confirmed and five withdrew from the study. Mean age 71 years, range 40-80 years. 54 males: 35 females.

**Study design** - Prospective study with consecutive recruitment. Bloods were taken for NT-proBNP testing on arrival.

**Target condition** - Heart failure.

**Reference standard** - The reference standard was two cardiologists opinions, following review of the patient notes. The cardiologists were blinded to the NT-proBNP results.

**Index and comparator tests**- Patients had an echocardiogram as an index test, spirometry, pulmonary volumes and blood gases. Left ventricular dilatation (LVEDD>55mm) and LVEF <45% were used as criteria for systolic LV dysfunction. Left ventricular hypertrophy and LVEF >45% with inversion or pseudo-normalisation of E/A ratio identified diastolic LV dysfunction.

**Follow up** - Patients reviewed at 7 days after admission. 6 patients died before confirmation of diagnosis and 5 patients withdrew from the study.

**Disclosures:** Drs Bayés-Genis, Santaló-Bel, Zapico-Muñiz and Ofdoñez-Llanos received honoraria from Roche Diagnostics for conferences. Roche Diagnostic provided the reagents for the study.

**Notes** - Stated that recruited consecutive patients but only 100 patients recruited in 11 months.

## **Boldanova 2010**

Further write up of Mueller 2005 paper but includes data about the diagnostic accuracy of the clinical diagnosis of heart failure and BNP while recorded in the emergency department.

## **Choi 2007**

**Clinical features** - Adult patients presenting because of dyspnoea. Excluded patients with history of trauma, cardiopulmonary resuscitation or patients who had been transferred from other hospitals.

**Settings** - Emergency departments in Korea from October 2004 to October 2005.

**Participants** - 1040 patients. 582 male: 458 female. Mean age 63 years,

**Study design** – consecutive? Prospective study.

**Target condition** - Heart failure.

**Reference standard** - The reference standard was an echocardiogram performed by two echo-cardiologists blind to the BNP result. The criteria for systolic dysfunction were LVEF <50% or the inner diameter of the left ventricle at diastole was >5.2cm or global hypokinesia was detected. The criteria for diastolic dysfunction was an E-wave / A-wave ratio of <1 and the deceleration time was >240ms or the E/A ratio was >1.5 and the deceleration time <150ms.

**Index and comparator tests**- BNP (Biosite Triage) was the index test. Collected blood on arrival, stored in EDTA tubes prior to analysis. Also looked at concurrent medical conditions.

**Follow up** - Until discharge from hospital.

**Disclosures** - None mentioned.

**Notes** - Korean hospital, mentions emergency departments in pleural but doesn't state which hospitals or how many.

## **Chung 2006**

**Clinical features** - Adult patients with acute dyspnoea whom the attending emergency physician felt required hospital admission. Exclusion criteria included AMI (including NSTEMI), unstable angina, or renal failure requiring dialysis.

**Setting** - Single emergency department in Sydney, Australia. Recruitment period not stated.

**Participants** - 143 patients. Mean age 79±10 years. 63 male: 80 female patients.

**Study design** - The emergency physician assessed the patient and then had to give a likeliest diagnosis as to acute heart failure, heart failure but another precipitation condition or no heart failure. The ED physician was then told the BNP and offered the opportunity to change their mind.

**Target condition** - Heart failure.

**Reference standard** - A single cardiologist who was aware of all the medical notes, investigations and BNP result. 105 patients had an echocardiogram, LVEF was assessed on a visual estimate.

**Index and comparator tests**- BNP was the Biosite Triage fluorescence immunoassay kit.

**Follow up** - Until discharge?

**Disclosures** - Study supported by untied grant from Servier Australia. Biosite P/L lent the authors the BNP measuring device.

**Notes** -

## **Collins 2006**

**Clinical features** - Adult patients who presented with signs or symptoms of heart failure (dyspnoea, extremity oedema, fatigue) and who had an ECG requested by the attending physician.

**Settings** - Four emergency departments in the US between September 2003 and June 2004.

**Participants** - 439 patients were recruited in total. The first 17 were part of a pilot study. Of the remainder, 343 were included in the study. Main reasons for exclusion were failure to get heart sound data recorded successfully, patients were also excluded if the heart sound could not be analysed (S3: ventricular rhythm/tachycardia, heart rate >115, PR interval <120ms; S4: all of the S3 excluders plus junctional rhythm, ventricular pacing, atrial fib/flutter/tachy). Excluded if an ECG had been performed and over 1h had passed since receiving vasodilators or diuretics. Median age 61 years. 230 female : 209 male patients.

**Study design** - Prospective convenience sample.

**Target condition** - Heart failure.

**Reference standard** - 2 senior cardiology fellows blinded to BNP and heart sound data.

**Index and comparator tests**- Looked at diagnostic value of electronically detected 3rd heart sound and physician detected 3rd heart sound. Also electronically detected S4.

**Follow up** - 30 day follow up by telephone.

**Disclosures** - The original data collection for the database was paid for by Inovise Medical Inc. Drs Collins and Lindsell received honoraria from Inovise Medical Inc. Dr Peacock has received research support and is a consultant for Inovise Medical, Inc. and received research support and honoraria from Biosite,

Inc. Dr Storrow received research support from Inovise Medical, Inc. and Biosite Inc. and honoraria from Biosite, Inc.

**Notes** - Large proportion of patients not analysed because of missing or uncollectable data.

## **Collins 2009**

**Clinical features** - Patients aged over 39 years presenting to EDs with acute dyspnoea. Excluded patients with dialysis dependent renal impairment or patients where there was another obvious cause.

**Settings** - 7 centres in USA, 1 in Switzerland and 1 in Taiwan between March and October 2006.

**Participants** - 995 patients included. Excluded 70 patients as no acoustic cardiography S3 determination and 11 patients as no criterion standard diagnosis. Median age 60 years (range 40 to 95 years). 547 male : 448 female patients.

**Study design** - Convenience sample. Acoustic cardiography carried out by study personnel. ED data collected prospectively: demographics, medical history, physical examination, ECG findings as documented by the attending ED physician. Laboratory tests, CXR reported by radiologist and echo reported by a cardiologist were obtained from the medical records. After history and examination but before any diagnostic test results were available, the treating physician had to use a VAS to document the likelihood of acute decompensated heart failure. The attending physician was then presented with the acoustic cardiography printout, which incorporates an ECG and a computer generated diagnosis, and then repeat the VAS score. The score was repeated again, once all the other investigations were back.

**Target condition** - Heart failure

**Reference standard** - The opinion of 2 cardiologists at each centre who reviewed all the data apart from the acoustic cardiography. Not clear if they were blinded to each other.

**Index and comparator tests**- Index test was acoustic cardiography.

**Follow up** - Followed up to 90 days by telephone.

**Disclosure** - Inovise Medical Inc. funded the study including provision of acoustic cardiography devices and sensors, providing study support for patient enrolment and data management and assisted with device training for study personnel.

**Notes** - Acoustic cardiography software did not analyse data if the heart rate was >115.

## **Coste 2006**

**Clinical features** - Adult patients with acute dyspnoea presenting to three hospitals as emergencies

**Settings** - Three hospitals in Pontoise, Clichy and Boulogne between September 2003 and February 2005.

**Participants** - 699 consecutive patients. Mean age  $72.8 \pm 14.3$  years. 477 male : 222 female patients.

**Study design** - Standardised clinical examination. Bloods taken before any therapeutic intervention.

**Target condition** - Heart failure.

**Reference standard** - 2 cardiologists who used all results except BNP.

**Index and comparator tests**- Bloods taken into a vial containing potassium EDTA and then immediately assayed using the triage BNP test. (Biosite Diagnostics). Mean confidence limit of the analytical sensitivity  $<5\text{ng/L}$  (95% CI 0.2-4.8 ng/L).

**Follow up** - Followed up until discharge.

**Disclosures** - None stated.

**Notes** - Cannot calculate sensitivity / specificity from data provided.

## **Davis 1994**

**Clinical features** - Patients with acute dyspnoea. Patients were recruited after clinical examination to exclude obvious pneumonia, pulmonary embolus or pneumothorax.

**Settings** - One hospital in New Zealand.

**Participants** - 52 patients. 31 female : 21 male patients. Mean age 74 years (range 49 – 89 years).

**Study design** -

**Target condition** - Heart failure.

**Reference standard** - Committee of physicians and a radiologist decided end-diagnosis.

**Index and comparator tests**- Measured routine bloods, had chest x-ray and blood gases. also measured ANP, BNP and endopeptidase. Within-assay coefficients of variation were 5% for ANP and 3.2% for BNP. LVEF

radionuclide scanning done with General Electric 400T gamma camera linked to a GE Star 2 computer. The CV for LVEF varies from 5-7.5% over the range of 35-75%. All results reviewed retrospectively by a committee of physicians for a decision on the final diagnosis.

**Follow up** - Follow up was attempted after a few days to repeat the tests. This was not possible in 5 patients who had died and 6 patients who were discharged prior to follow up.

**Disclosures** - None stated.

**Notes** -

## **Fleischer 1997**

**Clinical features** - Patients requiring admission due to acute dyspnoea.

**Settings** - One hospital in New Zealand recruited between November 1994 and January 1995.

**Participants** - 123 patients. Mean age 68.3 years (range 23 – 90 years). 54 female : 69 male patients.

**Study design** - Patients had a blood sample taken within 24 hours of admission. All patients also had a chest x-ray.

**Target condition** - Heart failure.

**Reference standard** - The discharge diagnosis.

**Index and comparator tests** - BNP levels were measured on admission and were usually available within 24h of attendance. The results were made available to the treating clinician at that time with information explaining that levels greater than 49 pg/ml were suggestive of heart failure.

**Follow up** - Until discharge.

**Disclosure** - Supported in part by the Health Research Council and the National Heart Foundation of New Zealand.

**Notes** - Poor reference standard. Small numbers. Developed own assay for BNP so may not be directly comparable to commercial assays.

## **Gargani 2008**

**Clinical features** - Inclusion criteria were presence of dyspnoea at admission, venous blood sample for NT-proBNP taken on day of admission, assessment of

ultrasound lung comets within 4h of the NT-proBNP assay, no diuretic therapy between the two measurements

**Settings** - . Patients admitted to the Cardiology and Pneumology Division of the Institute of Clinical Physiology in Pisa between November 2004 and March 2006.

**Participants** - 149 consecutive patients. Mean age  $71 \pm 11$  years. 51 female : 98 male patients.

**Study design** - Although patients were described as consecutive over 100 patients were excluded, often because the emergency physician did not notify the sonographer, so it is essentially a convenience sample. No exclusion criteria given. Peripheral venous blood was collected from each patient on admission. Blood samples were collected in polypropylene tubes containing aprotinin and ethylene diamine tetra acetic acid then analysed by electro-chemiluminescence sandwich immunoassay (ECLIA) method for NT-proBNP using an Elecsys 2010 analyser from Roche Diagnostics. Plasma samples were centrifuged for 15min at 4C then stored at -20C in polypropylene tubes until assay. The coefficient for inter and intra-assay precision is 4%.

Every patient had an ultrasound chest of the anterior and lateral chest walls, with a total of 28 sites examined. The total number of comets seen was recorded giving the 'comet score'. A score of  $\leq 5$  was considered normal. Intra and inter-observer variability of assessments are 5.1% and 7.4% in their institution.

The patients also echocardiograms done with the LV volumes and EFs measured according to the modified biplane Simpson's method and adjusted for body surface area. Diastolic function was determined from the pattern of mitral and pulmonary venous flow velocity by pulsed Doppler. Diastolic dysfunction was staged as absent, mild, moderate or severe.

**Target condition** - The target condition was heart failure (prevalence 81.9%).

**Reference Standard** - The standard reference was the opinion of two cardiologists who were blinded to the ULC and NT-proBNP results. A third expert was around for cases where there was disagreement.

**Index and comparator tests**- The index test was the presence of ultrasound comets on ultrasound examination of the chest. NT-proBNP was used as a comparator test.

**Follow up** - Not clear. Presumably until discharge when all relevant investigations had been carried out.

**Disclosures** - Supported by Institute of Clinical Physiology, National Research Council and a research grant from the Clinica "Montevergine".

**Notes** - Very high level of heart failure compared with other studies, 122 out of 149 patients felt to have heart failure.

## **Gegenhuber 2006**

**Clinical features and settings** -Same patients as in Mueller 2005 paper.  
Excluded patients with STEMI and non-STEMI and trauma

**Participants** - Patients presenting to the St. John of God Hospital in Linz, Austria. All patients presenting between October 2003 and February 2004 were recruited.

**Study design** -

**Target condition and reference standard** - The target condition was heart failure (prevalence 54.6%).

**Index and comparator tests**- All patients were examined and had blood taken for BNP and NT-proBNP. Estimated GFR by Cockcroft and Gault formula. The echo and classification of the patients had to be done within 3 days of the patients' attendance. The classification was done before the NT-proBNP and BNP results were made available. Set up to compare BNP and NT-proBNP but kept samples for further analysis. These samples were then used to look at mid-range proANP.

**Follow up** -

**Notes** -

## **Green 2008**

**Clinical features and settings** -

**Participants** -

**Study design** -

**Target condition and reference standard** -

**Index and comparator tests**- Study designed around NT-proBNP

**Follow up** -

**Notes** - Further analysis of Januzzi (PRIDE) study but looking at clinical opinion though hard work to extract.

## **Huang 2006**

**Clinical features** - Adult patients with main complaint of dyspnoea. Exclusion criteria were AMI, trauma, other obvious cause of dyspnoea e.g. pneumothorax.

Also excluded patients with poor echocardiographic windows, which was 12 out of 104 potential patients.

**Settings** - Tertiary medical centre from Jan-Oct 2005.

**Participants** - Identified 104 patients and recruited 92 of these. Mean age 72.3 years. 45 male : 47 female patients.

**Study design** - Prospective.

**Target condition** - The target condition was heart failure (prevalence 55.4%).

**Reference standard** - The opinion of 2 cardiologists blinded to the patients tissue Doppler echocardiography and BNP levels. Diagnosis based on the Framingham and NHANES criteria. In the event of a disagreement between the cardiologists a 3rd cardiologist was brought into the discussion. The decision was made as to whether patients had acute heart failure or not.

**Index and comparator tests**- The index test was tissue Doppler echocardiography. Not clear when the echo was performed. Patients also had BNP checked if the attending physician requested it. BNP assayed on an AzSYM analyser from Abbott laboratories. Echocardiography was carried out by a cardiologist using a System V, GE-Ving Med Sound AB machine with a 3.5-MHz multiphase array probe. Used pulsed Doppler to record trans-mitral and aortic flow in the apical 4 chamber view. Tissue Doppler velocities were acquired at the septal and lateral mitral annular sites of the left ventricle from the apical 4-chamber view with a spectral pulse-wave pattern. The studies were then analysed by an echocardiographer blinded to all clinical details. Mitral inflow data included peak early diastolic and peak late diastolic velocities, E:A ratios and the deceleration time of early diastolic flow. The tissue Doppler peak early diastolic velocities at the septal and lateral mitral annular sites were measured and the E:Ea ratios were calculated from the average of the septal and lateal Ea due to the optimal accuracy of this approach in patients with regional wall abnormalities.

**Follow up** - Until discharge.

**Disclosure** - None declared.

**Notes** - BNP only performed if requested by the attending physician so data from this not included although the index test and reference standard were blinded to the result.

Calculated the E:Ea ratio, i.e. the peak early diastolic mitral inflow against the peak early diastolic inflow at the septal and lateral mitral annular sites but used different cut-offs for patients with ejection fractions above or below 50%.

**Januzzi 2005**

**Clinical features** - Patients over 21 years presenting with acute dyspnoea. Exclusion criteria creatinine > 2.5 mg/dl, chest trauma, ST elevation or depression >0.1mV, >2h after intravenous loop diuretics and unblinded natriuretic peptide level measurement.

**Settings** - Massachusetts General Hospital, Boston over a 4 month period.

**Participants** - 600 patients recruited, 1 withdrew consent. Mean age 62.4 years. 305 male : 295 female patients.

**Study design** - Clinical characteristics, signs, symptoms, medical history, medications, results of diagnostic test in the ED, x-ray and standard blood tests. Another blood sample was taken for NT-proBNP following measurement.

**Target condition and reference standard** - The target condition was heart failure (prevalence 35%). States that study cardiologist reviewed all results including the 60 day phone call. Doesn't state how many cardiologists were involved.

**Index and comparator tests**- Blood samples taken after recruitment and frozen at -80C.

**Follow up** - 60 day follow up on every patient. Attempts made to contact each patient and a chart review undertaken to see if any clinical events had taken place since the index presentation.

**Disclosure** - Supported by a grant from Roche Diagnostics.

**Notes** -

## **Jourdain 2002**

**Clinical features** - Adult patients presenting with a primary complaint of acute dyspnoea.

**Settings** - One emergency department in France.

**Participants** - 125 patients recruited, 115 analysed. Mean age 71.4 years.

**Study design** - All patients had BNP checked on arrival. Attending physicians blind to results. End diagnosis by cardiologists, also blind to BNP. Patients also had echo done looking for systolic or diastolic dysfunction.

**Target condition** - Target condition was heart failure. Prevalence 77/115 (67.0%).

**Reference standard** - End diagnosis provided by the cardiology service.

**Index and comparator tests**- Looked at emergency physician clinical opinion and few other clinical and investigative findings.

**Follow up** - Until discharge.

**Disclosures** - None.

**Notes** - Authors state recruited 125 patients but only 115 patients information provided in table, not clear if this is an error or if the patients have been excluded but can't see this in the text.

## **Kang 2006**

**Clinical features** - Patients with acute dyspnoea suspected to be of cardiac origin.

**Settings** - one department in Korea. Patients recruited between November 2003 and May 2004.

**Participants** - 69 patients. Mean age  $68 \pm 12$  years. 31 male : 38 female patients.

**Study design** - Prospective diagnostic study. All patients had BNP levels checked and an echocardiogram.

**Target condition** - Heart failure.

**Reference standard** - Target condition was heart failure. The reference standard was based on a selection of clinical and investigative findings.

BNP used but have given cut-off values for systolic dysfunction an

**Index and comparator tests-**

**Follow up** - Patients with diastolic dysfunction were followed up for a year.

**Disclosures** - None declared.

**Notes** - Korean. Can't use BNP data as split results up for diastolic and systolic dysfunction and have used different cut-off values.

## **Klemen 2009**

**Clinical features** - Patients presenting with acute dyspnoea. Excluded if any other obvious cause for dyspnoea other than heart failure, e.g. pneumonia

**Settings** - Pre-hospital setting which was the centre for emergency medicine in Maribor, Slovenia.

**Participants** - 546 patients recruited. 105 patients were excluded. Mean age 59.8 years. 271 male: 170 female patients.

**Study design** - Prospective study with consecutive recruitment. All patients were admitted to the same hospital following assessment.

**Target condition** - Heart failure.

**Reference standard** - Discharge diagnosis based on CXR, echo, exercise test, pulmonary function test, FBC, biochemistry and invasive tests or angiography.

**Index and comparator tests**- NT-proBNP and quantitative capnometry.

**Follow up** - Until hospital discharge.

**Disclosures** - None.

**Notes** - I think pre-hospital means pre-admission? No table 1 provided. Also used partial-pressure of end-tidal CO<sub>2</sub> as a diagnostic test but not included in my table as not blinded and not routinely used in UK practice.

## **Knudsen 2004**

**Clinical features** - Patients with a principal complaint of shortness of breath, either of acute onset or an acute exacerbation. Patients with another obvious cause were excluded. The study also excluded patients with AMI or advanced renal failure.

**Settings** - Patients were recruited from 5 US and 2 European hospitals. Recruited from June 1999 to December 2000.

**Participants** - 1586 patients with a principal complaint of shortness of breath. This subgroup analysis excluded the patients with incomplete information from the analysis so only looked at 880 patients in total. Mean age 64 ± 16 years. 482 male: 398 female patients.

**Study design** - The patients all had an ECG and a CXR (reported by a radiologist at the recruiting hospital). Bloods were taken on a point-of-care BNP test done (Biosite). Information from the data was used to calculate the Framingham and NHANES scores for heart failure. Univariate analysis was made with chi-squared or independent sample t tests.

**Target condition**- Heart failure.

**Reference standard** - Provided by 2 cardiologists who reviewed all the information at around 30 days post-admission.

**Index and comparator tests**- Looked at BNP (cut-off 100 pg/ml), radiograph findings, ECG findings, historical findings and clinical findings. Also did block test to test the hypothesis that the prognostic ability of the model with BNP as a co-variable was significantly different from that of the model without BNP.

**Follow up** - Notes assessed at around 30 days following admission.

**Disclosure** - None stated.

**Notes** - Subgroup of Maisel 2002 study.

### **Lainchbury 2003**

**Clinical features** - Adult patients with shortness of breath as part of reason for attendance to emergency department.

**Settings** - Single emergency department in New Zealand.

**Participants** - 205 patients with dyspnoea. Mean age  $70 \pm 14$  years. 100 male : 105 female patients.

**Study design** - Prospective study.

**Target condition** - Heart failure.

**Reference standard** - Two cardiologists who had access to all results apart from the BNP / NT-proBNP. In cases of disagreement a 3rd cardiologist was the final adjudicator.

**Index and comparator tests**- Patients had medical history, physical examination, blood tests, chest x-ray, and other diagnostic tests. 171 patients had an echocardiogram and 5 patients had radionuclide ventriculography.

**Follow up** - Discharge?

**Disclosures** - Supported by the Health Research Council of New Zealand and the National Heart Foundation of New Zealand. Triage BNP test strips and analyser provided by Biosite Diagnostics.

**Notes** - Ethical approval obtained.

### **Liteplo 2009**

**Clinical features** - Adult patients presenting with dyspnoea who were undergoing a NT-proBNP test as part of their investigations.

**Settings** - Massachusetts General Hospital , Boston between December 2006 and June 2007.

**Participants** - 100 hundred patients enrolled, 6 excluded due to incomplete data. Mean age  $74 \pm 14$  years. 55 male : 41 female patients.

**Study design** - Convenience sample, patients recruited when investigator available. Each patient had an eight-zone thoracic US performed. Sonographers were trained for minimum of 2.5h. Sonographers blinded to NT-proBNP result.

Interpreted at time but also recorded and later reviewed by a single sonographer blinded to clinical parameters and

**Target condition** - Heart failure (prevalence 42.6%).

**Reference standard** - There was a 'criterion standard' diagnosis of CHF based on a consensus of chart review. 2 emergency physicians reviewed all data except US results, so included NT-proBNP results.

**Index and comparator tests-**

**Follow up** - Until discharge.

**Disclosures** - None stated.

**Notes** - Selection bias as doctors would know that if they checked a NT-proBNP test then the patient would be eligible for the study and vice versa.

## **Lo 2007**

**Clinical features** - Adult patients with one or more of the following criteria: SOB, respiratory rate >20bpm, oxygen saturations less than 90% in room air. Exclusion criteria included trauma and pregnancy.

**Setting** - Single emergency department at Linko Chang Gung Memorial Hospital, Taiwan.

**Participants** - 60 patients recruited, 8 excluded due to lack of definite diagnosis. Mean age 68.5 ±14.2 years. 14 female : 38 male patients.

**Study design** - Convenience sample. Uses impedance cardiography to measure beat-to-beat changes of thoracic bio-impedance to calculate haemodynamic parameters. Prospective blind study in convenience sample. All patients seen and examined by an emergency physician and had blood tests, ECG, CXR and ABG. Echocardiograms performed as felt necessary.

**Target condition** - Heart failure (prevalence 38.5%).

**Reference standard** - A review of medical record by an ED physician blind to ICG results and not involved in the treatment of the patient.

**Index and comparator tests-** Impedance cardiography data evaluated retrospectively. A cardiac index <2.4 or lower or a systolic time ratio of 0.55 or higher concurrent with a cardiac index of less than 3.0 used to define cardiac cause of dyspnoea.

**Follow up** - Until discharge.

**Disclosures** - None declared.

**Notes** - 6 patients excluded from authors calculations as no clear end-diagnosis. I think that the patients should have been included as non-cardiac breathlessness for the calculations. No ethical approval mentioned.

## **Logeart 2002**

**Clinical features** - All patients presenting to one emergency department with severe shortness of breath were recruited over a 2 year period.

**Setting** - One emergency department in Clichy, France between June 1999 and June 2001.

**Participants** - 235 patients were enrolled, 72 patients were excluded due to treatment already being started or because emergency echocardiography was not feasible. Mean age 67.4 years. 109 male : 56 female patients.

**Study design** - All patients had an ECG and chest x-ray and blood taken for BNP levels (Biosite). All included patients also had an echocardiogram done using a Hewlett-Packard Sonos 1500 with a 2.5 MHz probe. The echo was recorded on videotape and reviewed by a cardiologist.

**Target condition** - Heart failure.

**Reference standard** - The opinion of two cardiologists and one pneumologist who were blinded to the BNP and echocardiogram results.

**Index and comparator tests**- Data also compiled on historical and clinical findings. Ejection fraction visually estimated from echocardiogram.

**Follow up** - Not mentioned.

**Disclosure** - Supported in part by grant from the French Federation of Cardiology. BNP assays provided by Biosite.

Notes –

## **Lokuge 2010**

**Clinical features** - All patients presenting to one emergency department with severe shortness of breath were recruited between August 2005 and April 2007.

**Setting** – Emergency departments in two hospitals in Victoria, Australia.

**Participants** - 799 patients recruited, various patients excluded (reasons provided) so 612 patients left. 328 male: 284 female.

**Study design** – Half of the patients had a BNP test done. The majority of patients had a clinical estimation of the likelihood of heart failure by the attending physician.

**Target condition** - Heart failure.

**Reference standard** – A cardiologist and an emergency physician, blinded to the BNP results provided the reference standard based on the patient notes and investigation results. In the event of a disagreement a third physician made the final diagnosis.

**Index and comparator tests**- BNP was analysed using the Abbott Laboratories assay.

**Follow up** – Until discharge.

**Disclosure** – Supported by an unrestricted educational grant from Janssen-Cilag (part of Johnson & Johnson).

Notes -

Maisel 2002

**Clinical features** - Adult patients with acute dyspnoea of either acute onset or an acute exacerbation. Excluded patients with another obvious cause for dyspnoea, patients with AMI or advanced renal failure.

**Settings** - International, Patients recruited from 5 US and 2 European hospitals (France & Norway). April 1999 to December 2000.

**Participants** - 1586 patients. Mean age 64 ±17 years. 888 male: 698 female patients.

**Study design** - The patients all had an ECG and a CXR (reported by a radiologist at the recruiting hospital). Bloods were taken and a point-of-care BNP test done (Biosite). Information from the data was used to calculate the Framingham and NHANES scores for heart failure. Univariate analysis was made with chi-squared or independent sample t tests.

**Target condition** - Heart failure.

**Reference standard** - Two cardiologists who reviewed all the information at around 30 days post-admission.

**Index and comparator tests** - Comparing BNP against reference standard of cardiologists opinion.

**Follow up** - Notes assessed around 30 days following admission.

**Disclosures** - Biosite supplied the diagnostic kits and managed the technical aspects of the data accrual and storage. Biosite provided some financial support. Drs Maisel and McCullough have received honorariums from Biosite for speaking and consulting.

**Notes** - Study that McCullough and Knudsen's data derived from.

Maisel 2010

**Clinical features** - Adult patients with acute dyspnoea of either acute onset or an acute exacerbation. Excluded patients with another obvious cause for dyspnoea, patients with AMI or advanced renal failure.

**Settings** - International, Patients recruited from 8 US and 6 European hospitals and 1 hospital in New Zealand. March 2007 to February 2008.

**Participants** - 1641 patients. Mean age 64 ±17 years. 859 male: 782 female patients.

**Study design** - The patients all had an ECG and a CXR (reported by a radiologist at the recruiting hospital). Bloods were taken and a point-of-care BNP test done (Biosite). Information from the data was used to calculate the Framingham and NHANES scores for heart failure. Univariate analysis was made with chi-squared or independent sample t tests.

**Target condition** - Heart failure, prevalence 34.6%

**Reference standard** - Two cardiologists who reviewed all the information at around 30 days post-admission. Third cardiologist available in event of disagreement.

**Index and comparator tests** - Comparing MR-ANP against reference standard of cardiologists opinion and BNP.

**Follow up** – Followed up for 90 days.

**Disclosures** - Dr. Maisel has received research support from Roche, Biosite, and Bayer, and is a consultant to Biosite. Dr. Mueller has received research grants from the Swiss National Science Foundation, the Swiss Heart Foundation, the Novartis Foundation, the Krokus Foundation, Abbott, Biosite, BRAHMS, Roche, and the University of Basel. Dr. Peacock is on the Scientific Advisory Board of Abbott, Beckman-Coulter, Biosite, Inverness, Ortho Clinical Diagnostics, and Response Biomedical, and has received research grants from Abbott, Biosite, and Inverness. Dr. Ponikowski has received honoraria for serving as a consultant for Vifor and Athera, and as a speaker for Merck-Serono, Pfizer, and Sanofi-Aventis. Dr. Richards is on the Scientific Advisory Board of Inverness Medical, and has received travel support, honoraria, and research grants from Roche Diagnostics and Inverness Medical (Biosite). Dr. Filippatos has received research support from Biosite, BRAHMS, and Roche. Dr. Di Somma is a

consultant to Biosite. Dr. Ng is a consultant for Inverness Medical Innovations and BRAHMS AG. Dr. Daniels has received research support from Biosite Inc. and Roche Diagnostics. Dr. Neath is a consultant to BRAHMS USA. Dr. Christensen is a consultant for Siemens, BG-Medicine, Critical Care Diagnostics, Inverness Medical, and Abbott Diagnostics, and has received research support from Siemens, BG-Medicine, BRAHMS, Roche, Inverness Medical, and Nanosphere. Dr. McCord has received research funding from BRAHMS Diagnostics. Oliver Hartmann is an employee of BRAHMS AG, which is a company developing and marketing in vitro diagnostic products, including the MR-proANP and MR-proADM assays used in this study. Dr. von Haehling has received honoraria from BRAHMS. Dr. Bergmann is an employee of BRAHMS AG, holds patent applications related to this technology, and is a shareholder of BRAHMS AG. Dr. Morgenthaler is an employee of BRAHMS AG. Dr. Anker has received research support from BRAHMS, honoraria from Abbott and Biosite, and is a consultant to BRAHMS.

**Notes** – Some of the variables that were looked at were provided as percentages of the patient totals and so in converting them back to raw patient numbers there were some slight differences between the calculated and reported patient totals. This usually consisted of one or two patients but occasionally there were more suggesting some errors in the data provided in the table. This is not likely to significantly affect any of the variables.

Marantz 1990

**Clinical features** - Chief complaint of shortness of breath.

**Settings** - Patients presenting to one hospital with shortness of breath.

**Participants** - Patients had to be 40y or over, English speaking, have attended during study hours and consent. 103 patients were considered, 52 were excluded or refused. Mean age 64 years. 20 male : 31 female patients.

**Study design** -

**Target condition** - The target condition was heart failure (prevalence 45%).

**Reference standard** - The reference standard was the investigators blindly applying predetermined clinico-radiographic criteria, the Boston criteria. Although the Boston criteria has apparently been validated they authors modified it by removing wheezing as one of the criterion as they felt that there were patients with asthma or COPD who would be included wrongly, presumably the modified score has not been validated. No kappa value given for the application of these criteria.

**Index and comparator tests**- Looked at the presence of hepato-jugular reflux and the response of blood pressure to the Valsalva manoeuvre. The hepato-jugular reflux was elicited by pressing over the abdomen with a partially inflated

blood pressure cuff. Pressure was maintained over 10 seconds and an increase of at least 3cm water was considered positive. The Valsalva manoeuvre was performed with the blood pressure cuff inflated to 15 mmHg above the systolic blood pressure. An interpretable result required at least two beats to be heard. A normal response was to hear beats then for the beats to disappear but reappear once the manoeuvre was stopped. An abnormal result was when there was no reappearance of sounds after the release of the manoeuvre or when the sounds did not disappear.

**Follow up** - None mentioned.

**Disclosures** - None declared.

**Notes** - Consider removing this study as poor quality.

McCullough 2002

Clinical features and settings -

Participants -

Study design -

Target condition and reference standard -

Index and comparator tests-

Follow up -

**Notes** - Same study as Maisel 2002 but includes clinical certainty of heart failure.

Moe 2007

**Clinical features** - Adult patients with dyspnoea suspected to be of cardiac origin. Excluded patients with severe renal impairment, cancer, dyspnoea from clinically over causes such as trauma.

**Settings** - 7 participating emergency departments in Canada between December 2004 and December 2005.

**Participants** - 534 patients screened, 500 enrolled. Mean age 70.5 years. 258 male :242 female patients.

**Study design** - RCT, double-blind, prospective multi-centre trial. Patients were enrolled and had a history taken, examination and investigations. A separate sample was taken for NT-proBNP measurement. Patients were then randomised into two groups, those who had a known NT-proBNP and those whose NT-proBNP remained unknown.

**Target condition** - Heart failure (prevalence 46%).

**Reference standard** - Provided by 2 cardiologist who reviewed all the data from the admission up to the 60 day follow-up by telephone interview apart from the NT-proBNP result.

**Index and comparator tests** - NT-proBNP analysis performed using Elecsys 1010, 2010 or E170 proBNP assay from Roche diagnostics.

**Follow up** - 60 day patient telephone interview and chart records.

**Disclosures** - Dr Moe received a research grant and honoraria from Roche Diagnostics Canada. Dr Januzzi received research grants, honoraria, and consulting fees from Roche Diagnostics, Dade-Behring and Ortho Clinical Diagnostics. H. Zowall received a research grand and consulting fees from Roche Diagnostics Canada.

**Notes** - IMPROVE-CHF study.

Morrison 2002

**Clinical features** - Patients with presenting with a prominent complaint of dyspnoea at rest, on exertion or on lying down. Excluded patients with another obvious cause of the dyspnoea.

**Setting** - San Diego Veteran's Health Care System.

**Participants** - 321 patients were recruited. 305 male : 16 female patients. No details about age provided.

**Study design** - Prospective. Convenience sample.

**Target condition and reference standard** - The target condition was heart failure (prevalence 42%) and the reference standard was the opinion of 2 cardiologists based around the Framingham criteria. Any other investigations the patient had other than the BNP result were also available to the cardiologists.

**Index and comparator tests**- BNP was analysed using the Triage Biosite assay. The coefficient of variation for intra-assay precision is 9.5% fro 28.8 ng/l and 13.9% for 1,180 ng/l. The CV for inter-assay variation is 10% for 28.8 ng/l, 12.4% for 584 ng/l and 14.8% for 1,180 ng/l. The measurable range for BNP is 5-13,000 ng/l.

**Follow up** - No follow time period stated.

**Disclosure** - None declared.

**Notes** - According to email from Maisel to Schwam, these patients were not included in the Breathing Not Properly study but the Dao study are a subset of these patients. (Dao et al.; Schwam)

Mueller 2005

**Clinical features** - Patients presenting with dyspnoea. Excluded patients with STEMI and non-STEMI and trauma.

**Settings** - One emergency department in Austria. All patients presenting between October 2003 and February 2004 were recruited.

**Participants** - 293 patients presented with dyspnoea. 33 excluded. 17 did not meet the criteria, 11 had no reference standard, 5 did not have bloods done within four hours. Male 243 : female 17 patients.

**Study design** - All patients were examined and had blood taken for BNP and NT-proBNP. Estimated GFR by Cockcroft and Gault formula. The echo and classification of the patients had to be done within 3 days of the patients attendance. The classification was done before the NT-proBNP and BNP results were made available.

**Target condition** - The target condition was heart failure (prevalence 52.5%).

**Reference standard** -Reference standard was one of the investigators reviewing all the data about the patient except the BNP and NT-proBNP. The diagnosis was based on the Framingham score and/or echocardiography.

**Index and comparator tests**- History taken, clinically examined, 12 lead ECG, chest x-ray. LVEF was calculated at echo by biplane Simpson's method. If the LVEF was less than or equal to 50% then the patient had systolic dysfunction. If the LVEF was greater than 50% the patient was examined for diastolic dysfunction by looking at pulsed wave Doppler examination of mitral inflow before and during a Valsalva Manoeuvre. This could not be done in patients with AF. Diastolic dysfunction was graded from 1 to 4. BNP was assayed on a AxSYM analyser from Abbot Laboratories. The AxSYM BNP assay had a coefficient of variance of 8.1% at a mean concentration of 108 ng/l, a total CV of 7.5% at a concentration of 524 ng/l and a total CV of 10% at a mean concentration of 2117 ng/l.

NT-proBNP was measured on an Elecsys 2010 instrument from Roche Diagnostics. Fully automated 2 site enzyme immunoassay with electrochemiluminescent technology for use with human serum and plasma. The total CV was 3.8% at a mean concentration of 246 ng/l, 4.7% at a mean concentration of 891 ng/l and a 2.2% at a mean concentration of 10,666 ng/l.

**Follow up** - Classified and had all investigations by 3 days.

**Disclosures** - Supported in part by a grant from the Upper Austrian Government. Reagents provided by Abbott Diagnostics and Roche Diagnostics.

**Notes** - My results different from the authors as they have excluded patients with an unclear diagnosis from the analysis while I have treated it as intention to diagnose and included these patients as not heart failure.

Mueller, C 2005

**Clinical features** - Patients presenting with acute dyspnoea. Excluded patients with obvious traumatic cause, severe renal disease or cardiogenic shock.

**Setting** - Single emergency department at the University Hospital in Basel, Austria between

**Participants** - 452 patients. Mean age 71±15 years. 262 male : 190 female patients.

**Study design** - All patients were assessed and had history, examination, ecg, pulse oximetry, bloods and cxr. Signs and symptoms recorded by the ED resident. The original study randomised patients to have a BNP test or not to see if this improved outcome, so this was available for 225 of the patients.

**Target condition** - The target condition was heart failure (217/ 452 48% prevalence).

**Reference standard** - Reference standard was the opinion of an internal medical specialist based on all the available results, findings and outcomes for the patient.

**Index and comparator tests**- Variety of symptoms and signs assessed. Significant predictors of heart failure listed.

**Follow up** - To discharge.

**Disclosures** - None declared.

**Notes** - Original study published NEJM 2004.

Nazerian 2010

**Clinical features** - Patients with acute dyspnoea presenting to an ED of an adult tertiary centre in Italy. Patients with acute onset or worsening of chronic dyspnoea. Excluded patients with trauma, STEMI, obvious other cause. Also excluded if had intravenous treatment in the ED before echo and NT-proBNP checked. Excluded patients with poor acoustic window. Recruited Jan 2007 to March 2007.

**Setting** – Single ED in adult tertiary centre in Italy between Jan and March 2007.

**Participants** – 145 patients recruited (41 excluded). 74 female: 71 male patients. Mean age of 77.6 years.

**Study design** - Convenience sample of patients. Investigator present for average of 14h per day. Four investigators recruiting patients. Patients had

history taken, physical examination, ECG, ABG and routine blood tests. Treating physicians completed Boston criteria form to estimate risk of heart failure. Bloods and echo done within 30 minutes of inclusion. Treating physician blinded to NT-proBNP and echo results. NT-proBNP performed with Roche immunoassay on Elecsys 1010 analyser. Co-efficient for inter- and intra-assay precision 4% according to manufacturers data. Echo performed by emergency physicians and included 2D examination and pulsed Doppler analysis of mitral flow. Estimated LVEF using cross-sectional diameter measurements. Used apical view to take Doppler samples between mitral leaflet tips to look at E/A ratio to describe impaired relaxation, normal or restrictive pattern.

**Target condition** - Heart failure (prevalence 44%).

**Reference standard** - Reference standard provided by three independent reviewers, 2 cardiologists and 1 respiratory physician. Reviewers had access to ED notes, clinical notes and any other additional information from hospital stay. Blinded to NT-proBNP and echo results.

**Index and comparator tests**- Variety of symptoms and signs assessed Boston criteria as filled in by attending physician. NT-proBNP. ED echo.

**Follow up** - To discharge.

**Disclosures** - None declared.

**Notes** – Small study.

Parrinello 2008

**Clinical features** - Patients presenting with dyspnoea to the ED. Excluded patients with recent history of ACS/MI or cardiac surgery. Also excluded patients with pericardial effusions, PEs, trauma, liver cirrhosis, advanced renal failure and nephrotic syndrome.

**Setting** – the ED of the University Hospital Palermo between December 2004 and December 2006.

**Participants** – 292 adult patients, selected by physicians for the study, i.e. not consecutive. Mean age 67.5 years. 156 male:136 female patients.

**Study design** - Patients underwent clinical history, physical exam, ECG, pulse oxymetry, blood tests included BNP and CXR. Bioelectrical impedance analysis was also recorded on entry to the study.

**Target condition** - Target condition heart failure. Prevalence 58.9%.

**Reference standard** - Reference standard was by an independent team including a cardiologist and an internist who reviewed all the data including the BNP and echocardiography results. The patients who were diagnosed with acute heart failure were further divided into preserved or impaired systolic function.

**Index and comparator tests-** After the patients had been diagnosed by the independent group the researchers looked at the value of BIA in assisting with the diagnosis.

**Follow up** – 90 days.

**Disclosures** - None

**Notes** –

Potocki 2010

**Clinical features** - Patients presenting with dyspnoea to. Excluded patients on dialysis or with a history of trauma.

**Setting** – the ED of the University Hospital Basel from April 2006 to March 2007

**Participants** – 287 consecutive adult patients. Median age 77 years.

**Study design** - Patients underwent clinical history, physical exam, ECG, pulse oximetry, blood tests and CXR. ED physicians assessed the probability that the patient had heart failure based on all available information including BNP.

**Target condition** - Target condition heart failure. Prevalence 54%.

**Reference standard** - Reference standard two independent cardiologists who had access to all the information including BNP, but not NT-proBNP or MRANP. BNP considered as quantitative marker of heart failure and therefore interpreted as continuous variable. Absolute BNP variables also adjusted for the presence of obesity and renal function.

**Index and comparator tests-** Blood samples taken for determination of MRproANP and NTproBNP levels. MRproANP detected using sandwich immunoassay (MRproANP LIA, B.R.A.H.M.S, Hennigsdorf/Berlin, Germany). Functional assay sensitivity (interassay coefficient of variance <20%) is 20pmol/l. NT-proBNP levels were determined by a quantitative electrochemiluminescence immunoassay (Elecsys proBNP, Roche Diagnostics, Switzerland). BNP was measured by a microparticle enzyme immunoassay (Ax-Sym; Abbott Laboratories, Abbott Park, USA).

**Follow up** – 90 days.

**Disclosures** - Supported by research grants from Swiss National Science Foundation, Brahms, Brandenburg Ministry of Economics, European Regional

Development Fund. Dr Mueller has received research support from Abbott, Biosite, Brahms, Roche and Siemens. Dr Morgenthaler is an employer at Brahms. Dr Bergmann is a member of the board of directors and shareholder of B.R.A.H.M.S. Drs Bergmann and Morgenthaler are inventors of patents related to MRporANP.

Notes –

## Ray 2006

**Clinical features** - who were at least 65y; had dyspnoea for less than 2 weeks; and at least one of the following: a respiratory rate greater than 25 bpm; an Pao<sub>2</sub> of 70mmHg or less; a peripheral oxygen saturation of 92% or less; a PaCO<sub>2</sub> of 45mmHg or more with a pH of 7.35 or less. There were no exclusion criteria.

**Settings** - Patients presenting to the Emergency Department of a single hospital in Paris between Feb 2001 and Sept 2002.

**Participants** - 514 patients. Mean age 80 ±9 years. 253 male : 261 female patients.

**Study design** - The ED physician took history and examined the patients. ABG, CXR and bloods were taken. Dependence and quality of life were assessed by the activity of daily living score. Treatment and admission decisions were carried out as usual. Some patients had high-resolution CT scans, echocardiograms or pulmonary function tests. Some patients had BNP testing done and some had NT-proBNP testing plus BNP testing; the attending physician was not informed of the results.

**Target condition** - The target condition was heart failure (prevalence 43%).

**Reference standard** - The reference standard was the opinion of two senior physicians from an examination of the complete medical chart, laboratory tests, CXR data and the results of the HRCT, echo, PFTs and BNP/NT-proBNP when available. In cases of disagreement, a third physician was involved.

**Index and comparator tests**- In the 2005 paper, looking at 202 patients from this same group, the reference standard is given as the same except that the experts don't know the BNP or NT-proBNP results.

In the 2004 paper, looking at 380 patients from this same group, the reference standard is given as the same except that the experts don't know the BNP result.

The treating ED physician was asked to make a diagnosis while the patient was still in the ED. BNP was performed using the Triage BNP from Biosite. The biochemist was unaware of the diagnosis and the emergency physician and expert panel were unaware of the BNP levels.

**Follow up** - until discharge?

**Disclosures** - Biosite and Roche Diagnostics provided the diagnostic tests free of charge.

**Notes** - Ethics committee approval obtained, the need for informed consent waived. This study contained 514 patients. A subgroup of this study consisting of 308 patients was published by Ray in 2004 in Int Care Med who had BNP testing carried out. Another subgroup was published by Ray in 2005 in J Am Geriat Soc. This consisted of 202 patients who had NT-proBNP and BNP testing performed. The 202 patients were largely a subset of the 308 patients who had BNP testing done as it is mentioned in this paper, the 2006 paper, that 375 patients had BNP or NT-proBNP testing done. It is not clear from any of the published papers how patients were selected for the subgroup studies as the inclusion criteria and time periods are described identically in all 3 papers.

**Study design** – Patients taken from the Mueller, C 2005 and Potocki 2010 studies. Checked uric acid levels on all patients and looked for diagnostic and prognostic utility. 743 patients included from these two studies.

**Target condition** - The target condition was heart failure (prevalence 51%).

Robaei 2011

**Clinical features** – Adult patients presenting with dyspnoea.

**Settings** - Patients presenting to the Emergency Department of a single tertiary hospital in Australia. Time period not specified.

**Participants** - 68 patients. Mean age 73±16 years. 30 male : 38 female patients.

**Study design** - The ED physician took history and examined the patients then asked to decide if the patient had heart failure. In 38 of the patients a blood sample was taken for NT-proBNP and the results were then made available to the physician to see if this altered the presumed diagnosis.

**Target condition** - The target condition was heart failure (prevalence 40%).

**Reference standard** - The reference standard was the opinion of a single cardiologist blinded to the NT-proBNP results.

**Index and comparator tests**-It isn't really possible to use the NT-proBNP test results as the results were used with a range of age-dependent cut-off values and an overall sensitivity and specificity were provided.

**Follow up** - until discharge.

**Disclosures** – Partially supported by a grant from Roche Diagnostics Australia.

**Notes** – Small study. Difficult to elicit results due to the way that they were presented.

## **Shah 2009**

**Clinical features** - Patients 18y or older presenting to an ED setting with a primary complaint of dyspnoea. Exclusion criteria were overt causes of dyspnoea e.g. trauma, pneumothorax or upper airway obstruction and ACS. ACS defined as ST segment deviation more than 1mm or a serum troponin concentration rise above the upper reference limit for the local laboratory.

**Settings** - Patients recruited from 5 centres in US between May 2003 and December 2006.

**Participants** - 412 patients. Mean age  $58 \pm 14$  years. 251 male : 161 female patients.

**Study design** - Prospective cohort study. Data obtained by designated study coordinator from the medical records and the patients directly. Data collected included demographics, symptoms, physical examination results, observations, ECG, medical history, CXR and medication status. Data used to adjust multivariate models of prognosis and to calculate estimated GFR. 20ml of blood taken into plastic tubes, EDTA containing, heparin coated and plain tubes. Blood centrifuged within 2h and stored at  $-20^{\circ}\text{C}$  or colder. Then transferred on dry ice to main laboratory where stored at  $-70^{\circ}\text{C}$ . Samples thawed once in room-temperature water bath and assayed within 1h.

**Target condition** - Target condition was heart failure (prevalence 35.7%).

**Reference standard** - Reference standard was provided by a panel of experts. 2 clinicians reviewed all results by standardised adjudication form according to the Antihypertensive and Lipid Lowering Treatment to Prevent Heart Attack Trial criteria. A third clinician was used if there was disagreement.

**Index and comparator tests** - Measured myeloperoxidase concentrations from the EDTA samples, measured by immunoassay on the Dimension RxL system (Siemens Healthcare Diagnostics). MPO within run and total imprecision values of 3.8% and 4.8% respectively at 428 pmol/l and 3.3.% and 3.6% respectively at 3642 pmol/l. The detection limit was 13 pmol/l and the assay was linear to 5223 pmol/l.

The BNP was measured with Triage BNP 2-site immunoassay reagents (Biosite) formatted for Beckman Coulter instrumentation. Performed on an Access II system.

The NT-proBNP was measured with 1-step immunometric Dimension PBNP sandwich assay (Dimension RxL system; Siemens Healthcare Diagnostics).

All blood sample analysis performed blind to patient data.

Prospective echocardiography planned for all patients with 96h. Performed by expert sonographer and images sent to core laboratory where a single investigator, blinded to the clinical data, interpreted all the images and performed all the measurements. Simpson biplane analysis was used to determine LV end-systolic volume and the LV end-diastolic volume. LVEF was calculated as  $LVEDV - LVESV / LVEDV \times 100\%$ . LV mass was measured by M-mode criteria using formula derived by Devereux. The LV mass was indexed to body surface area.

Diastolic function assessed in the apical 4 chamber view. Doppler evaluations of mitral inflow velocities and tissue Doppler imaging sampling from the lateral mitral annulus were obtained to grade the severity as stage 1 - impaired relaxation, ratio of transmitral peak flow velocity in early diastole to peak flow velocity in late diastole (E/A ratio  $<1$ , deceleration time  $>220$ ms), stage II - pseudonormal, E/A ratio between 1 and 2, Dt  $<220$ ms, early diastolic mitral annular velocity  $<8.5$  ms or stage III - restrictive E/A ratio  $>2$ , Dt  $<150$  ms.

Follow up -

**Disclosures** - Dr Christenson is a consultant for Siemens Healthcare Diagnostics and Response Biomedical and has received honoraria from these companies and Inverness Medical Innovations and Instrumentation Laboratories. He has also received research funding from Siemens Healthcare, Roche Diagnostics, Response Biomedical, Brahms, Inverness Medical Innovations and Instrumentation Laboratories.

Dr Mehra is a consultant or has an advisory role for Roche, XDX, Debiopharm, Sovay, Orqis Medical, Medtronic, St. Jude Medical and Geron. He has received honoraria from Inverness Medical Innovations (formerly Biosite) and Roche diagnostics and has received research funding from the National Institutes of Health, Other Tobacco-Related Diseases Fund and Maryland Industrial Partnerships.

Dr deFillippi is a consultant or has an advisory role for Siemens Healthcare Diagnostics and Roche Diagnostics. He has received honoraria and research funding from Siemens Healthcare Diagnostics and Roche Diagnostics.

**Notes** - Data for BNP, NT-proBNP, MPO and echo given as Pearson coefficients etc. Not given raw figures to work out sensitivity and specificity for any given cut-offs.

## **Singer 2009**

**Clinical features** - Patients 18y or over with a primary complaint of SOB. Excluded patients with asthma exacerbations, trauma and chronic renal failure requiring haemodialysis.

**Settings** - 10 community and university hospitals throughout US in 2005. (New York , Boston, North Carolina, New Jersey, Massachusetts, Philadelphia.)

**Participants** - 301 patients. Mean age 61 ±18 years. 132 male : 169 female patients.

**Study design** - Prospective, multi-centre, observational study looking at the incremental benefit of a SOB point-of-care biomarker panel. Convenience sample of patients, recruited when a member of the investigator team available. Demographic and clinical data collected by a trained research assistant or nurse on all patients using standardised data collection sheets. After initial evaluation, consisting of history and examination, radiography and ECG, the attending physician estimated probability of HF, MI and PE. These estimates were dichotomised into high (≥80%) and low or medium (<80%).

**Target condition** - The target conditions were myocardial infarction, heart failure (prevalence 30.2%) and pulmonary embolism.

**Reference standard** - The reference standard was provided by an investigator blinded to the biomarker panels results but able to view central laboratory BNP results.

**Index and comparator tests**- Blood samples collected and analysed using the Triage Profiler SOB point-of-care biomarker panel. Fluorescence-based immunoassay that measures TnI, myoglobin, CK-MB, D-dimer and BNP simultaneously. The following values were considered positive myoglobin >107 ng/ml, CK-MB >4.2ng/ml, TnI >0.05 ng/ml, D-dimer >200 ng/ml and BNP >100 ng/ml. The results were not made available to the attending physician.

**Follow up** - At 30 days the inpatient records were reviewed and telephone interviews were performed to determine the presence of MI, HF and PE based on standardised criteria.

**Disclosures** - Funded by Biosite Inc., Sand Diego, CA.

Notes -

Villacorta 2002

**Clinical features** - Patients presenting with dyspnoea.

**Settings** - A single ED with acute dyspnoea from April to July 2001. The centre is a described as a referral centre for cardiology in that city.

**Participants** - 70 consecutive patients. Mean age 72.4 ± 15.9 years. 33 male : 37 female patients.

**Study design** - Excluded patients with a diagnosis of a condition causing the dyspnoea such as tracheal stenosis or cardiac tamponade. Also excluded patients with acute coronary syndromes whose prominent complaint was not dyspnoea. All patients underwent laboratory tests and chest x-rays. Echocardiography was recommended. After initial assessment the emergency medicine physician assigned an initial diagnosis, blinded to BNP values.

**Target condition** - The target condition was heart failure (prevalence 51.4%).

**Reference standard** - The reference standard was provided by a cardiologist who reviewed all patients' data and was blinded to the BNP result. Patients who had an echocardiogram showing left ventricular fractional shortening equal to or less than 25% were considered as having systolic dysfunction.

**Index and comparator tests**- The BNP was measured using the Triage B-Type Natriuretic Peptide test by Biosite Diagnostics Inc., San Diego. 5ml of blood collected in tubes containing EDTA. The majority of samples were processed immediately, a few (not clear how many) were spun, frozen and then analysed 1-3 days later. BNP is known to be stable so this should not affect the results.

**Follow up** - Results until discharge were reviewed.

**Disclosures** - None declared.

**Notes** - Small study.

Wang 2010

**Clinical features** - Patients over 17 years presenting with acute dyspnoea. Exclusion criteria included acute MI, trauma or another obvious cause of dyspnoea.

**Settings** – Tertiary medical centre in Taiwan over a 15 month period.

**Participants** – 95 patients screened, 84 enrolled.

**Study design** – Various signs and symptoms recorded while in the emergency department. Echocardiogram performed by emergency physician. BNP taken while in ED.

**Target condition and reference standard** - The target condition was heart failure (prevalence 58.3%). Two cardiologists blinded to echo and BNP results reviewed all other data and made the diagnosis.

**Index and comparator tests**- BNP analysed using Abbott laboratories system.

**Follow up** – Until discharge.

**Disclosure** – None declared.

**Notes** – Small study. Does not mention any failure to record echocardiogram findings.

## 10.8 Appendix VII - References

Ababsa, R., Jourdain, P., Funck, F., Deschamps, P. and Sadeg, N. (2005) '[BNP and dyspnea: proposition of a diagnostic strategy based on two cut-off].' *Annales de Biologie Clinique*, 63(2), Apr 01, pp. 213-216.

Abdelhafiz, A. H. (2002) 'Heart failure in older people: causes, diagnosis and treatment.' *Age and Ageing*, 31(1), Feb, pp. 29-36.

Abraham, W. T., Adams, K. F., Fonarow, G. C., Costanzo, M. R., Berkowitz, R. L., LeJemtel, T. H., Cheng, M. L., Wynne, J., Committee, A. S. A., Investigators and Group, A. S. (2005) 'In-hospital mortality in patients with acute decompensated heart failure requiring intravenous vasoactive medications: an analysis from the Acute Decompensated Heart Failure National Registry (ADHERE).' *J Am Coll Cardiol*, 46(1), Jul 5, pp. 57-64.

Acanfora, D., Trojano, L., Maggi, S., Furgi, G., Rengo, C., Iannuzzi, G., Papa, A. and Rengo, F. (1998) 'Development and validation of a Clinical History Form for the diagnosis of congestive heart failure.' *Aging - Clinical and Experimental Research*, 10(1), Mar 01, pp. 39-47.

Adams, K. F., Jr., Fonarow, G. C., Emerman, C. L., LeJemtel, T. H., Costanzo, M. R., Abraham, W. T., Berkowitz, R. L., Galvao, M., Horton, D. P., Committee, A. S. A. and Investigators (2005) 'Characteristics and outcomes of patients hospitalized for heart failure in the United States: rationale, design, and preliminary observations from the first 100,000 cases in the Acute Decompensated Heart Failure National Registry (ADHERE).' *Am Heart J*, 149(2), Feb, pp. 209-216.

Evaluation and Management of Chronic Heart Failure in the Adult 1-22 (2002).

Ahluwalia, A., Macallister, R. J. and Hobbs, A. J. (2006) 'Vascular actions of natriuretic peptides. Cyclic GMP-dependent and -independent mechanisms.' *Basic Research in Cardiology*, 99(2), Jun 25, pp. 83-89.

Ahmed, A., Aboshady, I., Munir, S. M., Gondi, S., Brewer, A., Gertz, S. D., Lai, D., Shaik, N. A., Shankar, K. J., Deswal, A. and Casscells, S. W. (2008) 'Decreasing body temperature predicts early rehospitalization in

congestive heart failure.' *Journal of Cardiac Failure*, 14(6), Aug 01, pp. 489-496.

Alibay, Y., Beauchet, A., el Mahmoud, R., Schmitt, C., Brun-Ney, D., Benoit, M., Dubourg, O., Boileau, C., Jondeau, G. and Puy, H. (2005) 'Plasma N-terminal pro-brain natriuretic peptide and brain natriuretic peptide in assessment of acute dyspnea.' *Biomedicine & Pharmacotherapy*, 59(1-2) pp. 20-24.

Allen, L., Hernandez, A., O'Connor, C. and Felker, G. (2009) 'End Points for Clinical Trials in Acute Heart Failure Syndromes.' *Journal of the American College of Cardiology*, 53(24) pp. 2248-2258.

Anderson, W. (2005) 'Statistical techniques for validating logistic regression models.' *Ann Thorac Surg*, (80) p. 1169.

Anwaruddin, S., Lloyd-Jones, D., Baggish, A., Chen, A., Krauser, D., Tung, R., Chae, C. and Januzzi, J. (2006) 'Renal function, congestive heart failure, and amino-terminal pro-brain natriuretic peptide measurement: results from the ProBNP Investigation of Dyspnea in the Emergency Department (PRIDE) Study.[see comment].' *Journal of the American College of Cardiology*, 47(1), Feb 03, pp. 91-97.

Apple, F. S., Murakami, M. M., Pearce, L. A. and Herzog, C. A. (2004) 'Multi-biomarker risk stratification of N-terminal pro-B-type natriuretic peptide, high-sensitivity C-reactive protein, and cardiac troponin T and I in end-stage renal disease for all-cause death.' *Clinical Chemistry*, 50(12), Dec, pp. 2279-2285.

Apple, F. S., Panteghini, M., Ravkilde, J., Mair, J., Wu, A. H. B., Tate, J., Pagan, F., Christenson, R. H., Jaffe, A. S. and IFCC, C. o. S. o. M. o. C. D. o. t. (2005) 'Quality specifications for B-type natriuretic peptide assays.' *Clinical Chemistry*, 51(3), Apr, pp. 486-493.

Aurigemma, G. P., Gottdiener, J. S., Shemanski, L., Gardin, J. and Kitzman, D. (2001) 'Predictive value of systolic and diastolic function for incident congestive heart failure in the elderly: the cardiovascular health study.' *JAC*, 37(4), Apr 15, pp. 1042-1048.

Azzazy, H. M. E. and Christenson, R. H. (2003) 'B-type natriuretic peptide: physiologic role and assay characteristics.' *Heart failure reviews*, 8(4) pp. 315-320.

Badgett, R. G., Mulrow, C. D., Otto, P. M. and Ramírez, G. (1996) 'How well can the chest radiograph diagnose left ventricular dysfunction?' *Journal of General Internal Medicine*, 11(10), Oct, pp. 625-634.

Baldasseroni, S., Gentile, A., Gorini, M., Marchionni, N., Marini, M., Masotti, G., Porcu, M., Maggioni, A. and Italian Network on Congestive Heart Failure, I. (2003) 'Intraventricular conduction defects in patients with congestive heart failure: left but not right bundle branch block is an independent predictor of prognosis. A report from the Italian Network on Congestive Heart Failure (IN-CHF database).' *Italian Heart Journal: Official Journal of the Italian Federation of Cardiology*, 4(9), Sep 01, pp. 607-613.

Bando, M., Ishii, Y., Sugiyama, Y. and Kitamura, S. (1999) 'Elevated plasma brain natriuretic peptide levels in chronic respiratory failure with cor pulmonale.' *Respiratory Medicine*, 93(7), Jul, pp. 507-514.

Barcarse, E., Kazanegra, R. and Chen, A. (2004) 'Combination of B - Type Natriuretic Peptide Levels and Non - Invasive Hemodynamic Parameters in Diagnosing Congestive Heart Failure in the Emergency ....' *Congestive Heart ...*,

Baxter, G. F. (2004) 'The natriuretic peptides.' *Basic Research in Cardiology*, 99(2) pp. 71-75.

Bayes-Genis, A., Santalo-Bel, M., Zapico-Muniz, E., Lopez, L., Cotes, C., Bellido, J., Leta, R., Casan, P. and Ordonez-Llanos, J. (2004) 'N-terminal probrain natriuretic peptide (NT-proBNP) in the emergency diagnosis and in-hospital monitoring of patients with dyspnoea and ventricular dysfunction.' *European Journal of Heart Failure*, 6(3), 15 Mar, pp. 301-308.

Bayes-Genis, A., Lloyd-Jones, D., van Kimmenade, R., Lainchbury, J., Richards, A., Ordonez-Llanos, J., Santalo, M., Pinto, Y. and Januzzi Jr, J. (2007) 'Effect of body mass index on diagnostic and prognostic usefulness of amino-terminal pro-brain natriuretic peptide in patients with acute dyspnea.' *Archives of Internal Medicine*, 167(4) pp. 400-407.

Begum, F., Ghosh, D., Tseng, G. C. and Feingold, E. (2012) 'Comprehensive literature review and statistical considerations for GWAS meta-analysis.' *Nucleic Acids Research*, 40(9), Jun 11, pp. 3777-3784.

Behnes, M., Brueckmann, M., Ahmad-Nejad, P., Lang, S., Wolpert, C., Elmas, E., Kaelsch, T., Gruettner, J., Weiss, C., Borggrefe, M. and Neumaier, M. (2009) 'Diagnostic performance and cost effectiveness of measurements of plasma N-terminal pro brain natriuretic peptide in patients presenting with acute dyspnea or peripheral edema.' *International Journal of Cardiology*, 135(2) pp. 165-174.

Bellenger, N. G., Burgess, M. I., Ray, S. G., Lahiri, A., Coats, A. J., Cleland, J. G. and Pennell, D. J. (2000) 'Comparison of left ventricular ejection fraction and volumes in heart failure by echocardiography, radionuclide ventriculography and cardiovascular magnetic resonance; are they interchangeable?' *European Heart Journal*, 21(16), Aug, pp. 1387-1396.

Berdagué, P., Caffin, P.-Y., Barazer, I., Vergnes, C., Sedighian, S., Letrillard, S., Pilosof, R., Goutorbe, F., Piot, C. and Reny, J.-L. (2006) 'Use of N-terminal prohormone brain natriuretic peptide assay for etiologic diagnosis of acute dyspnea in elderly patients.' *American Heart Journal*, 151(3), Apr 01, pp. 690-698.

Bewick, V., Cheek, L. and Ball, J. (2004) 'Statistics review 13: receiver operating characteristic curves.' *Crit Care*, 8(6), Dec, pp. 508-512.

Bhalla, V., Willis, S. and Maisel, A. (2004) 'B-Type Natriuretic Peptide: The Level and the Drug - Partners in the Diagnosis and Management of Congestive Heart Failure.' *Congestive Heart Failure*, 10(S1) pp. 3 - 27.

BHF (2008) 'BHF Coronary Heart Disease Statistics.'

Binanay, C., Califf, R. M. and Hasselblad, V. (2005) 'Evaluation study of congestive heart failure and pulmonary artery catheterization effectiveness: the ESCAPE trial.' *JAMA: the journal of ...*,

Biosite. (2007) *Triage BNP product insert*. San Diego, USA: Biosite Incorporated. 2007 viewed

Bland, M. (2000) *An Introduction To Medical Statistics*. Oxford: Oxford University Press.

Boldanova, T., Noveanu, M., Breidthardt, T., Potocki, M., Reichlin, T., Taegtmeyer, A., Christ, M., Laule, K., Stelzig, C. and Mueller, C. (2010) 'Impact of history of heart failure on diagnostic and prognostic value of

BNP: Results from the B-type Natriuretic Peptide for Acute Shortness of Breath Evaluation (BASEL) Study☆.' *International Journal of Cardiology*, 142(3), Jul 23, pp. 265-272.

Boomsma, F. and van den Meiracker, A. (2001) 'Plasma A- and B-type natriuretic peptides: physiology, methodology and clinical use.' *Cardiovascular Research*, 51 pp. 442-449.

Boyd, K. J., Murray, S. A., Kendall, M., Worth, A., Benton, T. F. and Clausen, H. (2004) 'Living with advanced heart failure: a prospective, community based study of patients and their carers.' *European Journal of Heart Failure*, 6(5), Jun 01, pp. 585-591.

Breiman, L., Friedman, J., Olshen, R. and Stone, C. (1995). CART: Tree-Structured Non-Parametric Data Analysis. San Diego, USA: Salford Systems.

Brenner, H. (1996) 'Measures of differential diagnostic value of diagnostic procedures.' *Journal of clinical epidemiology*, 49(12) pp. 1435-1439.

Brooksby, P., Batin, P., Nolan, J., Lindsay, S., Andrews, R., Mullen, M., Baig, W., Flapan, A., Prescott, R., Neilson, J., Cowley, A. and Fox, K. (1999) 'The relationship between QT intervals and mortality in ambulant patients with chronic heart failure. The united kingdom heart failure evaluation and assessment of risk trial (UK-HEART)[see comment].' *European Heart Journal*, 20(18), Sep 01, pp. 1335-1341.

Campbell, D. J., Munir, V., Hennessy, O. F. and Dent, A. W. (2001) 'Plasma amino-terminal pro-brain natriuretic peptide levels in subjects presenting to the Emergency Department with suspected acute coronary syndrome: possible role in selecting patients for follow up?' *Internal Medicine Journal*, 31(4), Jun, pp. 211-219.

Carlson, K., Lee, D., Goroll, A., Leahy, M. and Johnson, R. (1985) 'An analysis of physicians' reasons for prescribing long-term digitalis therapy in outpatients\* 1.' *Journal of chronic diseases*, 38(9) pp. 733-739.

Casscells, W., Vasseghi, M. F., Siadaty, M. S., Madjid, M., Siddiqui, H., Lal, B. and Payvar, S. (2005) 'Hypothermia is a bedside predictor of imminent death in patients with congestive heart failure.' *American Heart Journal*, 149(5), Jun, pp. 927-933.

CEM. (2014) *College of Emergency Medicine Curriculum*. London:  
[Online] [Accessed on Feb]  
<http://secure.collemergencymed.ac.uk/Training-Exams/Curriculum/CurriculumfromAugust2010/default.asp>.

Cepeda, M. S. (2003) 'Comparison of Logistic Regression versus Propensity Score When the Number of Events Is Low and There Are Multiple Confounders.' *American Journal of Epidemiology*, 158(3), Aug 01, pp. 280-287.

Chappell, R. and Fine, J. (2010). Logistic Regression. In Chow, S. (Ed.), *Encyclopedia of Biopharmaceutical Statistics* (Third ed., pp. 725-729): CRC Press.

Chen, J. T. (1992) 'Radiographic diagnosis of heart failure.' *Heart disease and stroke : a journal for primary care physicians*, 1(2), Apr, pp. 58-63.

Chenevier-Gobeaux, C., Claessens, Y., Voyer, S., Desmoulins, D. and Ekindjian, O. (2005) 'Influence of renal function on N-terminal pro-brain natriuretic peptide (NT-proBNP) in patients admitted for dyspnoea in the Emergency Department: Comparison with brain natriuretic peptide (BNP).' *Clinica Chimica Acta*, 361(1-2), Nov, pp. 167-175.

Choi, S., Park, D., Lee, S., Hong, Y., Kim, S. and Lee, J. (2007) 'Cut-off values of B-type natriuretic peptide for the diagnosis of congestive heart failure in patients with dyspnoea visiting emergency departments: a study on Korean patients visiting emergency departments.[see comment].' *Emergency Medicine Journal*, 24(5), May, pp. 343-347.

Chung, J. M., Lee, S. D., Kang, D. I., Kwon, D. D. and Kim, K. S. (2009) 'ScienceDirect.com - Urology - Prevalence and Associated Factors of Overactive Bladder in Korean Children 5-13 Years Old: A Nationwide Multicenter Study.' *Urology*,

Chung, P. and Hermann, L. (2006) 'Acute decompensated heart failure: formulating an evidence-based approach to diagnosis and treatment (part I).' *Mount Sinai Journal of Medicine*, 73(2), Apr 01, pp. 506-515.

Cinar, O., Cevik, E., Acar, A., Kaya, C., Ardic, S., Comert, B., Yokusoglu, M., Bilgi, C., Meisner, M. and Madsen, T. (2012) 'Evaluation of mid-regional pro-atrial natriuretic peptide, procalcitonin, and mid-regional pro-adrenomedullin for the diagnosis and risk stratification of dyspneic ED patients.' *Am J Emerg Med*, 30(9), Nov, pp. 1915-1920.

Clarke, S., Parris, R. and Reynard, K. (2005) 'Core-peripheral temperature gradient as a diagnostic test in dyspnoea.' *Emergency Medicine Journal*, 22(9), Sep 01, pp. 633-635.

Cleland, J. G. and Habib, F. (1996) 'Assessment and diagnosis of heart failure.' *Journal of Internal Medicine*, 239(4), May, pp. 317-325.

Cleland, J. G. F., Dargie, H. J., Hardman, S., McDonagh, T., Mitchell, P. and Wales, N. H. F. A. T. f. E. a. (2012). National Heart Failure Audit April 2011 - March 2012 (pp. 1-62): Heart.

Clerico, A. (2002) 'Pathophysiological and clinical relevance of circulating levels of cardiac natriuretic hormones: are they merely markers of cardiac disease?' *Clinical chemistry and laboratory medicine : CCLM / FESCC*, 40(8), Aug, pp. 752-760.

Clerico, A. (2004) 'Diagnostic Accuracy and Prognostic Relevance of the Measurement of Cardiac Natriuretic Peptides: A Review.' *Clinical Chemistry*, 50(1), Sep 19, pp. 33-50.

Clerico, A., Fontana, M., Zyw, L., Passino, C. and Emdin, M. (2007) 'Comparison of the diagnostic accuracy of brain natriuretic peptide (BNP) and the N-terminal part of the propeptide of BNP immunoassays in chronic and acute heart failure: a systematic review.' *Clinical Chemistry*, 53(5) pp. 813-822.

Coats, T. J. (2006) 'Consent for emergency care research: the Mental Capacity Act 2005.' *Emergency Medicine Journal*, 23(12), Dec 01, pp. 893-894.

Cochran, W. G. (1954) 'The combination of estimates from different experiments.' *Biometrics*, 10(1) pp. 101-129.

Cochrane. (2012). Review Manager (RevMan). In Centre, T. N. C. (Ed.). Copenhagen: The Cochrane Collaboration.

Coderre, S., Mandin, H., Harasym, P. and Fick, G. (2003) 'Diagnostic reasoning strategies and diagnostic success.' *Medical Education*, 37 pp. 695-703.

Collins, S., Lindsell, C., Peacock, W., Eckert, D., Askew, J. and Storrow, A. (2006) 'Clinical characteristics of emergency department heart failure patients initially diagnosed as non-heart failure.' *BMC Emergency Medicine*, 6(11)

Collins, S., Lindsell, C., Storrow, A., Abraham, W., Adhere Scientific Advisory Committee, I. and Study, G. (2006) 'Prevalence of negative chest radiography results in the emergency department patient with decompensated heart failure.[see comment].' *Annals of Emergency Medicine*, 47(1), Feb 01, pp. 13-18.

Collins, S., Peacock, W., Lindsell, C., Clopton, P., Diercks, D., Hiestand, B., Hogan, C., Kontos, M., Mueller, C., Nowak, R., Chen, W.-J., Huang, C.-H., Abraham, W., Amsterdam, E., Breidthardt, T., Daniels, L., Hasan, A., Hudson, M., McCord, J., Naz, T., Wagoner, L. and Maisel, A. (2009) 'S3 detection as a diagnostic and prognostic aid in emergency department patients with acute dyspnea.' *Annals of Emergency Medicine*, 53(6), Jul 01, pp. 748-757.

Collins, S. P., Lindsell, C. J., Peacock, W. F., Hedger, V. D., Askew, J., Eckert, D. C. and Storrow, A. B. (2006) 'The combined utility of an S3 heart sound and B-type natriuretic peptide levels in emergency department patients with dyspnea.' *Journal of Cardiac Failure*, 12(4), Jun 01, pp. 286-292.

Collinson, P. O., Barnes, S. C., Gaze, D. C., Galasko, G., Lahiri, A. and Senior, R. (2004) 'Analytical performance of the N terminal pro B type natriuretic peptide (NT-proBNP) assay on the Elecsys 1010 and 2010 analysers.' *European journal of heart failure : journal of the Working Group on Heart Failure of the European Society of Cardiology*, 6(3), Apr 15, pp. 365-368.

Colucci, W. S., Elkayam, U., Horton, D. P., Abraham, W. T., Bourge, R. C., Johnson, A. D., Wagoner, L. E., Givertz, M. M., Liang, C. S., Neibaur, M., Haught, W. H. and LeJemtel, T. H. (2000) 'Intravenous nesiritide, a natriuretic peptide, in the treatment of decompensated congestive heart failure. Nesiritide Study Group.' *The New England Journal of Medicine*, 343(4), Jul 27, pp. 246-253.

Concato, J., Feinstein, A. R. and Holford, T. R. (1993) 'The risk of determining risk with multivariable models.' *Annals of internal medicine*, 118(3) p. 201.

Coste, J., Jourdain, P. and Pouchot, J. (2006) 'A gray zone assigned to inconclusive results of quantitative diagnostic tests: Application to the use of brain natriuretic peptide for diagnosis of heart failure in acute dyspneic patients.' *Clinical Chemistry*, 52(12), Dec 01, pp. 2229-2235.

Costello-Boerrigter, L. C., Boerrigter, G., Cataliotti, A., Harty, G. J. and Burnett, J. C. (2010) 'Renal and anti-aldosterone actions of vasopressin-2 receptor antagonism and B-type natriuretic peptide in experimental heart failure.' *Circulation. Heart failure*, 3(3), Jun, pp. 412-419.

Couchoud, C., Pozet, N., Labeeuw, M. and Pouteil-Noble, C. (1999) 'Screening early renal failure: Cut-off values for serum creatinine as an indicator of renal impairment.' *Kindney International*, 55 pp. 1878-1884.

Cowie, M. and Chandrasekaran, B. (2011). Clinical features and medical treatments. In Warrell, D. (Ed.), *Oxford Textbook of Medicine*. Oxford, UK: 'Oxford University Press'.

Cowie, M., Wood, D. A., Coats, A. J., Thompson, S. G., Poole-Wilson, P. A., Suresh, V. and Sutton, G. C. (1999) 'Incidence and aetiology of heart failure; a population-based study.' *European Heart Journal*, 20(6), Apr 01, pp. 421-428.

Craig, J., Bradbury, I., Cummins, E., Downie, S., Foster, L. and Stout, A. (2005) 'Health Technology Assessment Report 6 The use of B-type natriuretic peptides (BNP and NT-proBNP) in the investigation of patients with suspected heart failure.' pp. 1-142.

Croskerry, P. (2005) 'Advances in Patient Safety Diagnostic Failure: A Cognitive and Affective Approach.' In Henriksen, K., Battles, J. B., Marks, E. S. and Lewin, D. I. (eds.) *Advances in Patient Safety: From Research to Implementation (Volume 2: Concepts and Methodology)*. Rockville (MD): Agency for Healthcare Research and Quality (US),

Croskerry, P. (2013) 'From mindless to mindful practice--cognitive bias and clinical decision making.' *N Engl J Med*, 368(26), Jun 27, pp. 2445-2448.

Daniels, L., Bhalla, V., Clopton, P., Hollander, J., Guss, D., McCullough, P., Nowak, R., Green, G., Saltzberg, M., Ellison, S., Bhalla, M., Jesse, R. and Maisel, A. (2006) 'B-type natriuretic peptide (BNP) levels and ethnic disparities in perceived severity of heart failure: results from the Rapid

Emergency Department Heart Failure Outpatient Trial (REDHOT) multicenter study of BNP levels and emergency department decision making in patients presenting with shortness of breath.' *Journal of Cardiac Failure*, 12(4), Jun 01, pp. 281-285.

Daniels, L., Clopton, P., Bhalla, V., Krishnaswamy, P., Nowak, R., McCord, J., Hollander, J., Duc, P., Omland, T., Storrow, A., Abraham, W., Wu, A., Steg, P., Westheim, A., Knudsen, C., Perez, A., Kazanegra, R., Herrmann, H., McCullough, P. and Maisel, A. (2006) 'How obesity affects the cut-points for B-type natriuretic peptide in the diagnosis of acute heart failure. Results from the Breathing Not Properly Multinational Study.' *American Heart Journal*, 151(5), Jun 01, pp. 999-1005.

Dao, Q., Krishnaswamy, P., Kazanegra, R., Harrison, A., Amirnovin, R., Lenert, L., Clopton, P., Alberto, J., Hlavin, P. and Maisel, A. S. (2001) 'Utility of B-type natriuretic peptide in the diagnosis of congestive heart failure in an urgent-care setting.' *Journal of the American College of Cardiology*, 37(2), Feb, pp. 379-385.

Davenport, C., Cheng, E. Y. L., Kwok, Y. T. T., Lai, A. H. O., Wakabayashi, T., Hyde, C. and Connock, M. (2006) 'Assessing the diagnostic test accuracy of natriuretic peptides and ECG in the diagnosis of left ventricular systolic dysfunction: a systematic review and meta-analysis.' *The British journal of general practice : the journal of the Royal College of General Practitioners*, 56(522), Feb 01, pp. 48-56.

Davie, A. P., Francis, C. M., Caruana, L., Sutherland, G. R. and McMurray, J. J. (1997) 'Assessing diagnosis in heart failure: which features are any use?' *Qjm*, 90(5) pp. 335-339.

Davie, A. P., Francis, C. M., Love, M. P., Caruana, L., Starkey, I. R., Shaw, T. R., Sutherland, G. R. and McMurray, J. J. (1996) 'Value of the electrocardiogram in identifying heart failure due to left ventricular systolic dysfunction.' *BMJ*, 312(7025), Feb 27, p. 222.

Davis, M., Espiner, E., Richards, G., Billings, J., Town, I., Neill, A., Drennan, C., Richards, M., Turner, J. and Yandle, T. (1994) 'Plasma brain natriuretic peptide in assessment of acute dyspnoea.' *Lancet*, 343(8895), Mar 19, pp. 440-444.

de Bold, A. J., Borenstein, H. B., Veress, A. T. and Sonnenberg, H. (1981) 'A rapid and potent natriuretic response to intravenous injection of atrial myocardial extract in rats.' *Life sciences*, 28(1), Feb 05, pp. 89-94.

- de Lemos JA, M. (2003) 'B-type natriuretic peptide in cardiovascular disease.' *Lancet*, 362 pp. 316-322.
- Deakin, M. A. B. (1981) 'The development of the Laplace transform, 1737-1937.' *Archive for History of Exact Sciences*, 25(4), Dec, pp. 343-390.
- Deeks, J. J. (2001) 'Systematic reviews in health care: Systematic reviews of evaluations of diagnostic and screening tests.' *BMJ*, 323(7305), Jul 21, pp. 157-162.
- deFilippi, C., Seliger, S., Maynard, S. and Christenson, R. (2007) 'Impact of renal disease on natriuretic peptide testing for diagnosing decompensated heart failure and predicting mortality.' *Clinical Chemistry*, 53(8), Aug 01, pp. 1511-1519.
- Demšar, J., Curk, T., Erjavec, A., Gorup, Č., Hočevar, T., Milutinovič, M., Možina, M., Polajnar, M., Toplak, M., Starič, A., Štajdohar, M., Umek, L., Žagar, L., Žbontar, J., Žitnik, M. and Zupan, B. (2013) 'Orange: data mining toolbox in python.' *The Journal of Machine Learning Research*, 14(1), Feb 01, pp. 2349-2353.
- DerSimonian, R. and Laird, N. (1986) 'Meta-analysis in clinical trials.' *Controlled clinical trials*, 7(3) pp. 177-188.
- Devillé, W. L., Buntinx, F., Bouter, L. M., Montori, V. M., de Vet, H. C. W., van der Windt, D. A. W. M. and Bezemer, P. D. (2002) 'BMC Medical Research Methodology 2002 29 2002 DevilléConducting systematic reviews of diagnostic.' *BMC Medical Research Methodology*, 2, Jul 17, pp. 1-13.
- Dhalla, N. S., Afzal, N., Beamish, R. E., Naimark, B., Takeda, N. and Nagano, M. (1993) 'Pathophysiology of cardiac dysfunction in congestive heart failure.' *The Canadian journal of cardiology*, 9(10) p. 873.
- Di Bari, M., Pozzi, C., Cavallini, M. C., Innocenti, F., Baldereschi, G., de Alfieri, W., Antonini, E., Pini, R., Masotti, G. and Marchionni, N. (2004) 'The diagnosis of heart failure in the community. Comparative validation of four sets of criteria in unselected older adults: the ICARE Dicomano Study.' *Journal of the American College of Cardiology*, 44(8), Oct 19, pp. 1601-1608.

DiagnosisPro. (2014) *DiagnosisPro*. [Online] [Accessed on Feb] [http://en.diagnosispro.com/differential\\_diagnosis-for/poisoning-specific-agent-dyspnea/25103-154-100.html](http://en.diagnosispro.com/differential_diagnosis-for/poisoning-specific-agent-dyspnea/25103-154-100.html)

Dieplinger, B., Gegenhuber, A., Haltmayer, M. and Mueller, T. (2009) 'Evaluation of novel biomarkers for the diagnosis of acute destabilised heart failure in patients with shortness of breath.' *Heart*, 95(18), Sep, pp. 1508-1513.

Dillingham, M. A. and Anderson, R. J. (1986) 'Inhibition of vasopressin action by atrial natriuretic factor.' *Science*, 231(4745) pp. 1572-1573.

DOH. (2014) *AccidentAndEmergency*. [Online] [Accessed [http://www.dh.gov.uk/en/Publicationsandstatistics/Statistics/Performance dataandstatistics/AccidentandEmergency/index.htm](http://www.dh.gov.uk/en/Publicationsandstatistics/Statistics/Performance%20dataandstatistics/AccidentandEmergency/index.htm)]

Domingos, P. and Pazzani, M. (1996) 'Beyond independence: Conditions for the optimality of the simple bayesian classifier.' *Proc 13th Intl Conf Machine Learning*,

Drazner, M. H., Rame, J. E., Stevenson, L. W. and Dries, D. L. (2001) 'Prognostic importance of elevated jugular venous pressure and a third heart sound in patients with heart failure.' *New England Journal of Medicine*, 345(8), Aug 23, pp. 574-581.

Durbridge, T. C., Edwards, F., Edwards, R. G. and Atkinson, M. (1976) 'Evaluation of benefits of screening tests done immediately on admission to hospital.' *Clinical Chemistry*, 22(7) pp. 968-971.

Eckstein, J., Potocki, M., Murray, K., Breidhardt, T., Ziller, R., Mosimann, T., Klima, T., Hoeller, R., Moehring, B., Sou, S. M., Rubini Gimenez, M., Morgenthaler, N. G. and Mueller, C. (2012) 'Direct comparison of mid-regional pro-atrial natriuretic peptide with N-terminal pro B-type natriuretic peptide in the diagnosis of patients with atrial fibrillation and dyspnoea.' *Heart*, 98(20), Oct, pp. 1518-1522.

el Mahmoud, R., Alibay, Y., Brun-Ney, D., Boulard, J., Dubourg, O., Puy, H. and Jondeau, G. (2006) 'Type B natriuretic peptide (BNP) versus n-terminal type B natriuretic propeptide in the diagnosis of cardiac failure in the elderly over 75 population. [French].' *Archives des Maladies du Coeur et des Vaisseaux*, 99(3), Apr 01, pp. 201-207.

Emdin, M., Passino, C., Prontera, C., Fontana, M., Poletti, R., Gabutti, A., Mammini, C., Giannoni, A., Zyw, L., Zucchelli, G. and Clerico, A. (2007) 'Comparison of brain natriuretic peptide (BNP) and amino-terminal ProBNP for early diagnosis of heart failure.' *Clinical Chemistry*, 53(7), Jul, pp. 1289-1297.

Eriksson, H., Svärdsudd, K., Larsson, B., Ohlson, L. O., Welin, L., Tibblin, G. and Wilhelmsen, L. (1987) 'Dyspnoea in a cross-sectional and a longitudinal study of middle-aged men: the Study of Men Born in 1913 and 1923.' *European Heart Journal*, 8(9), Sep 01, pp. 1015-1023.

Eriksson, H., Caidahl, K., Larsson, B., Ohlson, L. O., Welin, L., Wilhelmsen, L. and Svärdsudd, K. (1987) 'Cardiac and pulmonary causes of dyspnoea--validation of a scoring test for clinical-epidemiological use: the Study of Men Born in 1913.' *European Heart Journal*, 8(9), Sep 01, pp. 1007-1014.

Eurlings, L. W., Sanders-van Wijk, S., van Kimmenade, R., Osinski, A., van Helmond, L., Vallinga, M., Crijns, H. J., van Dieijen-Visser, M. P., Brunner-La Rocca, H. P. and Pinto, Y. M. (2012) 'Multimarker strategy for short-term risk assessment in patients with dyspnea in the emergency department: the MARKED (Multi mARKer Emergency Dyspnea)-risk score.' *J Am Coll Cardiol*, 60(17), Oct 23, pp. 1668-1677.

Fagan, T. (1975) 'Nomogram for Bayes's Theorem.' *New England Journal of Medicine*, 293(5), Jul 31, pp. 257-257.

Feinstein, A. R. (1987) *Clinimetrics*. US: Yale University.

Feinstein, A. R. (1990) 'The inadequacy of binary models for the clinical reality of three-zone diagnostic decisions.' *Journal of clinical epidemiology*, 43(1), Feb 01, pp. 109-113.

Felker, G. M., Adams Jr, K. F., Konstam, M. A., O'Connor, C. M. and Gheorghade, M. (2003) 'The problem of decompensated heart failure: Nomenclature, classification, and risk stratification.' *American Heart Journal*, 145(2, Supplement), 2//, pp. S18-S25.

Felker, G. M., Hasselblad, V., Tang, W. H. W., Hernandez, A. F., Armstrong, P. W., Fonarow, G. C., Voors, A. A., Metra, M., McMurray, J. J. V., Butler, J., Heizer, G. M., Dickstein, K., Massie, B. M., Atar, D., Troughton, R. W., Anker, S. D., Califf, R. M., Starling, R. C. and O'Connor, C. M. (2012) 'Troponin I in acute decompensated heart

failure: insights from the ASCEND-HF study.' *European journal of heart failure : journal of the Working Group on Heart Failure of the European Society of Cardiology*, 14(11), Nov, pp. 1257-1264.

Fleischer, D., Espiner, E., Yandle, T., Livesey, J., Billings, J., Town, I. and Richards, A. (1997) 'Rapid assay of plasma brain natriuretic peptide in the assessment of acute dyspnoea.' *The New Zealand medical journal*, 110(1039) p. 71.

Fletcher, L. and Thomas, D. (2001) 'Congestive heart failure: understanding the pathophysiology and management.' *Journal of the American Academy of Nurse Practitioners*, 13(6), Jul, pp. 249-257.

Fonseca, C., Morais, H., Mota, T., Matias, F., Costa, C., Gouveia-Oliveira, A. and Ceia, F. (2004) 'The diagnosis of heart failure in primary care: value of symptoms and signs.' *European journal of heart failure : journal of the Working Group on Heart Failure of the European Society of Cardiology*, 6(6) pp. 795-800.

Forgacs, P. (1978) 'The Functional Basis of Pulmonary Sounds.' *Chest*, 73(2) pp. 399-405.

Fox, K. F., Cowie, M. R., Wood, D. A., Coats, A. J., Gibbs, J. S., Underwood, S. R., Turner, R. M., Poole-Wilson, P. A., Davies, S. W. and Sutton, G. C. (2001) 'Coronary artery disease as the cause of incident heart failure in the population.' *European Heart Journal*, 22(3), Mar 01, pp. 228-236.

Frassi, F., Gargani, L., GLIGOROVA, S., CIAMPI, Q., MOTTOLA, G. and Picano, E. (2007) 'Clinical and echocardiographic determinants of ultrasound lung comets☆.' *European Journal of Echocardiography*, 8(6), Dec, pp. 474-479.

Friedman, J. and Popescu, B. (2003) *Importance Sampled Learning Ensembles*. Stanford University, CA: Stanford University.

Fromm, R. E., Varon, J. and Gibbs, L. R. (1995) 'Congestive heart failure and pulmonary edema for the emergency physician.' *The Journal of emergency medicine*, 13(1), Feb, pp. 71-87.

Garcia, X., Simon, P., Guyette, F. X., Ramani, R., Alvarez, R., Quintero, J. and Pinsky, M. R. (2013) 'Noninvasive assessment of acute dyspnea in the ED.' *Chest*, 144(2), Aug, pp. 610-615.

Gargani, L., Frassi, F., Soldati, G., Tesorio, P., Gheorghide, M. and Picano, E. (2008) 'Ultrasound lung comets for the differential diagnosis of acute cardiogenic dyspnoea: a comparison with natriuretic peptides.' *European Journal of Heart Failure*, 10(1), Feb 01, pp. 70-77.

Gatsonis, C. and Paliwal, P. (2006) 'Meta-analysis of diagnostic and screening test accuracy evaluations: methodologic primer.' *American Journal of Roentgenology*, 187(2) pp. 1-11.

Gatt, M. E. (2003) 'Chest radiographs in the emergency department: is the radiologist really necessary?' *Postgraduate medical journal*, 79(930), May 01, pp. 214-217.

Gegenhuber, A., Struck, J., Poelz, W., Pacher, R., Morgenthaler, N., Bergmann, A., Haltmayer, M. and Mueller, T. (2006) 'Midregional pro-A-type natriuretic peptide measurements for diagnosis of acute destabilized heart failure in short-of-breath patients: comparison with B-type natriuretic peptide (BNP) and amino-terminal proBNP.' *Clinical Chemistry*, 52(5), Jun 01, pp. 827-831.

Gheorghide, M. (2005) 'Reassessing treatment of acute heart failure syndromes: the ADHERE Registry.' *European Heart Journal Supplements*, 7(Suppl B) pp. B13-B19.

Gheorghide, M., Gattis Stough, W., Adams, K. F., Jaffe, A. S., Hasselblad, V. and O'Connor, C. M. (2005) 'The Pilot Randomized Study of Nesiritide Versus Dobutamine in Heart Failure (PRESERVED-HF).' *The American Journal of Cardiology*, 96(6A), Sep 19, pp. 18G-25G.

Gillespie, N., McNeill, G., Pringle, T., Ogston, S., Struthers, A. and Pringle, S. (1997) 'Cross sectional study of contribution of clinical assessment and simple cardiac investigations to diagnosis of left ventricular systolic dysfunction in patients admitted with acute dyspnoea.[see comment].' *BMJ*, 314(7085), Apr 29, pp. 936-940.

Glas, A. S., Lijmer, J. G., Prins, M. H., Bossel, G. J. and Bossuyt, P. M. M. (2003) 'The diagnostic odds ratio: a single indicator of test performance.' *Journal of clinical epidemiology*, 56(11), Nov, pp. 1129-1135.

Glass, G. V. (1976) 'Primary, secondary, and meta-analysis of research.' *Educational researcher*, 5(10) pp. 3-8.

Gluecker, T., Capasso, P., Schnyder, P., Gudinchet, F., Schaller, M. D., Revelly, J. P., Chiolero, R., Vock, P. and Wicky, S. (1999) 'Clinical and radiologic features of pulmonary edema.' *Radiographics*, 19(6), Oct, pp. 1507-1531; discussion 1532-1503.

Gnani, S. and Ellis, C. (2002) 'Trends in hospital admissions and case fatality due to heart failure in England, 1990/91 to 1999/2000.' *Health Statistics Quarterly*, 13

Go, A. S., Mozaffarian, D., Roger, V. L., Benjamin, E. J., Berry, J. D., Borden, W. B., Bravata, D. M., Dai, S., Ford, E. S. and Fox, C. S. (2013) 'Heart Disease and Stroke Statistics—2013 Update A Report From the American Heart Association.' *Circulation*, 127(1) pp. e6-e245.

Goetze, J. P., Videbaek, R., Boesgaard, S., Aldershvile, J., Rehfeld, J. F. and Carlsen, J. (2004) 'Pro-brain natriuretic peptide as marker of cardiovascular or pulmonary causes of dyspnea in patients with terminal parenchymal lung disease.' *The Journal of Heart and Lung Transplantation*, 23(1), Feb, pp. 80-87.

Goodman, S. (1999) 'Toward Evidence-Based Medical Statistics. 1: The P Value Fallacy.' *Annals of internal medicine*, 130 pp. 995-1004.

Green, S., Martinez-Rumayor, A., Gregory, S., Baggish, A., O'Donoghue, M., Green, J., Lewandrowski, K. and Januzzi, J. (2008) 'Clinical uncertainty, diagnostic accuracy, and outcomes in emergency department patients presenting with dyspnea.' *Archives of Internal Medicine*, 168(7), May 14, pp. 741-748.

Groen, G. and Patel, V. (1985) 'Medical problem-solving: some questionable assumptions.' *Medical Education*, 19 pp. 95-100.

Groenning, B. A., Raymond, I., Hildebrandt, P. R., Nilsson, J. C., Baumann, M. and Pedersen, F. (2004) 'Diagnostic and prognostic evaluation of left ventricular systolic heart failure by plasma N-terminal pro-brain natriuretic peptide concentrations in a large sample of the general population.' *Heart*, 90(3), Apr, pp. 297-303.

Groenning, B. A., Nilsson, J. C., Sondergaard, L., Pedersen, F., Trawinski, J., Baumann, M., Larsson, H. B. W. and Hildebrandt, P. R. (2002) 'Detection of left ventricular enlargement and impaired systolic function with plasma N-terminal pro brain natriuretic peptide concentrations.' *American Heart Journal*, 143(5), Jun, pp. 923-929.

Gruson, D., Rousseau, M. F., Ahn, S., van Linden, F., Thys, F., Ketelslegers, J.-M. and Verschuren, F. (2008) 'Accuracy of N - terminal - pro - atrial natriuretic peptide in patients admitted to emergency department.' *Scandinavian Journal of Clinical & Laboratory Investigation*, 68(5), Feb, pp. 410-414.

Hammerer-Lercher, A., Puschendorf, B., Mair, J. and Innsbruck, A. (2010) 'B-type natriuretic peptides as powerful markers in cardiac diseases—analytical and clinical aspects.' *J Lab Med*, 30, Jun 26, pp. 165-184.

Hammerer-Lercher, A., Neubauer, E., Muller, S., Pachinger, O., Puschendorf, B. and Mair, J. (2001) 'Head-to-head comparison of N-terminal pro-brain natriuretic peptide, brain natriuretic peptide and N-terminal pro-atrial natriuretic peptide in diagnosing left ventricular dysfunction.' *Clinica chimica acta; international journal of clinical chemistry*, 310(2), Aug 20, pp. 193-197.

An empirical comparison of methods for meta-analysis of diagnostic accuracy showed hierarchical models are necessary, 61 Cong. Rec. 1095-1103 (2008).

Hardy, R. J. and Thompson, S. G. (1998) 'Detecting and describing heterogeneity in meta-analysis.' *Statistics in Medicine*, 17(8), May 30, pp. 841-856.

Harrison, A., Morrison, L. K., Krishnaswamy, P., Kazanegra, R., Clopton, P., Dao, Q., Hlavin, P. and Maisel, A. S. (2002) 'B-type natriuretic peptide predicts future cardiac events in patients presenting to the emergency department with dyspnea.' *YMEM*, 39(2) pp. 131-138.

Havelka, E. G., Rzechula, K. H., Bryant, T. O., Anneken, S. M. and Kulstad, E. B. (2011) 'Correlation between impedance cardiography and B-type natriuretic peptide levels in dyspneic patients.' *J Emerg Med*, 40(2), Feb, pp. 146-150.

Haynes, B. and Haines, A. (1998) 'Getting research findings into practice. Barriers and bridges to evidence based clinical practice.' *BMJ*, 317 pp. 273-276.

He, J., Ogden, L., Bazzano, L., Vupputuri, S., Loria, C. and Whelton, P. (2001) 'Risk factors for congestive heart failure in US men and women: NHANES I epidemiologic follow-up study.' *Archives of Internal Medicine*, 161(7), May 09, pp. 996-1002.

Heidenreich, P. A., Albert, N. M., Allen, L. A., Bluemke, D. A., Butler, J., Fonarow, G. C., Ikonomidis, J. S., Khavjou, O., Konstam, M. A., Maddox, T. M., Nichol, G., Pham, M., Pina, I. L., Trogdon, J. G., American Heart Association Advocacy Coordinating, C., Council on Arteriosclerosis, T., Vascular, B., Council on Cardiovascular, R., Intervention, Council on Clinical, C., Council on, E., Prevention and Stroke, C. (2013) 'Forecasting the impact of heart failure in the United States: a policy statement from the American Heart Association.' *Circ Heart Fail*, 6(3), May, pp. 606-619.

Hendel, R. C., Berman, D. S., Di Carli, M. F., Heidenreich, P. A., Henkin, R. E., Pellikka, P. A., Pohost, G. M. and Williams, K. A. (2009) 'ACCF/ASNC/ACR/AHA/ASE/SCCT/SCMR/SNM 2009 Appropriate Use Criteria for Cardiac Radionuclide Imaging.' *JAC*, 53(23), Jul 09, pp. 2201-2229.

Henriksson, L., Sundin, A., Smedby, O. and Albrektsson, P. (1990) 'Assessment of congestive heart failure in chest radiographs. Observer performance with two common film-screen systems.' *Acta radiologica (Stockholm, Sweden : 1987)*, 31(5), Sep, pp. 469-471.

Henry, J. and Pearce, J. (1956) 'THE POSSIBLE ROLE OF CARDIAC ATRIAL STRETCH RECEPTORS IN THE INDUCTION OF CHANGES IN URINE FLOW.' *J. Physiol.*, 131 pp. 572-585.

Herman, P., Khan, A., Kallman, C., Rojas, K., Carmody, D. and Bodenheimer, M. (1990) 'Limited correlation of left ventricular end-diastolic pressure with radiographic assessment of pulmonary hemodynamics.' *Radiology*, 174(3) pp. 721-724.

Herrmann, M., Stanger, O., Paulweber, B., Hufnagl, C. and Herrmann, W. (2007) 'Effect of folate supplementation on N-terminal pro-brain natriuretic peptide.' *Int J Cardiol*, 118(2), May 31, pp. 267-269.

Higgins, J. P. T. (2003) 'Measuring inconsistency in meta-analyses.' *BMJ*, 327(7414), Sep 06, pp. 557-560.

Higgins, J. P. T. and Thompson, S. G. (2002) 'Quantifying heterogeneity in a meta-analysis.' *Statistics in Medicine*, 21(11), Jul 15, pp. 1539-1558.

Ho, K. K., Pinsky, J. L., Kannel, W. B. and Levy, D. (1993) 'The epidemiology of heart failure: the Framingham Study.' *JAC*, 22(4 Suppl A), Oct, pp. 6A-13A.

Hobbs, F. D. R., Davis, R. C., Roalfe, A. K., Hare, R., Davies, M. K. and Kenkre, J. E. (2002) 'Reliability of N-terminal pro-brain natriuretic peptide assay in diagnosis of heart failure: cohort study in representative and high risk community populations.' *BMJ*, 324(7352), Jul 22, p. 1498.

Høilund-Carlsen, P., Gadsbøll, N., Hein, E., Stage, P., Badsberg, J. and Jensen, B. (2005) 'Assessment of left ventricular systolic function by the chest x-ray: comparison with radionuclide ventriculography.' *Journal of Cardiac Failure*, 11(4) pp. 299-305.

Hosmer, D., Taber, S. and Lemeshow, S. (1991) 'The Importance of Assessing the Fit of Logistic Regression Models: A Case Study.' *American Journal of Public Health*, 81(2) pp. 1630-1635.

Houghton, A., Sparrow, N., Toms, E. and Cowley, A. (1997) 'Should general practitioners use the electrocardiogram to select patients with suspected heart failure for echocardiography?' *International Journal of Cardiology*, 62(1), Oct 31, pp. 31-36.

Centre, H. S. C. I. (2011) *Accident and Emergency Attendances in England - 2009-2010*. Cmnd. UK Government (HSCIC Report).

Huang, C., Tsai, M., Hsieh, C., Wang, T., Chang, W. and Chen, W. (2006) 'Diagnostic accuracy of tissue Doppler echocardiography for patients with acute heart failure.' *Heart*, 92(12), Dec 01, pp. 1790-1794.

Hui, S. and Zhou, X. (1998) 'Evaluation of diagnostic tests without gold standards.' *Statistical Methods in Medical Research*, 7, Jul 02, pp. 354-370.

Hundley, W. G., Hamilton, C. A., Thomas, M. S., Herrington, D. M., Salido, T. B., Kitzman, D. W., Little, W. C. and Link, K. M. (1999) 'Utility of Fast Cine Magnetic Resonance Imaging and Display for the Detection of Myocardial Ischemia in Patients Not Well Suited for Second Harmonic Stress Echocardiography.' *Circulation*, 100(16) pp. 1697-1702.

Hunt, P. J., Espiner, E. A., Richards, A. M., Yandle, T. G., Frampton, C. and Nicholls, M. G. (1995) 'Interactions of atrial and brain natriuretic peptides at pathophysiological levels in normal men.' *Am J Physiol*, 269(6 Pt 2), Dec, pp. R1397-1403.

Ioannidis, J. P. A., Patsopoulos, N. A. and Rothstein, H. R. (2008) 'Reasons or excuses for avoiding meta-analysis in forest plots.' *BMJ*, 336(7658), Jul 21, pp. 1413-1415.

Irwig, L., Macaskill, P., Glasziou, P. and Fahey, M. (2002) 'Meta-analytic methods for diagnostic test accuracy.' *Journal of clinical epidemiology*, 48(1), Feb 26, pp. 119-130; discussion 131-112.

Irwig, L., Macaskill, P., Walter, S. and Houssami, N. (2006) 'New methods give better estimates of changes in diagnostic accuracy when prior information ....' *Journal of clinical epidemiology*, Feb 01,

Irwin, R. S. and Madison, J. M. (2000) 'The diagnosis and treatment of cough.' *The New England Journal of Medicine*, 343(23), Dec 07, pp. 1715-1721.

Jacobs, I. and Bast, R. C. (1989) 'The CA 125 tumour-associated antigen: a review of the literature.' *Human Reproduction*, 4(1), January 1, 1989, pp. 1-12.

Jambrik, Z. (2004) 'Usefulness of ultrasound lung comets as a nonradiologic sign of extravascular lung water.' *The American Journal of Cardiology*, 93(10), Jun 15, pp. 1265-1270.

Jang, T., Aubin, C., Naunheim, R., Lewis, L. M. and Kaji, A. H. (2010) 'Jugular venous distension on ultrasound: sensitivity and specificity for heart failure in patients with dyspnea.' *The American Journal of Emergency Medicine*, Oct 15,

Jang, T., Aubin, C., Naunheim, R., Lewis, L. M. and Kaji, A. H. (2012a) 'Sonographic assessment of jugular venous distension and B-type

natriuretic peptide levels in patients with dyspnoea.' *Emerg Med J*, 29(6), Jun, pp. 477-481.

Jang, T. B., Aubin, C., Naunheim, R., Lewis, L. M. and Kaji, A. H. (2012b) 'The predictive value of physical examination findings in patients with suspected acute heart failure syndrome.' *Intern Emerg Med*, 7(3), Jun, pp. 271-274.

Januzzi, J., van Kimmenade, R., Lainchbury, J., Bayes-Genis, A., Ordonez-Llanos, J., Santalo-Bel, M., Pinto, Y. and Richards, M. (2006) 'NT-proBNP testing for diagnosis and short-term prognosis in acute destabilized heart failure: an international pooled analysis of 1256 patients: the International Collaborative of NT-proBNP Study.' *European Heart Journal*, 27(3), Feb 11, pp. 330-337.

Januzzi, J., Camargo, C., Anwaruddin, S., Baggish, A., Chen, A., Krauser, D., Tung, R., Cameron, R., Nagurney, J. and Chae, C. (2005) 'The N-terminal Pro-BNP investigation of dyspnea in the emergency department (PRIDE) study.' *The American Journal of Cardiology*, 95(8) pp. 948-954.

Javadzadeh, J., Collen, A.-C., Manhem, K. and Swedberg, K. (2008) '[Safe diagnosis in acute phase with T-proBNP. Heart failure and acute coronary syndrome can be identified very early].' *Lakartidningen*, 105(9), Mar 27, pp. 616-619.

Jessup, M., Abraham, W. T., Casey, D. E., Feldman, A. M., Francis, G. S., Ganiats, T. G., Konstam, M. A., Mancini, D. M., Rahko, P. S., Silver, M. A., Stevenson, L. W. and Yancy, C. W. (2009) '2009 Focused Update: ACCF/AHA Guidelines for the Diagnosis and Management of Heart Failure in Adults: A Report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines: Developed in Collaboration With the International Society for Heart and Lung Transplantation.' *Circulation*, 119(14), May 13, pp. 1977-2016.

Jones, A. E. and Kline, J. A. (2003) 'Elevated brain natriuretic peptide in septic patients without heart failure.' *Annals of Emergency Medicine*, 42(5), Nov, pp. 714-715.

Jones, C. M. and Athansiou, T. (2009) 'Diagnostic accuracy meta-analysis: review of an important tool in radiological research and decision making.' *British Journal of Radiology*, 82(978), Jun 18, pp. 441-446.

Jortani, S. A., Prabhu, S. D. and Valdes, R. (2004) 'Strategies for developing biomarkers of heart failure.' *Clinical Chemistry*, 50(2), Mar, pp. 265-278.

Jose, J. V., Gupta, S. N. and Selvakumar, D. (2003) 'Utility of N-terminal pro-brain natriuretic peptide for the diagnosis of heart failure.' *Indian heart journal*, 55(1), Feb 01, pp. 35-39.

Joshua, A. M., Celermajer, D. S. and Stockler, M. R. (2005) 'Beauty is in the eye of the examiner: reaching agreement about physical signs and their value.' *Internal Medicine Journal*, 35(3), Apr, pp. 178-187.

Jourdain, P., Funck, F., Canault, E., Bellorini, M., Deschamps, P., Duval, G., Duboc, D. and Desnos, M. (2002) 'Value of brain peptide in the management of suspected cardiac failure in the emergency room. [French].' *Archives des Maladies du Coeur et des Vaisseaux*, 95(9) pp. 763-767.

Kahneman, D. (2012) *Thinking, Fast and Slow*. Penguin Books Ltd.

Kang, D., Kim, M., Kang, S., Kim, Y., Song, J., Song, J. and Kim, J. (2006) 'Role of B-type natriuretic peptide in diagnosis and follow-up of diastolic heart failure. [Korean].' *Korean Circulation Journal*, 36(5), May 1, pp. 359-365.

Kannel, W. B. (1999) 'Current status of the epidemiology of heart failure.' *Current cardiology reports*, 1(1), Jun, pp. 11-19.

Kardaun, J. and Kardaun, O. (1990) 'Comparative diagnostic performance of three radiological procedures for the detection of lumbar disk herniation.' *Methods of Information in Medicine*, 29, Jun 22, pp. 12-22.

Kass, G. (1980) 'An Exploratory Technique for Investigating Large Quantities of Categorical Data.' *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 29(2) pp. 119-127.

Katz, M. H. (2003) 'Multivariable analysis: a primer for readers of medical research.' *Annals of internal medicine*, 138(8), May 15, pp. 644-650.

Khunti, K., Baker, R. and Grimshaw, G. (2000) 'Diagnosis of patients with chronic heart failure in primary care: usefulness of history, examination, and investigations.' *The British Journal of General Practice*, 50(450), Feb 01, pp. 50-54.

Khunti, K., Squire, I., Abrams, K. R. and Sutton, A. J. (2004) 'Accuracy of a 12-lead electrocardiogram in screening patients with suspected heart failure for open access echocardiography: a systematic review and meta-analysis.' *European journal of heart failure : journal of the Working Group on Heart Failure of the European Society of Cardiology*, 6(5), Aug, pp. 571-576.

Killip, T. and Kimball, J. T. (1967) 'Treatment of myocardial infarction in a coronary care unit: a two-year experience with 250 patients. .' *Am J Cardiol*, 20, Jun 18, pp. 457-464.

Kim, J. (2006) 'Prognostic Value of a Novel Classification Scheme for Heart Failure: The Minnesota Heart Failure Criteria.' *American Journal of Epidemiology*, 164(2), Jul 02, pp. 184-193.

Kirkpatrick, J. N., Vannan, M. A., Narula, J. and Lang, R. M. (2007) 'Echocardiography in heart failure: applications, utility, and new horizons.' *J Am Coll Cardiol*, 50(5), Jul 31, pp. 381-396.

Kjaer, A. and Hesse, B. (2001) 'Heart failure and neuroendocrine activation: diagnostic, prognostic and therapeutic perspectives.' *Clinical physiology (Oxford, England)*, 21(6), Nov, pp. 661-672.

Klemen, P., Golub, M. and Grmec, S. (2009) 'Combination of quantitative capnometry, N-terminal pro-brain natriuretic peptide, and clinical assessment in differentiating acute heart failure from pulmonary disease as cause of acute dyspnea in pre-hospital emergency setting: study of diagnostic accuracy.' *Croatian Medical Journal*, 50(2), May 01, pp. 133-142.

Kline, J. A., Courtney, D. M., Than, M. P., Hogg, K., Miller, C. D., Johnson, C. L. and Smithline, H. A. (2010) 'Accuracy of very low pretest probability estimates for pulmonary embolism using the method of attribute matching compared with the Wells score.' *Academic emergency medicine : official journal of the Society for Academic Emergency Medicine*, 17(2), Mar, pp. 133-141.

Klocke, F. J., Baird, M. G., Lorell, B. H., Bateman, T. M., Messer, J. V., Berman, D. S., O'Gara, P. T., Carabello, B. A., Russell, R. O., Jr., Cerqueira, M. D., St John Sutton, M. G., DeMaria, A. N., Udelson, J. E., Kennedy, J. W., Verani, M. S., Williams, K. A., Antman, E. M., Smith, S. C., Jr., Alpert, J. S., Gregoratos, G., Anderson, J. L., Hiratzka, L. F., Faxon, D. P., Hunt, S. A., Fuster, V., Jacobs, A. K., Gibbons, R. J., Russell, R. O., American College of, C., American Heart Association Task Force on Practice, G. and American Society for Nuclear, C. (2003) 'ACC/AHA/ASNC guidelines for the clinical use of cardiac radionuclide imaging--executive summary: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (ACC/AHA/ASNC Committee to Revise the 1995 Guidelines for the Clinical Use of Cardiac Radionuclide Imaging).' *Circulation*, 108(11), Sep 16, pp. 1404-1418.

Klok, F. A., Mos, I. C. M. and Huisman, M. V. (2008) 'Brain-Type Natriuretic Peptide Levels in the Prediction of Adverse Outcome in Patients with Pulmonary Embolism: A Systematic Review and Meta-analysis.' *American journal of respiratory and critical care medicine*, 178(4), Jun 29, pp. 425-430.

Klokkers, J., Langehanenberg, P., Kemper, B., Kosmeier, S., von Bally, G., Riethmüller, C., Wunder, F., Sindic, A., Pavenstädt, H. and Schlatter, E. (2009) 'Atrial natriuretic peptide and nitric oxide signaling antagonizes vasopressin-mediated water permeability in inner medullary collecting duct cells.' *American Journal of Physiology- Renal Physiology*, 297(3) pp. F693-F703.

Knudsen, C., Omland, T., Clapton, P. and Henriksson, P. (2004) 'Chest radiographs and BNP levels provided complementary information beyond clinical findings for diagnosing heart failure.' *Evidence Based Medicine*, 9(5), Sep 01, p. 152.

Knudsen, C., Riis, J., Finsen, A., Eikvar, L., Muller, C., Westheim, A. and Omland, T. (2004) 'Diagnostic value of a rapid test for B-type natriuretic peptide in patients presenting with acute dyspnoea: effect of age and gender.' *European Journal of Heart Failure*, 6(1), Feb 01, pp. 55-62.

Knudsen, C. W., Omland, T., Clapton, P., Westheim, A., Abraham, W. T., Storrow, A. B., McCord, J., Nowak, R. M., Aumont, M.-C., Duc, P., Hollander, J. E., Wu, A. H. B., McCullough, P. A. and Maisel, A. S. (2004) 'Diagnostic value of B-Type natriuretic peptide and chest radiographic findings in patients with acute dyspnea.' *The American Journal of Medicine*, 116(6), Apr 15, pp. 363-368.

Kossman, C. (1964) *Nomenclature and Criteria for Disease*. Little Brown and Company.

Kruger, S., Filzmaier, K., Graf, J., Kunz, D., Stickel, T., Hoffmann, R., Hanrath, P. and Janssens, U. (2004) 'QRS prolongation on surface ECG and brain natriuretic peptide as indicators of left ventricular systolic dysfunction.' *Journal of Internal Medicine*, 255(2), Mar 01, pp. 206-212.

Krum, H., Sackner-Bernstein, J. D., Goldsmith, R. L., Kukin, M. L., Schwartz, B., Penn, J., Medina, N., Yushak, M., Horn, E. and Katz, S. D. (1995) 'Double-blind, placebo-controlled study of the long-term efficacy of carvedilol in patients with severe chronic heart failure.' *Circulation*, 92(6), Sep, pp. 1499-1506.

Kucher, N., Printzen, G. and Goldhaber, S. Z. (2003) 'Prognostic role of brain natriuretic peptide in acute pulmonary embolism.' *Circulation*, 107(20), Jun 27, pp. 2545-2547.

Lainchbury, J., Campbell, E., Frampton, C., Yandle, T., Nicholls, M. and Richards, A. (2003) 'Brain natriuretic peptide and n-terminal brain natriuretic peptide in the diagnosis of heart failure in patients with acute shortness of breath.' *Journal of the American College of Cardiology*, 42(4) pp. 728-735.

Lakhdhar, R., Hamouda, C., Ammar, L., Majed, K., Moncel, F., Kaabachi, N., Drissa, H. and Falfoul, N. (2013) 'Est-ce que le dosage du brain natriuretic peptide (BNP) est utile dans le diagnostic de la dyspnée aiguë aux urgences.' *La Tunisie Medicale*, 91 pp. 527-533.

Lau, J., Ioannidis, J. P. and Schmid, C. H. (1998) 'Summing up evidence: one answer is not always enough.' *Lancet*, 351(9096), Feb 10, pp. 123-127.

Laupacis, A., Sekar, N. and Stiell, I. (1997) 'Clinical prediction rules. A review and suggested modifications of methodological standards.' *JAMA: The Journal of the American Medical Association*, 277(6) pp. 488-494.

Lavdaniti, M. (2008) 'Invasive and non-invasive methods for cardiac output measurement.' *International Journal for Caring Sciences*,

Lee, S.-C., Stevens, T. L., Sandberg, S. M., Heublein, D. M., Nelson, S. M., Jougasaki, M., Redfield, M. M. and Burnett, J. C. (2002) 'The potential

of brain natriuretic peptide as a biomarker for New York Heart Association class during the outpatient treatment of heart failure.' *Journal of Cardiac Failure*, 8(3), Jul, pp. 149-154.

Leeflang, M. M. G., Moons, K. G. M., Reitsma, J. B. and Zwinderman, A. H. (2008) 'Bias in Sensitivity and Specificity Caused by Data-Driven Selection of Optimal Cutoff Values: Mechanisms, Magnitude, and Solutions.' *Clinical Chemistry*, 54(4), May 01, p. 729.

Leeflang, M. M. G., Deeks, J. J., Gatsonis, C., Bossuyt, P. M. M. and Group, C. D. T. A. W. (2008) 'Systematic reviews of diagnostic test accuracy.' *Annals of internal medicine*, 149(12), Dec 16, pp. 889-897.

Leemis, L. and Trivedi, K. (1996) 'A Comparison of Approximate Interval Estimations for the Bernoulli Parameter.' *Am Stat*, 50(63-68)

Lehr, R. and Pong, A. (2010). ROC Curve. In Chow, S. (Ed.), *Encyclopedia of Biopharmaceutical Statistics* (Third ed., pp. 1185-1191): CRC Press.

Lemeshow, S. and Hosmer, D. (1982) 'A REVIEW OF GOODNESS OF FIT STATISTICS FOR USE IN THE DEVELOPMENT OF LOGISTIC REGRESSION MODELS.' *American Journal of Epidemiology*, 115(1) pp. 92-106.

Leuppi, J., Dieterle, T., Koch, G., Martina, B., Tamm, M., Perruchoud, A., Wildeisen, I. and Leimenstoll, B. (2005) 'Diagnostic value of lung auscultation in an emergency room setting.' *Swiss Medical Weekly*, 135(35-36), Sep 03, pp. 520-524.

Levin, E. R., Gardner, D. G. and Samson, W. K. (1998) 'Natriuretic peptides.' *New England Journal of Medicine*, 339(5), Jul 30, pp. 321-328.

Levy, D., Kenchaiah, S., Larson, M. G., Benjamin, E. J., Kupka, M. J., Ho, K. K. L., Murabito, J. M. and Vasan, R. S. (2002) 'Long-term trends in the incidence of and survival with heart failure.' *The New England Journal of Medicine*, 347(18), Oct 31, pp. 1397-1402.

Lewis, P. and Stevens, J. (1991) 'Nonlinear Modeling of Time Series Using Multivariate Adaptive Regression Splines (MARS).' *Journal of the American Statistical Association*, 86(416) pp. 864-877.

Lip, G. Y. H., Beevers, D. G., Singh, S. P. and Watson, R. D. S. (1995) 'ABC of Atrial Fibrillation: AETIOLOGY, PATHOPHYSIOLOGY, AND CLINICAL FEATURES.' *BMJ*, 311(7017), November 25, 1995, pp. 1425-1428.

Liteplo, A., Marill, K., Villen, T., Miller, R., Murray, A., Croft, P., Capp, R. and Noble, V. (2009) 'Emergency thoracic ultrasound in the differentiation of the etiology of shortness of breath (ETUDES): sonographic B-lines and N-terminal pro-brain-type natriuretic peptide in diagnosing congestive heart failure.' *Academic Emergency Medicine*, 16(3), Apr 01, pp. 201-210.

Lloyd-Jones, D. M. (2002) 'Lifetime Risk for Developing Congestive Heart Failure: The Framingham Heart Study.' *Circulation*, 106(24), Dec 10, pp. 3068-3072.

Lo, H., Liao, S., Ng, C., Kuan, J., Chen, J. and Chiu, T. (2007) 'Utility of impedance cardiography for dyspneic patients in the ED.' *American Journal of Emergency Medicine*, 25(4), Jun 01, pp. 437-441.

Logeart, D., Saudubray, C., Beyne, P., Thabut, G., Ennezat, P.-V., Chavelas, C., Zanker, C., Bouvier, E. and Solal, A. C. (2002) 'Comparative value of Doppler echocardiography and B-type natriuretic peptide assay in the etiologic diagnosis of acute dyspnea.' *Journal of the American College of Cardiology*, 40(10), Nov 20, pp. 1794-1800.

Lok, C., Morgan, C. and Ranganathan, N. (1998) 'The accuracy and interobserver agreement in detecting the 'gallop sounds' by cardiac auscultation.' *Chest*, Feb 01,

Lokuge, A., Lam, L., Cameron, P., Krum, H., de Villiers, S., Bystrycki, A., Naughton, M. T., Eccleston, D., Flannery, G., Federman, J. and Schneider, H.-G. (2010) 'B-type natriuretic peptide testing and the accuracy of heart failure diagnosis in the emergency department.' *Circulation. Heart failure*, 3(1), Feb, pp. 104-110.

Macaskill, P. (2010) 'Empirical Bayes estimates generated in a hierarchical summary ROC analysis agreed closely with those of a full Bayesian analysis.' *Journal of clinical epidemiology*, 57(9), Jul 18, pp. 925-932.

Madhok, V., Falk, G., Rogers, A. and Struthers, A. D. (2008) 'The accuracy of symptoms, signs and diagnostic tests in the diagnosis of left

ventricular dysfunction in primary care: a diagnostic accuracy systematic review.' *BMC family ...*,

Madsen, E. B., Gilpin, E., Slutsky, R. A., Ahnve, S., Henning, H. and Ross, J. (1984) 'Usefulness of the chest x-ray for predicting abnormal left ventricular function after acute myocardial infarction.' *American Heart Journal*, 108(6), Dec 01, pp. 1431-1436.

Maeda, K., Tsutamoto, T., Wada, A., Hisanaga, T. and Kinoshita, M. (1998) 'Plasma brain natriuretic peptide as a biochemical marker of high left ventricular end-diastolic pressure in patients with symptomatic left ventricular dysfunction.' *American Heart Journal*, 135(5 Pt 1), Jun 01, pp. 825-832.

Magga J, M. (1994) 'Brain natriuretic peptide in plasma, atria and ventricles of vasopressin- and phenylephrine-infused conscious rats.' *Endocrinology*, 134 pp. 2505-2515.

Maisel, A., Doyle, J. and Schwam, E. (2005) 'B-type natriuretic peptide in the emergency department: A valuable diagnostic aid [1] (multiple letters).' *Academic Emergency Medicine*, 12(6), Jul 01, pp. 572-574.

Maisel, A., Mueller, C., Nowak, R. and Peacock, W. (2010) 'Mid-Region Pro-Hormone Markers for Diagnosis and Prognosis in Acute Dyspnea:: Results From the BACH (Biomarkers in Acute Heart Failure) Trial.' *Journal of the American ...*,

Primary results of the Rapid Emergency Department Heart Failure Outpatient Trial (REDHOT). A multicenter study of B-type natriuretic peptide levels, emergency department decision making, and outcomes in patients presenting with shortness of breath., 44 Cong. Rec. 1328-1333 (2004).

Maisel, A. S., Krishnaswamy, P., Nowak, R. M., McCord, J., Hollander, J. E., Duc, P., Omland, T., Storrow, A. B., Abraham, W. T., Wu, A. H., Clopton, P., Steg, P. G., Westheim, A., Knudsen, C. W., Perez, A., Kazanegra, R., Herrmann, H. C., McCullough, P. A. and Investigators, B. N. P. M. S. (2002) 'Rapid measurement of B-type natriuretic peptide in the emergency diagnosis of heart failure.' *The New England Journal of Medicine*, 347(3), Jul 18, pp. 161-167.

Malas, O., Caglayan, B., Fidan, A., Ocal, Z., Ozdogan, S. and Torun, E. (2003) 'Cardiac or pulmonary dyspnea in patients admitted to the

emergency department.' *Respiratory Medicine*, 97(12), Dec 01, pp. 1277-1281.

Mandavia, D. P., Hoffner, R. J., Mahaney, K. and Henderson, S. O. (2001) 'Bedside echocardiography by emergency physicians.' *Annals of Emergency Medicine*, 38(4), Oct, pp. 377-382.

Mann, D. L. (1999) 'Mechanisms and models in heart failure: A combinatorial approach.' *Circulation*, 100(9), Aug 31, pp. 999-1008.

Mant, J., Doust, J., Roalfe, A., Barton, P., Cowie, M., Glasziou, P., Mant, D., McManus, R., Holder, R. and Deeks, J. (2010) 'Systematic review and individual patient data meta-analysis of diagnosis of heart failure, with modelling of implications of different diagnostic strategies in primary care.' *Health Technol Assess*, 13, Jun 26, pp. 1-207.

Marantz, P., Kaplan, M. and Alderman, M. (1990) 'Clinical diagnosis of congestive heart failure in patients with acute dyspnea.' *Chest*, 97(4) pp. 776-781.

Mariano-Goulart, D., Eberlé, M. C., Boudousq, V., Hejazi-Moughari, A., Piot, C., de Kerleau, C. C., Verdier, R., Barge, M. L., Comte, F. and Bressot, N. (2003) 'Major increase in brain natriuretic peptide indicates right ventricular systolic dysfunction in patients with heart failure.' *European journal of heart failure : journal of the Working Group on Heart Failure of the European Society of Cardiology*, 5(4) pp. 481-488.

Marx, K. (1859) *A Contribution to the Critique of Political Economy*. Moscow: Progress Publishers.

Mathews, L. and Singh, K. R. (2008) 'Cardiac output monitoring.' *Annals of cardiac anaesthesia*, 11(1) p. 56.

McCord, J., Nowak, R., Jacobsen, G., Sallach, J., Wu, A., Perez, A., Omland, T., Knudsen, C., Westheim, A., Duc, P., Steg, P., Hollander, J., Herrmann, H., Storrow, A., Abraham, W., Lamba, S., McCullough, P. and Maisel, A. (2005) 'B-type natriuretic peptide levels in patients in the emergency department with possible heart failure and previous stable angina pectoris and/or healed myocardial infarction.' *American Journal of Cardiology*, 96(10) pp. 1370-1373.

McCullough, P., Hollander, J., Nowak, R., Storrow, A., Duc, P., Omland, T., McCord, J., Herrmann, H., Steg, P., Westheim, A., Knudsen, C., Abraham, W., Lamba, S., Wu, A., Perez, A., Clopton, P., Krishnaswamy, P., Kazanegra, R., Maisel, A. and Investigators, B. (2003) 'Uncovering heart failure in patients with a history of pulmonary disease: rationale for the early use of B-type natriuretic peptide in the emergency department.[see comment].' *Academic Emergency Medicine*, 10(3), Apr 01, pp. 198-204.

McCullough, P. A., Nowak, R. M., McCord, J., Hollander, J. E., Herrmann, H. C., Steg, P. G., Duc, P., Westheim, A., Omland, T., Knudsen, C. W., Storrow, A. B., Abraham, W. T., Lamba, S., Wu, A. H. B., Perez, A., Clopton, P., Krishnaswamy, P., Kazanegra, R. and Maisel, A. S. (2002) 'B-type natriuretic peptide and clinical judgment in emergency diagnosis of heart failure: analysis from Breathing Not Properly (BNP) Multinational Study.' *Circulation*, 106(4), Jul 23, pp. 416-422.

McGinn, T. G., Guyatt, G. H., Wyer, P. C., Naylor, C. D., Stiell, I. G. and Richardson, W. S. (2000) 'Users' guides to the medical literature: XXII: how to use articles about clinical decision rules. Evidence-Based Medicine Working Group.' *JAMA*, 284(1), Jul 5, pp. 79-84.

McKee, P. A., Castelli, W. P., McNamara, P. M. and Kannel, W. B. (1971) 'The Natural History of Congestive Heart Failure: The Framingham Study.' *The New England Journal of Medicine*, 285(26), Dec 23, pp. 1441-1446.

McMurray, J. J. and Stewart, S. (2000) 'Epidemiology, aetiology, and prognosis of heart failure.' *Heart*,

McMurray, J. J. V., Adamopoulos, S., Anker, S. D., Auricchio, A., Böhm, M., Dickstein, K., Falk, V., Filippatos, G., Fonseca, C., Gomez-Sanchez, M. A., Jaarsma, T., Køber, L., Lip, G. Y. H., Maggioni, A. P., Parkhomenko, A., Pieske, B. M., Popescu, B. A., Ronnevik, P. K., Rutten, F. H., Schwitter, J., Seferovic, P., Stepinska, J., Trindade, P. T., Voors, A. A., Zannad, F., Zeiher, A., (CPG), E. C. f. P. G., Bax, J. J., Baumgartner, H., Ceconi, C., Dean, V., Deaton, C., Fagard, R., Funck-Brentano, C., Hasdai, D., Hoes, A., Kirchhof, P., Knuuti, J., Kolh, P., McDonagh, T., Moulin, C., Reiner, Z., Sechtem, U., Sirnes, P. A., Tendera, M., Torbicki, A., Vahanian, A., Windecker, S., Reviewers, D., Bonet, L. A., Avraamides, P., Ben Lamin, H. A., Brignole, M., Coca, A., Cowburn, P., Dargie, H., Elliott, P., Flachskampf, F. A., Guida, G. F., Hardman, S., lung, B., Merkely, B., Mueller, C., Nanas, J. N., Nielsen, O. W., Orn, S., Parissis, J. T. and Ponikowski, P. (2012) 'ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure 2012: The Task Force for the Diagnosis and Treatment of Acute and Chronic Heart

Failure 2012 of the European Society of Cardiology. Developed in collaboration with the Heart Failure Association (HFA) of the ESC.' *European Heart Journal*, 33(14), Jul 14, pp. 1787-1847.

McNamara, R. and Cionni, D. (1992) 'Utility of the peak expiratory flow rate in the differentiation of acute dyspnea; Cardiac vs pulmonary origin.' *Chest*, 101(1) pp. 129-132.

Mendez Bailon, M., Munoz Rivas, N., Romero Roman, C., Perez De Oteyza, C., Conthe Gutierrez, P. and Audibert Mena, L. (2007) 'Measurement of NT-proBNP in elderly patients with acute dyspnea: Diagnostic and pronostic value. [Spanish].' *Medicina Clinica*, 128(12) pp. 453-455.

Mikkelsen, K., Bie, P., Moller, J., Ryde, H. and Videbaek, L. (2007) 'Diagnostic accuracy of plasma brain natri- uretic peptide and aminoterminal-proBNP in mild heart failure depends on assay and introduction of therapy.' *Scand J Clin Lab Invest* 65, Jun 11, pp. 633-647.

Miller, J. B., Sen, A., Strote, S. R., Hegg, A. J., Farris, S., Brackney, A., Amponsah, D. and Mossallam, U. (2012) 'Inferior vena cava assessment in the bedside diagnosis of acute heart failure.' *Am J Emerg Med*, 30(5), Jun, pp. 778-783.

Moe, G. W., Howlett, J., Januzzi, J. L. and Zowall, H. (2007) 'N-Terminal Pro-B-Type Natriuretic Peptide Testing Improves the Management of Patients With Suspected Acute Heart Failure.' *Circulation*, 115(24) pp. 3103-3110.

Mohler, E. R., Ryan, T., Segar, D. S., Sawada, S. G., Sonel, A. F., Perkins, L., Fineberg, N., Feigenbaum, H. and Wilensky, R. L. (1998) 'Clinical utility of troponin T levels and echocardiography in the emergency department.' *American Heart Journal*, 135(2 Pt 1), Mar, pp. 253-260.

Moon, K., Harrell, F. and Ewout, S. (2002) 'Should scoring rules be based on odds ratios or regression coefficients?' *Journal of clinical epidemiology*, pp. 1054-1055.

Moons, K. G. and Grobbee, D. E. (2002) 'When should we remain blind and when should our eyes remain open in diagnostic studies?' *Journal of Clinical Epidemiology*, 55 pp. 633-636.

Morello, A., Lloyd-Jones, D., Chae, C., van Kimmenade, R., Chen, A., Baggish, A., O'Donoghue, M., Lee-Lewandrowski, E. and Januzzi Jr, J. (2007) 'Association of atrial fibrillation and amino-terminal pro-brain natriuretic peptide concentrations in dyspneic subjects with and without acute heart failure: Results from the ProBNP Investigation of Dyspnea in the Emergency Department (PRIDE) study.' *American Heart Journal*, 153(1), Feb 01, pp. 90-97.

Morrison, L. K., Harrison, A., Krishnaswamy, P., Kazanegra, R., Clopton, P. and Maisel, A. (2002) 'Utility of a rapid B-natriuretic peptide assay in differentiating congestive heart failure from lung disease in patients presenting with dyspnea.' *Journal of the American College of Cardiology*, 39(2), Feb 16, pp. 202-209.

Moses, L. E., Shapiro, D. and Littenberg, B. (1993) 'Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations.' *Stat Med*, 12(14), Jul 30, pp. 1293-1316.

Moynihan, R., Doust, J. and Henry, D. (2012) 'Preventing overdiagnosis: how to stop harming the healthy.' *BMJ*, 344 p. e3502.

Možina, M., Demšar, J., Kattan, M. and Zupan, B. (2004) 'Nomograms for visualization of naive Bayesian classifier.'

Mueller, C. and Laule-Kilian, K. (2004) 'BNP improves outcomes in evaluation of dyspnoea.' *BMJ*, 328, Jun 29, pp. 1-1.

Mueller, C., Frana, B., Rodriguez, D., Laule-Kilian, K. and Perruchoud, A. (2005) 'Emergency diagnosis of congestive heart failure: impact of signs and symptoms.' *Canadian Journal of Cardiology*, 21(11), Sep 01, pp. 921-924.

Mueller, C., Scholer, A., Laule-Kilian, K., Martina, B., Schindler, C., Buser, P., Pfisterer, M. and Perruchoud, A. P. (2004) 'Use of B-type natriuretic peptide in the evaluation and management of acute dyspnea.' *The New England Journal of Medicine*, 350(7), Mar 12, pp. 647-654.

Mueller, C., Laule-Kilian, K., Schindler, C., Klima, T., Frana, B., Rodriguez, D., Scholer, A., Christ, M. and Perruchoud, A. P. (2006) 'Cost-effectiveness of B-type natriuretic peptide testing in patients with acute dyspnea.' *Archives of Internal Medicine*, 166(10), Jun 22, pp. 1081-1087.

Mueller, T., Gegenhuber, A., Poelz, W. and Haltmayer, M. (2005) 'Diagnostic accuracy of B type natriuretic peptide and amino terminal proBNP in the emergency diagnosis of heart failure.' *Heart*, 91(5), Jun 01, pp. 606-612.

Mueller-Lenke, N., Rudez, J., Staub, D., Laule-Kilian, K., Klima, T., Perruchoud, A. and Mueller, C. (2006) 'Use of chest radiography in the emergency diagnosis of acute congestive heart failure.' *Heart*, 92(5) pp. 695-696.

Murkofsky, R. L., Dangas, G., Diamond, J. A., Mehta, D., Schaffer, A. and Ambrose, J. A. (2005) 'A prolonged QRS duration on surface electrocardiogram is a specific indicator of left ventricular dysfunction [see comment].' *Journal of the American College of Cardiology*, 32(2), Nov 29, pp. 476-482.

Murray, S. A., Boyd, K., Kendall, M., Worth, A., Benton, T. F. and Clausen, H. (2002) 'Dying of lung cancer or cardiac failure: prospective qualitative interview study of patients and their carers in the community.' *BMJ*, 325(7370), Oct 26, pp. 929-929.

Nazerian, P., Vanni, S., Zanobetti, M., Polidori, G., Pepe, G., Federico, R., Cangioli, E. and Grifoni, S. (2010) 'Diagnostic Accuracy of Emergency Doppler Echocardiography for Identification of Acute Left Ventricular Heart Failure in Patients with Acute Dyspnea: Comparison with Boston Criteria and N-terminal Prohormone Brain Natriuretic Peptide.' *Academic Emergency Medicine*, 17(1), Jul 30, pp. 18-26.

Neyman, J. and Pearson, E. S. (1933) 'On the Problem of the Most Efficient Tests of Statistical Hypotheses.' *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 231(Chapter 6) pp. 73-108.

Ng, L., Loke, I., Davies, J., Khunti, K., Stone, M., Abrams, K., Chin, D. and Squire, I. (2003) 'Identification of previously undiagnosed left ventricular systolic dysfunction: community screening using natriuretic peptides and electrocardiography.' *European Journal of Heart Failure*, 5(6), Dec 01, pp. 775-782.

NHSIC. (2010). National Heart Failure Audit 2010 (pp. 1-39). Leeds : The NHS Information Centre.

NICE. (2010) 'CHRONIC HEART FAILURE

National clinical guideline for diagnosis and management in primary and secondary care.' [Accessed

Nicol, E., Fittall, B., Roughton, M., Cleland, J., Dargie, H. and Cowie, M. (2008) 'NHS heart failure survey: a survey of acute heart failure admissions in England, Wales and Northern Ireland.' *Heart*, 94 pp. 172-177.

Nieminen, M. S., Brutsaert, D., Dickstein, K., Drexler, H., Follath, F., Harjola, V. P., Hochadel, M., Komajda, M., Lassus, J., Lopez-Sendon, J. L., Ponikowski, P., Tavazzi, L., EuroHeart Survey, I. and Heart Failure Association, E. S. o. C. (2006) 'EuroHeart Failure Survey II (EHFS II): a survey on hospitalized acute heart failure patients: description of population.' *Eur Heart J*, 27(22), Nov, pp. 2725-2736.

O'Connor, C. (2005) *Managing Acute Decompensated Heart Failure*. London and New York: Taylor & Francis.

Effect of Nesiritide in Patients with Acute Decompensated Heart Failure, 365 Cong. Rec. 32-43 (2011).

O'Gara, P. T., American College of Emergency, P., Society for Cardiovascular, A., Interventions, Kushner, F. G., Ascheim, D. D., Casey, D. E., Jr., Chung, M. K., de Lemos, J. A., Ettinger, S. M., Fang, J. C., Fesmire, F. M., Franklin, B. A., Granger, C. B., Krumholz, H. M., Linderbaum, J. A., Morrow, D. A., Newby, L. K., Ornato, J. P., Ou, N., Radford, M. J., Tamis-Holland, J. E., Tommaso, C. L., Tracy, C. M., Woo, Y. J., Zhao, D. X., Anderson, J. L., Jacobs, A. K., Halperin, J. L., Albert, N. M., Brindis, R. G., Creager, M. A., DeMets, D., Guyton, R. A., Hochman, J. S., Kovacs, R. J., Kushner, F. G., Ohman, E. M., Stevenson, W. G. and Yancy, C. W. (2013) '2013 ACCF/AHA guideline for the management of ST-elevation myocardial infarction: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines.' *J Am Coll Cardiol*, 61(4), Jan 29, pp. e78-140.

Ogawa, Y., Itoh, H. and Nakao, K. (1995) 'MOLECULAR BIOLOGY AND BIOCHEMISTRY OF NATRIURETIC PEPTIDE FAMILY.' *Clinical and Experimental Pharmacology and Physiology*, 22 pp. 49-53.

Ogawa, Y., Tamura, N., Chusho, H. and Nakao, K. (2001) 'Brain natriuretic peptide appears to act locally as an antifibrotic factor in the heart.' *Canadian Journal of Physiology and Pharmacology*, 79(8), 2001/08/01, pp. 723-729.

ONS. (2012) *Office of National Statistics*. [Online] [Accessed on 1/7/2012] [http://www.ons.gov.uk/onse/dcp171780\\_242394.pdf](http://www.ons.gov.uk/onse/dcp171780_242394.pdf)

Ozaki, J., Shimizu, H., Hashimoto, Y., Itoh, H., Nakao, K. and Inui, K. (1999) 'Enzymatic inactivation of major circulating forms of atrial and brain natriuretic peptides.' *European journal of pharmacology*, 370(3), May 16, pp. 307-312.

Packer, M. (1995a) 'Evolution of the neurohormonal hypothesis to explain the progression of chronic heart failure.' *European Heart Journal*, 16(suppl F), Jul 02, pp. 4-6.

Evolution of the neurohormonal hypothesis to explain the progression of chronic heart failure, 16 Cong. Rec. 4-6 (1995b).

Parrinello, G., Paterna, S., Di Pasquale, P., Torres, D., Fatta, A., Mezzero, M., Scaglione, R. and Licata, G. (2008) 'The usefulness of bioelectrical impedance analysis in differentiating dyspnea due to decompensated heart failure.' *Journal of Cardiac Failure*, 14(8), Oct 01, pp. 676-686.

Pascual Figal, D., Cerdan Sanchez, M., Noguera Velasco, J., Casas Pina, T., Munoz Gimeno, L., Garcia Rodriguez, R., Ruiperez Abizanda, J., Martinez Hernandez, P. and Valdes Chavarri, M. (2005) '[Usefulness of NTproBNP in the emergency management of patients with severe dyspnea and an uncertain heart failure diagnosis].[see comment].'  
*Revista espanola de cardiologia*, 58(10), Oct 01, pp. 1155-1161.

Pattenden, J., Roberts, H. and Lewin, R. (2007) 'Living with heart failure; patient and carer perspectives.' *European journal of cardiovascular nursing : journal of the Working Group on Cardiovascular Nursing of the European Society of Cardiology*, 6(4), Dec, pp. 273-279.

Paulus, W. (1998) 'How to diagnose diastolic heart failure.' *European Heart Journal*, 19, Jul 19, pp. 990-1003.

Payvar, S., Orlandi, C., Stough, W. G., Elkayam, U., Ouyang, J., Casscells, S. W. and Gheorghide, M. (2006) 'Comparison of 60-Day Mortality in Hospitalized Heart Failure Patients With Versus Without Hypothermia.' *The American Journal of Cardiology*, 98(11), Dec, pp. 1485-1488.

Peacock, W. F., Summers, R. L., Vogel, J. and Emerman, C. E. (2006) 'Impact of impedance cardiography on diagnosis and therapy of emergent dyspnea: the ED-IMPACT trial.' *Academic emergency medicine : official journal of the Society for Academic Emergency Medicine*, 13(4), May 01, pp. 365-371.

Pearsall, J. (Ed.). (1998). *The New Oxford Dictionary of English*: Oxford Univ Press.

Peng, C. and al, e. (2002) 'An introduction to logistic regression analysis and reporting.' *The Journal of Educational ...*,

Phillips, B. (2010) 'Clinical decision rules: how to build them.' *Archives of disease in childhood. Education and practice edition*, 95(3), Jul, pp. 83-87.

Pitt, B., Zannad, F., Remme, W. J., Cody, R., Castaigne, A., Perez, A., Palensky, J. and Wittes, J. (1999) 'The effect of spironolactone on morbidity and mortality in patients with severe heart failure. Randomized Aldactone Evaluation Study Investigators.' *New England Journal of Medicine*, 341(10), Sep 02, pp. 709-717.

Porter, J. G., Arfsten, A., Fuller, F., Miller, J. A., Gregory, L. C. and Lewicki, J. A. (1990) 'Isolation and functional expression of the human atrial natriuretic peptide clearance receptor cDNA.' *Biochemical and Biophysical Research Communications*, 171(2), 9/14/, pp. 796-803.

Potocki, M., Breidthardt, T., Reichlin, T., Hartwiger, S., Morgenthaler, N. G., Bergmann, A., Noveanu, M., Freidank, H., Taegtmeyer, A. B., Wetzel, K., Boldanova, T., Stelzig, C., Bingisser, R., Christ, M. and Mueller, C. (2010) 'Comparison of midregional pro-atrial natriuretic peptide with N-terminal pro-B-type natriuretic peptide in the diagnosis of heart failure.' *J Intern Med*, 267(1), Jan, pp. 119-129.

Price, R. (1763a) 'An Essay towards solving a Problem in the Doctrine of Chances. By the late Rev. Mr. Bayes, communicated by Mr. Price, in a letter to John Canton, M. A. and F. R. S.' *Philosophical Transactions of the Royal Society of London*, 53, Aug 23, pp. 370-418.

An Essay towards solving a Problem in the Doctrine of Chances. By the late Rev. Mr. Bayes, communicated by Mr. Price, in a letter to John Canton, M. A. and F. R. S., 53 Cong. Rec. 370-418 (1763b).

Prosen, G., Klemen, P. and Strnad, M. (2011) 'Combination of lung ultrasound (a comet-tail sign) and N-terminal pro-brain natriuretic peptide in differentiating acute heart failure from chronic obstructive pulmonary ....' *Critical Care*,

Ray, P. and Lefort, Y. (2006) '[Usefulness of B-type natriuretic peptide in emergency medicine].' *Revue de Medecine Interne*, 27(11), Nov 01, pp. 858-864.

Ray, P., Arthaud, M., Lefort, Y., Birolleau, S., Beigelman, C. and Riou, B. (2004) 'Usefulness of B-type natriuretic peptide in elderly patients with acute dyspnea.' *Intensive Care Medicine*, 30(12), Dec 22, pp. 2230-2236.

Ray, P., Arthaud, M., Birolleau, S., Isnard, R., Lefort, Y., Boddaert, J. and Riou, B. (2005) 'Comparison of brain natriuretic peptide and probrain natriuretic peptide in the diagnosis of cardiogenic pulmonary edema in patients aged 65 and older.[see comment].' *Journal of the American Geriatrics Society*, 53(4), May 01, pp. 643-648.

Ray, P., Birolleau, S., Lefort, Y., Becquemin, M.-H., Beigelman, C., Isnard, R., Teixeira, A., Arthaud, M., Riou, B. and Boddaert, J. (2006) 'Acute respiratory failure in the elderly: etiology, emergency diagnosis and prognosis.[see comment].' *Critical Care (London, England)*, 10(3) p. R82.

Reichlin, T., Potocki, M., Breidthardt, T., Noveanu, M., Hartwiger, S., Burri, E., Klima, T., Stelzig, C., Laule, K., Mebazaa, A., Christ, M. and Mueller, C. (2010) 'Diagnostic and prognostic value of uric acid in patients with acute dyspnea.' *The American Journal of Medicine*, 122(11), Jul 30, pp. 1054.e1057-1054.e1014.

Reitsma, J. B., Glas, A. S., Rutjes, A. W. S., Scholten, R. J. P. M., Bossuyt, P. M. and Zwinderman, A. H. (2005) 'Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews.' *Journal of clinical epidemiology*, 58(10), Mar 22, pp. 982-990.

Remes, J., MIETTINEN, H., Reunanen, A. and Pyorala, K. (1991) 'Validity of clinical diagnosis of heart failure in primary health care.' *European Heart Journal*, 12(3) pp. 315-321.

RightCare. (2013) *The NHS Atlas of Variation in Diagnostic Services*. London: November 2013 viewed

Riley, R. D., Abrams, K. R., Sutton, A. J., Lambert, P. C. and Thompson, J. R. (2007) 'Bivariate random-effects meta-analysis and the estimation of between-study correlation.' *BMC Medical Research Methodology*, 7(1) p. 3.

Robaei, D., Koe, L., Bais, R., Gould, I., Stewart, T. and Tofler, G. H. (2011) 'Effect of NT-proBNP testing on diagnostic certainty in patients admitted to the emergency department with possible heart failure.' *Annals of clinical biochemistry*, 48(Pt 3), Jun, pp. 212-217.

Robinson, P. J., Wilson, D., Coral, A. and Murphy, A. (1999) 'Variation between experienced observers in the interpretation of accident and emergency radiographs.' *British journal of ...*,

Roche. (2010) *proBrain Natriuretic Peptide*. [Information Leaflet] 2010 viewed

Roger, V. L., Weston, S. A., Redfield, M. M., Hellermann-Homan, J. P., Killian, J., Yawn, B. P. and Jacobsen, S. J. (2004) 'Trends in heart failure incidence and survival in a community-based population.' *JAMA: The Journal of the American Medical Association*, 292(3), Jul 21, pp. 344-350.

Roger, V. L., Go, A. S., Lloyd-Jones, D. M., Benjamin, E. J., Berry, J. D., Borden, W. B., Bravata, D. M., Dai, S., Ford, E. S., Fox, C. S., Fullerton, H. J., Gillespie, C., Hailpern, S. M., Heit, J. A., Howard, V. J., Kissela, B. M., Kittner, S. J., Lackland, D. T., Lichtman, J. H., Lisabeth, L. D., Makuc, D. M., Marcus, G. M., Marelli, A., Matchar, D. B., Moy, C. S., Mozaffarian, D., Mussolino, M. E., Nichol, G., Paynter, N. P., Soliman, E. Z., Sorlie, P. D., Sotoodehnia, N., Turan, T. N., Virani, S. S., Wong, N. D., Woo, D., Turner, M. B., American Heart Association Statistics, C. and Stroke Statistics, S. (2012) 'Heart disease and stroke statistics--2012 update: a report from the American Heart Association.' *Circulation*, 125(1), Jan 3, pp. e2-e220.

Rogers, R. K., Stoddard, G. J., Greene, T., Michaels, A. D., Fernandez, G., Freeman, A., Nord, J. and Stehlik, J. (2009) 'Usefulness of adjusting

for clinical covariates to improve the ability of B-type natriuretic peptide to distinguish cardiac from noncardiac dyspnea.' *The American Journal of Cardiology*, 104(5), Jul 30, pp. 689-694.

Ronco, C., Haapio, M., House, A. A., Anavekar, N. and Bellomo, R. (2008) 'Cardiorenal Syndrome.' *Journal of the American College of Cardiology*, 52(19), Nov, pp. 1527-1539.

Rondot, E. (1898) 'Le Reflux Hepato-Juglaire.' *Gazette hebdomadaire des science médicales de Bordeaux*, 19 p. 567.

Rudiger, A., Harjola, V.-P., Muller, A., Mattila, E., Saila, P., Nieminen, M. and Follath, F. (2005) 'Acute heart failure: clinical presentation, one-year mortality and prognostic factors.' 7(4), Jul 01, pp. 662-670.

Rutjes, A., Reitsma, J. and Coomarasamy, A. (2007) 'Evaluation of diagnostic tests when there is no gold standard. A review of methods.' *Health Technol ...*, 11(50)

Sackett, D., Haynes, R., Guyatt, H. and Tugwell, P. (2002) *Clinical epidemiology*. Second ed.: Lippincott Williams and Wilkins.

Sackner-Bernstein, J. and Aaronson, K. D. (2005) 'Nesiritide--not verified.' *The New England Journal of Medicine*, 353(14), Oct 06, pp. 1525-1527; author reply 1525-1527.

Sackner-Bernstein, J., Skopicki, H. and Aaronson, K. (2005) 'Risk of worsening renal function with nesiritide in patients with acutely decompensated heart failure.' *Circulation*, 111(12) pp. 1487-1491.

Sanz, M. P., Borque, L., Rus, A., Vicente, B., Ramírez, Y. and Lasa, L. (2006) 'Comparison of BNP and NT-proBNP assays in the approach to the emergency diagnosis of acute dyspnea.' *Journal of Clinical Laboratory Analysis*, 20(6), Feb 01, pp. 227-232.

Saracino, A., Weiland, T., Dent, A. and Jolly, B. (2008) 'Validation of a verbal dyspnoea rating scale in the emergency department.' *Emergency medicine Australasia : EMA*, 20(6), Jul 16, pp. 475-481.

Scal, P. (2005) 'Addressing Transition to Adult Health Care for Adolescents With Special Health Care Needs.' *PEDIATRICS*, 115(6), Jul 01, pp. 1607-1612.

Scharhag, J., Herrmann, M., Weissinger, M., Herrmann, W. and Kindermann, W. (2007) 'N-terminal B-type natriuretic peptide concentrations are similarly increased by 30 minutes of moderate and brisk walking in patients with coronary artery disease.' *Clin Res Cardiol*, 96(4), Apr, pp. 218-226.

Schmidt, D., Delany, D. and McLaurin, L. (1977) 'The clinical recognition of congestive heart failure.' *Journal of Family Practice*, 5(2), Aug 01, pp. 193-197.

Schneider, H., Lam, L., Lokuge, A., Krum, H., Naughton, M., Smit, P., Bystrycki, A., Eccleston, D., Federman, J., Flannery, G. and Cameron, P. (2009) 'B-type natriuretic peptide testing, clinical outcomes, and health services use in emergency department patients With dyspnea a randomized trial.' *Annals of internal medicine*, 150(6) pp. 365-371.

Schwam, E. (2004) 'B-type natriuretic peptide for diagnosis of heart failure in emergency department patients: a critical appraisal.' *Academic Emergency Medicine*, 11(6) pp. 686-691.

Selvais, P. L., Donckier, J. E., Robert, A., Laboux, O., Linden, F., Ahn, S., Ketelslegers, J. M. and Rousseau, M. F. (1998) 'Cardiovascular Diseases-Cardiac natriuretic peptides for diagnosis and risk stratification in heart failure: Influences of left ventricular dysfunction and coronary artery disease on cardiac.' *European Journal of Clinical Investigation*, 28(8) pp. 636-642.

Seronde, M.-F., Gayat, E., Logeart, D., Lassus, J., Laribi, S., Boukef, R., Sibellas, F., Launay, J.-M., Manivet, P., Sadoune, M., Noura, S., Solal, A. C., Mebazaa, A. and network, G. (2013) 'Comparison of the diagnostic and prognostic values of B-type and atrial-type natriuretic peptides in acute heart failure.' *International Journal of Cardiology*, Jun 13, pp. 1-8.

Shah, K., Kop, W., Christenson, R., Diercks, D., Kuo, D., Henderson, S., Hanson, K., Mehra, M. and deFilippi, C. (2009) 'Lack of diagnostic and prognostic utility of circulating plasma myeloperoxidase concentrations in patients presenting with dyspnea.[see comment].' *Clinical Chemistry*, 55(1), Feb 01, pp. 59-67.

Shah, R. V., Truong, Q. A., Gaggin, H. K., Pfannkuche, J., Hartmann, O. and Januzzi, J. L., Jr. (2012) 'Mid-regional pro-atrial natriuretic peptide and pro-adrenomedullin testing for the diagnostic and prognostic evaluation of patients with acute dyspnoea.' *Eur Heart J*, 33(17), Sep, pp. 2197-2205.

Shamim, W., Francis, D., Yousufuddin, M., Varney, S., Pieopli, M., Anker, S. and Coats, A. (1999) 'Intraventricular conduction delay: a prognostic marker in chronic heart failure.' *International Journal of Cardiology*, 70(2), Jul 31, pp. 171-178.

Silver, M., Maisel, A., Yancy, C. W., McCullough, P. A., Burnett, J. C., Francis, G. S., Mehra, M. R., IV, W. F. P., Fonarow, G., Gibler, W. B., Morrow, D. A. and Hollander, J. (2004) 'A Clinical Approach for the Diagnostic, Prognostic, Screening, Treatment Monitoring, and Therapeutic Roles of Natriuretic Peptides in Cardiovascular Diseases.' *Congestive Heart Failure*, 10(5 Suppl 3) pp. 1-30.

Silver, M. A., Horton, D. P., Ghali, J. K. and Elkayam, U. (2002) 'Effect of nesiritide versus dobutamine on short-term outcomes in the treatment of patients with acutely decompensated heart failure.' *JAC*, 39(5), Apr 06, pp. 798-803.

Singer, A., Thode Jr, H., Green, G., Birkhahn, R., Shapiro, N., Cairns, C., Baumann, B., Aghababian, R., Char, D. and Hollander, E. (2009) 'The incremental benefit of a shortness-of-breath biomarker panel in emergency department patients with dyspnea.' *Academic Emergency Medicine*, 16(6), Jul 01, pp. 488-494.

Singh, S. P. and Haider, R. (1973) 'The lack of value of hepato-jugular reflux as a sign of heart failure.' *Postgraduate medical journal*, 49(567), Feb, pp. 10-13.

Song, F., Khan, K. S., Dinnes, J. and Sutton, A. J. (2002) 'Asymmetric funnel plots and publication bias in meta-analyses of diagnostic accuracy.' *International journal of epidemiology*, 31(1), Mar, pp. 88-95.

Sousa, M. R. d. and Ribeiro, A. L. P. (2009) 'Systematic review and meta-analysis of diagnostic and prognostic studies: a tutorial.' *Arquivos Brasileiros de Cardiologia*, 92(3), Apr, pp. 229-238- 235-245.

Sox, C. M., Doctor, J. N., Koepsell, T. D. and Christakis, D. A. (2009) 'The influence of types of decision support on physicians' decision making.' *Arch Dis Child*, 94(3), Mar, pp. 185-190.

SPSS. (2008). PASW Statistics For Windows (18.0 ed.). Chicago: SPSS Inc.

Starling, E. H. (1896) 'On the Absorption of Fluids from the Connective Tissue Spaces.' *The Journal of physiology*, 19(4), Jun 05, pp. 312-326.

Sterne, J. A. and Davey Smith, G. (2001) 'Sifting the evidence-what's wrong with significance tests?' *BMJ*, 322(7280), Feb 27, pp. 226-231.

Steyerberg, E. W., Eijkemans, M. J., Harrell, F. E. and Habbema, J. D. (2000) 'Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets.' *Statistics in Medicine*, 19(8), May 30, pp. 1059-1079.

Stiell, I. and Wells, G. (1999a) 'Methodologic standards for the development of clinical decision rules in emergency medicine.' *Annals of Emergency Medicine*, 33(4) pp. 437-447.

Stiell, I. G. and Wells, G. A. (1999b) 'Methodologic standards for the development of clinical decision rules in emergency medicine.' *Annals of Emergency Medicine*, 33(4), May, pp. 437-447.

Strunk, A., Bhalla, V., Clopton, P., Nowak, R. M., McCord, J., Hollander, J. E., Duc, P., Storrow, A. B., Abraham, W. T., Wu, A. H. B., Steg, G., Perez, A., Kazanegra, R., Herrmann, H. C., Aumont, M.-C., McCullough, P. A. and Maisel, A. (2006) 'Impact of the history of congestive heart failure on the utility of B-type natriuretic peptide in the emergency diagnosis of heart failure: results from the Breathing Not Properly Multinational Study.' *American Journal of Medicine*, 119(1), Feb 01, pp. 69.e61-69.e11.

Struthers, A. D. (2000) 'The diagnosis of heart failure.' *Heart*, 84(3), Sep, pp. 334-338.

Studler, U., Kretzschmar, M., Christ, M., Breidhardt, T., Noveanu, M., Schoetzau, A., Perruchoud, A., Steinbrich, W. and Mueller, C. (2008)

'Accuracy of chest radiographs in the emergency diagnosis of heart failure.' *European Radiology*, 18(8), Aug 01, pp. 1644-1652.

Styron, J., Jois-Bilowich, P., Starling, R., Hobbs, R., Kontos, M., Pang, P. and Peacock, W. (2009) 'Initial emergency department systolic blood pressure predicts left ventricular systolic function in acute decompensated heart failure.' *Congestive Heart Failure*, 15(1) pp. 9-13.

Sudoh, T., Kangawa, K., Minamino, N. and Matsuo, H. (1988) 'A new natriuretic peptide in porcine brain.' *Nature*, 332(6159), Jul 15, pp. 78-81.

Suga S, N. (1992) 'Receptor selectivity of natriuretic peptide family, atrial natriuretic peptide, brain natriuretic peptide, and C-type natriuretic peptide.' *Endocrinology*, 130 pp. 229-239.

Sumida H, Y. (1995) 'Comparison of secretion pattern between A-type and B-type natriuretic peptides in patients with old myocardial infarction.' *Journal of the American College of Cardiology*, 25 pp. 1105-1110.

Sundsfjord, J., Thibault, G. and Cantin, M. (1988) 'Identification and Plasma Concentrations of the N-Terminal Fragment of Proatrial Natriuretic Factor in Man.' *JCEM*, 66, Jun 08, pp. 605-610.

Tamura, N., Itoh, H., Ogawa, Y., Nakagawa, O., Harada, M., Chun, T.-H., Suga, S.-i., Yoshimasa, T. and Nakao, K. (1996) 'cDNA Cloning and Gene Expression of Human Type I $\alpha$  cGMP-Dependent Protein Kinase.' *Hypertension*, 27(3), March 1, 1996, pp. 552-557.

Tang, W. H. W. and Mullens, W. (2010) 'Cardiorenal syndrome in decompensated heart failure.' *Heart*, 96(4), Apr 01, pp. 255-260.

Tayal, V. S. and Kline, J. A. (2003) 'Emergency echocardiography to detect pericardial effusion in patients in PEA and near-PEA states.' *Resuscitation*, 59(3), Dec, pp. 315-318.

Thygesen, K., Alpert, J. and White, H. (2007) 'Universal definition of myocardial infarction.' *Journal of the American College of Cardiology*, 50(22) pp. 2173-2195.

Topol, E. J. (2011) 'The lost decade of nesiritide.' *The New England Journal of Medicine*, 365(1), Jul 07, pp. 81-82.

Townsend, N., Wickramasinghe, K., Bhatnagar, P., Smolina, K., M, N., Leal, J., Luengo-Fernandez, R. and Rayner, M. (2012). Coronary heart disease statistics 2012 edition. London: British Heart Foundation.

Trikalinos, T. A., Balion, C. M., Coleman, C. I., Griffith, L., Santaguida, P. L., Vandermeer, B. and Fu, R. (2012) 'Chapter 8: Meta-analysis of Test Performance When There is a "Gold Standard".' *Journal of General Internal Medicine*, 27(S1), Jun 31, pp. 56-66.

Tsuji, T., Imagawa, K., Masuda, H., Haraikawa, M., Shibata, K., Kono, M., Inouye, K. and Uchida, K. (1994a) 'Stabilization of human brain natriuretic peptide in blood samples.' *Clinical Chemistry*, 40(4) pp. 672-673.

Stabilization of human brain natriuretic peptide in blood samples., 40 Cong. Rec. 672-673 (1994b).

Tsunda T, B. (2002) 'Brain natriuretic peptide is produced in cardiac fibroblasts and induces matrix metalloproteinases.' *Circ Res*, 91 pp. 1127-1134.

Tudor, G. R., Finlay, D. and Taub, N. (1997) 'An assessment of inter-observer agreement and accuracy when reporting plain radiographs.' *Clinical radiology*, 52(3), Apr, pp. 235-238.

van der Burg-de Graauw, N., Cobbaert, C. M., Middelhoff, C. J. F. M., Bantje, T. A. and van Guldener, C. (2009) 'The additive value of N-terminal pro-B-type natriuretic peptide testing at the emergency department in patients with acute dyspnoea.' *European Journal of Internal Medicine*, 20(3), Jun, pp. 301-306.

van Kimmenade, R., Januzzi Jr, J., Ellinor, P., Sharma, U., Bakker, J., Low, A., Martinez, A., Crijns, H., MacRae, C., Menheere, P. and Pinto, Y. (2006) 'Utility of Amino-Terminal Pro-Brain Natriuretic Peptide, Galectin-3, and Apelin for the Evaluation of Patients With Acute Heart Failure.' *Journal of the American College of Cardiology*, 48(6) pp. 1217-1224.

van Kraaij, D., van Pol, P., Ruiters, A., de Swart, J., Lips, D., Lencer, N. and Doevendans, P. (2002) 'Diagnosing diastolic heart failure.' *European Journal of Heart Failure*, 4(4) pp. 419-430.

Vasan, R. S., Benjamin, E. J., Larson, M. G. and Leip, E. P. (2002) 'Plasma natriuretic peptides for community screening for left ventricular hypertrophy and systolic dysfunction: the Framingham heart study.' *JAMA: The Journal of the American Medical Association*,

Vesely, D. (2003) 'Natriuretic peptides and acute renal failure.' *American Journal of Physiology- Renal Physiology*, Feb 01,

Vesely, D. L., Douglass, M. A., Dietz, J. R., Gower, J., William R, McCormick, M. T., Rodriguez-Paz, G. and Schocken, D. D. (1994) 'Clinical Investigation and Reports Three Peptides From the Atrial Natriuretic Factor Prohormone Amino Terminus Lower Blood Pressure and Produce Diuresis, Natriuresis, and/or Kaliuresis in Humans.' *Circulation*, 90 pp. 1129-1140.

Villacorta, H., Duarte, A., Duarte, N., Carrano, Â., Mesquita, E., Dohmann, H. and Ferreira, F. (2002) 'The role of B-type natriuretic peptide in the diagnosis of congestive heart failure in patients presenting to an emergency department with dyspnea.' *Arquivos Brasileiros de Cardiologia*, 79 pp. 569-572.

VMAC (2002) 'Intravenous nesiritide vs. nitroglycerin for treatment of decompensated congestive heart failure. A randomized controlled trial☆.' *JAMA: The Journal of the American Medical Association*, 287(12), Apr 27, pp. 1531-1540.

Wahl, H. G., Graf, S., Renz, H. and Fassbinder, W. (2004) 'Elimination of the cardiac natriuretic peptides B-type natriuretic peptide (BNP) and N-terminal proBNP by hemodialysis.' *Clinical Chemistry*, 50(6), Jul 01, pp. 1071-1074.

Wallace, B. C., Schmid, C. H., Lau, J. and Trikalinos, T. A. (2009) 'Meta-Analyst: software for meta-analysis of binary, continuous and diagnostic data.' *BMC Medical Research Methodology*, 9 p. 80.

Walter, S. D. (2002) 'Properties of the summary receiver operating characteristic (SROC) curve for diagnostic test data.' *Statistics in Medicine*, 21(9), Jun 15, pp. 1237-1256.

Wang, C., FitzGerald, J., Schulzer, M., Mak, E. and Ayas, N. (2005) 'Does this dyspneic patient in the emergency department have congestive heart failure?' *JAMA: The Journal of the American Medical Association*, 294(15) pp. 1944-1956.

Wang, F., Wu, Y., Tang, L., Zhu, W., Chen, F., Xu, T., Bo, L., Li, J. and Deng, X. (2012) 'Brain natriuretic peptide for prediction of mortality in patients with sepsis: a systematic review and meta-analysis.' *Critical Care (London, England)*, 16(3), Jun 06, p. R74.

Wang, H.-K., Tsai, M.-S., Chang, J.-H., Wang, T.-D., Chen, W.-J. and Huang, C.-H. (2010) 'Cardiac ultrasound helps for differentiating the causes of acute dyspnea with available B-type natriuretic peptide tests.' *The American Journal of Emergency Medicine*, 28(9), Nov, pp. 987-993.

Wang, T. J., Gona, P., Larson, M. G., Tofler, G. H., Levy, D., Newton-Cheh, C., Jacques, P. F., Rifai, N., Selhub, J., Robins, S. J., Benjamin, E. J., D'Agostino, R. B. and Vasan, R. S. (2006) 'Multiple biomarkers for the prediction of first major cardiovascular events and death.' *The New England Journal of Medicine*, 355(25), Dec 21, pp. 2631-2639.

Warrell, D., Cox, T. and Firth, J. (Eds.). (2010). *Oxford Textbook of Medicine* (5th ed.). Oxford.

Wasson, J., Sox, H., Neff, R. and Goldman, L. (1985) 'Clinical prediction rules. Applications and methodological standards.' *New England Journal of Medicine*, 313(13) pp. 793-799.

Weber, C. K., Miglioranza, M. H., Moraes, M. A., Sant'anna, R. T., Rover, M. M., Kalil, R. A. and Leiria, T. L. (2014) 'The five-point Likert scale for dyspnea can properly assess the degree of pulmonary congestion and predict adverse events in heart failure outpatients.' *Clinics*, 69(5) pp. 341-346.

Wegner, M., Ganten, D. and Stasch, J. P. (1996) 'Neutral endopeptidase inhibition potentiates the effects of natriuretic peptides in renin transgenic rats.' *Hypertension research : official journal of the Japanese Society of Hypertension*, 19(4), Dec, pp. 229-238.

Weidmann, P., Hellmueller, B., Uehlinger, D., Lang, R., Gnaedinger, M., Hasler, L., Shaw, S. and Bachmann, C. (1986) 'Plasma Levels and Cardiovascular, Endocrine, and Excretory Effects of Atrial Natriuretic Peptide during Different Sodium Intakes in Man.' *Journal of Clinical Endocrinology and Metabolism*, 62(5) pp. 1027-1036.

Wells, P. S., Owen, C., Doucette, S., Fergusson, D. and Tran, H. (2006) 'Does this patient have deep vein thrombosis?' *JAMA*, 295(2) pp. 199-207.

Whiting, P., Rutjes, A. W. S., Reitsma, J. B., Bossuyt, P. M. M. and Kleijnen, J. (2003) 'The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews.' *BMC Medical Research Methodology*, 3(1), Nov 10, p. 25.

Whiting, P., Weswood, M., Rutjes, A., Reitsma, J., Bossuyt, P. and Kleijnen, J. (2006) 'Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies.' *BMC Medical Research Methodology*, 6 p. 9.

Wieczorek, S. J., Wu, A. H. B., Christenson, R., Krishnaswamy, P., Gottlieb, S., Rosano, T., Hager, D., Gardetto, N., Chiu, A., Bailly, K. R. and Maisel, A. (2002) 'A rapid B-type natriuretic peptide assay accurately diagnoses left ventricular dysfunction and heart failure: a multicenter evaluation.' *American Heart Journal*, 144(5), Nov 01, pp. 834-839.

Wilczynski, N. and Haynes, R. (2005) 'EMBASE search strategies for identifying methodologically sound diagnostic studies for use ....' *BMC Medicine*, Feb 01,

Willis, B. H. (2012) 'Empirical evidence that disease prevalence may affect the performance of diagnostic tests with an implicit threshold: a cross-sectional study.' *BMJ Open*, 2(1), Feb 04, pp. e000746-e000746.

Willis, B. H. and Quigley, M. (2011) 'Uptake of newer methodological developments and the deployment of meta-analysis in diagnostic test research: a systematic review.' *BMC Medical Research Methodology*, 11(1), Apr 14, p. 27.

Wu, A. H. B., Packer, M., Smith, A., Bijou, R., Fink, D., Mair, J., Wallentin, L., Johnston, N., Feldcamp, C. S. and Haverstick, D. M. (2004) 'Analytical and clinical evaluation of the Bayer ADVIA Centaur automated B-type natriuretic peptide assay in patients with heart failure: a multisite study.' *Clinical Chemistry*, 50(5) pp. 867-873.

Yamamoto, K., Burnett, J. C., Jougasaki, M., Nishimura, R. A., Bailey, K. R., Saito, Y., Nakao, K. and Redfield, M. M. (1996) 'Superiority of brain natriuretic peptide as a hormonal marker of ventricular systolic and

diastolic dysfunction and ventricular hypertrophy.' *Hypertension*, 28(6), Dec, pp. 988-994.

Yan, A., Yan, R., Spinale, F., Afzal, R., Gunasinghe, H., Stroud, R., McKelvie, R. and Liu, P. (2008) 'Relationships between plasma levels of matrix metalloproteinases and neurohormonal profile in patients with heart failure☆.' *European Journal of Heart Failure*, 10(2), Mar, pp. 125-128.

Yap, L. B., Ashrafian, H., Mukerjee, D., Coghlan, J. G. and Timms, P. M. (2004) 'The natriuretic peptides and their role in disorders of right heart dysfunction and pulmonary hypertension.' *Clinical Biochemistry*, 37(10), Oct, pp. 847-856.

Yasue, H., Yoshimura, M., Sumida, H., Kikuta, K., Kugiyama, K., Jougasaki, M., Ogawa, H., Okumura, K., Mukoyama, M. and Nakao, K. (1994) 'Localization and mechanism of secretion of B-type natriuretic peptide in comparison with those of A-type natriuretic peptide in normal subjects and patients with heart failure.' *Circulation*, 90(1), Jul, pp. 195-203.

Youden, W. (1950) 'Index for rating diagnostic tests.' *Cancer*, 3(1) pp. 32-35.

Zamora, J., Abairra, V., Muriel, A., Khan, K. and Coomarasamy, A. (2006) 'Meta-DiSc: a software for meta-analysis of test accuracy data.' *BMC Medical Research Methodology*, 6(1) p. 31.

Zannad, F., Briancon, S., Juilliere, Y., Mertes, P.-M., Villemot, J.-P., Alla, F. and Virion, J. (1999) 'Incidence, Clinical and Etiologic Features, and Outcomes of Advanced Chronic Heart Failure: The EPICAL Study.' *Journal of the American College of Cardiology*, 33(3) pp. 734-742.

Zeidel, M., Kikeri, D., Silva, P., Burrowes, M. and Brenner, B. (1988) 'Atrial Natriuretic Peptides Inhibit Conductive Sodium Uptake by Rabbit Inner Medullary Collecting Duct Cells.' *J. Clin. Invest.*, 82 pp. 1067-1074.

Zimmerman, S. K. and Vacek, J. L. (2011) 'Imaging techniques in acute coronary syndromes: a review.' *ISRN Cardiol*, 2011 p. 359127.