

Please cite the Published Version

Schmidtke, KA, Watson, DG and Vlaev, I (2017) The use of Control Charts by Laypeople and Hospital Decision-Makers for Guiding Decision Making. *Quarterly Journal of Experimental Psychology*, 70 (7). pp. 1114-1128. ISSN 1747-0218

DOI: <https://doi.org/10.1080/17470218.2016.1172096>

Publisher: SAGE Publications

Version: Accepted Version

Downloaded from: <https://e-space.mmu.ac.uk/622427/>

Additional Information: This is an Author Accepted Manuscript of a paper accepted for publication in *Quarterly Journal of Experimental Psychology*, published by and copyright SAGE Publications.

Enquiries:

If you have questions about this document, contact openresearch@mmu.ac.uk. Please include the URL of the record in e-space. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from <https://www.mmu.ac.uk/library/using-the-library/policies-and-guidelines>)

The Use of Control Charts by Lay-people and Hospital decision-makers for Guiding Decision Making

Word count: 7820

Schmidtke, K. A.,¹ Watson, D. G.,² Vlaev, I.³¹ University of Warwick, Behavioural Science Group; Kelly.Schmidtke@wbs.ac.uk² University of Warwick, Psychology, D.G.Watson@warwick.ac.uk³ University of Warwick, Behavioural Science Group, Ivo.Vlaev@wbs.ac.uk

*Corresponding author is Kelly Ann Schmidtke

Behavioural Science Group
Warwick Business School
The University of Warwick
Coventry, CV4 7AL, UK,
e-mail: Kelly.Schmidtke@wbs.ac.uk.
Phone: 07758933026

Acknowledgement

This article presents independent research commissioned by the National Institute for Health Research (NIHR) under the Collaborations for Leadership in Applied Health Research and Care (CLAHRC) programme, West Midlands. The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. We also thank Dr Richard Lilford for his encouragement throughout this project.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19

Abstract

Graphs presenting healthcare data are increasingly available to support lay-people and hospital staffs' decision-making. When making these decisions, hospital staff should consider the role of chance, i.e., random variation. Given random variation, decision-makers must distinguish signals (sometimes called special-cause data) from noise (common-cause data). Unfortunately, many graphs do not facilitate the statistical reasoning necessary to make such distinctions. Control charts are a less commonly used type of graph that support statistical thinking by including reference lines that separate data more likely to be signals from those more likely to be noise. The current work demonstrates for whom (lay-people and hospital staff) and when (treatment and investigative decisions) control charts strengthen data driven decision-making. We present two experiments that compare people's use of control and non-control charts to make decisions between hospitals (Funnel charts vs League tables) and to monitor changes across time (Run charts with control lines vs Run charts without control lines). As expected, participants more accurately identified the outlying data using a control chart than a non-control chart, but their ability to then apply that information to more complicated questions (e.g., where should I go for treatment?, and should I investigate?) was limited. The discussion highlights some common concerns about using control charts in hospital settings.

Keywords: statistics, decision-making, healthcare

20 **The Use of Control Charts by Lay-people and Hospital decision-makers for Guiding Decision Making**

21 Graphs containing healthcare information are increasingly made available to support data
22 driven decision-making (Wainer, 2013). Unfortunately, such graphs commonly do not facilitate
23 the statistical reasoning necessary for such decisions (Chance, 2002). The present study
24 examines both lay-people's (i.e., university students) use and hospital staff (people positioned to
25 make decisions based on safety and quality performance measures) use of a type of graph
26 designed to support statistically informed decision-making, specifically control charts.

27 **Control charts support statistical reasoning.**

28 Control charts support statistically informed decision-making by including reference lines
29 that highlight the role of chance (Shewart, 1939; Woodall, 2006).¹ Control charts typically
30 include a centre line indicating the central tendency of data and at least two control lines
31 signifying chance variation (Thor, et al., 2007). Examples of such charts are provided in the top
32 half of Figure 1.

33 **The data falling between the upper and lower control lines are likely due to**
34 **common-cause variation. Common-cause variation are naturally anticipated fluctuations in**
35 **any working process, i.e., chance. Note that by 'chance' we do not mean that the data has**
36 **no cause, its cause is likely something that can be expected within the working process. If**
37 **any common-cause data are unacceptable then something about the process that underlies**
38 **all the common-cause data needs to be adjusted for total quality improvement. In contrast,**

¹ One may note that hospital data are often presented as tables and that people often prefer tables over graphs (Hildon, Allwood & Black, 2012). Past research comparing people's use of tables and graphs has found that the two presentation methods are best suited to answer different types of questions (Speier, 2006; Vessey I. 1991). While tables help decision-makers better identify past, unique data, graphs are better at portraying patterns in the data, e.g., 'In what month was the percent of patients waiting more than 4 hours to be seen in A&E highest?' Versus 'Is this percent increasing?' As quality improvement relies on recognizing patterns in data, we concentrate on graphs in the current work.

39 **data falling outside the control lines are more likely to be the result of a special-cause,**
40 **which an investigation could discover (Deming, 1979). Logically any data point that is not**
41 **due to common cause must be due to special cause. However, in practice all control charts**
42 **reveal is the likelihood that data are one or the other and so guide when further**
43 **investigative action is most likely to be useful.** For graphs presenting between-groups
44 comparisons, special-cause data indicate any group performing better or worse than expected by
45 chance, see Figure 1A. For graphs presenting time-series comparisons special-cause data indicate
46 a time when performance was better or worse than expected by chance, see Figure 1B.

47 For proportion data, the width of the upper and lower control lines can be adjusted for
48 unequal sample-sizes point by point. When between-group comparisons are considered, as in
49 Figure 1A, the adjusted control lines take on a funnel-like appearance by arranging the groups on
50 the horizontal axis so that the group representing the smallest sample-size appears first followed
51 by those representing larger sample-sizes. Such charts are called funnel charts (Spiegelhalter,
52 2005). When time-series comparisons are considered, as in a run chart, the adjusted control lines
53 take on a step-like appearance. The steps' extremity depends on the diversity of the sample-sizes;
54 as the sample-sizes in Figure 1B are similar the steps' extremity are slight (See Polit &
55 Chaboyer, 2012, Figure 3 for an example of control charts with more extreme steps).

56 Determining where the control lines are set on charts for different measures should reflect
57 the cost of investigating and the cost of not investigating, in terms of financial cost, quality, and
58 harm. The control lines in Figure 1 are set at 3 standard deviations from the mean. More cautious
59 decision-makers may think this is too lenient and prefer a two standard deviation control line.
60 Such adjustments are not without consequence, as this will increase the chance of a false positive
61 up to 25% depending on the underlying distribution (Kvanli, et al. 2006). Thus without

62 knowledge of the underlying distribution, as is common in healthcare, many prefer to set the
63 control lines at 3 standard deviations. Further instruction on control charts' construction and
64 interpretation is outside the scope of the current paper; for further details please see Amin, 2001
65 or Muhammad, Worthington & Woodall, 2008.

66 **Common graphical methods that do not support statistical reasoning.**

67 League tables and run charts without control lines are more commonly used types of
68 graphs, both of which often do not highlight the role of chance. Examples of such charts are
69 provided in the bottom half of Figure 1. League tables present between-groups comparisons by
70 listing each group's performance in rank-order, see Figure 1C. Such tables are used widely by
71 healthcare regulators and may be used to inform patients (Peymané, et al, 2002). While the
72 ordering in league tables renders the best and worst performers clear, it does not highlight chance
73 variation (Goldstein & Spiegelhalter, 1996). As a result, people using such tables may deem the
74 worst performing hospital as uniquely bad when it is in fact within chance. This is especially a
75 problem when the groups contain different sample-sizes (Tversky & Kahneman, 1974). As
76 discussed previously, funnel charts mitigate this problem by surrounding each group with
77 statistical limits based on sample-size. Lay-people's use of funnel charts and league tables are
78 compared in Experiment 1.

79 Run charts present time-series comparisons, see Figure 1D. Analyses of run charts are
80 often guided by four basic rules to determine whether a process is unstable (Perla, Provost &
81 Murray, 2011):

- 82 • shift - six or more consecutive points either all above or all below the median
- 83 • trend - five or more consecutive points all going up or all going down

84 • runs- too few or too many runs or crossings of the median line, with ‘too many’ given
85 according to tabled critical values one must look up

86 • astronomical point – a point that is different from the rest of the points

87 Notable, precise definitions are available for the first three rules, but the last rule depends on
88 chance variation, which is often not obvious in run charts without control lines.

89 Sometimes people mistakenly identify common-cause data as special-cause data, i.e.,
90 false positives. This is true for both between-groups and time series analyses (Marshall,
91 Mohammed, & Rouse, 2004; Speekenbrink, Twyman & Harvey, 2012). This is a problem
92 because investigations of unique data points within chance variation are unlikely to find special
93 causes, and could divert resources from investigating/altering the entire process. The addition of
94 control lines to run charts mitigates this problem by including statistical reference lines that
95 contextualise all data as falling within or outside of chance variation.² People’s use of run charts
96 with and without control lines are compared in Experiments 1 (lay-people) and Experiment 2
97 (hospital staff).

98 **Can people use control charts effectively?**

99 The current paper compares people’s use of between-group and time-series control charts
100 to commonly used non-control charts. Below we briefly review recent research that has explored
101 the use of control charts.

² For simplicity, the current study focuses on a single type of control chart and a single rule Western Electronic rule to identify irregular data, i.e., any data point outside 3SD is irregular. More complicated run charts that include multiple sets of control lines (e.g., one set at 2 SD and another set at 3 SD’s) enhance decision-makers’ ability to identify lower level statistically irregular trends. For example, while one data point outside of three SD is statistically irregular, two consecutive data outside 2 SD are irregular. While it is tempting to apply as many rules as possible, decision-makers ought to remain cautious as the more rules applied the greater the probability of a false positive. For more information see Amin, 2001.

102 *Between-groups comparisons.* More and more, patients are being allowed to choose
103 which hospital they would like to be treated at for non-emergency procedures. To appreciate how
104 lay-people make this choice, Hildon, Allwood and Black (2012) asked people to examine
105 different displays of hospitals' performance measures to decide which hospital they would prefer
106 for treatment. People had more difficulty understanding, and largely did not prefer funnel charts
107 compared to other presentation methods (e.g., bar charts). However, after researchers explained
108 the purpose of and how to interpret funnel charts, more numerate people warmed to them. But
109 this should not be taken so far as to suggest that people can interpret them properly without
110 assistance (Zikmund-Fisher, Smith, Ubel & Fagerlin, 2007).

111 Hospital decision-makers' use of funnel charts and league tables has already been
112 compared in a randomized controlled trial (Marshall, et al., 2004). In that paper, board members
113 were presented with either three funnel charts or three league tables displaying health service
114 providers' 30-day mortality rates. After examining each graph board members were asked, if
115 "they would take action as a result of the data, and if so to identify the service providers towards
116 whom action would be directed" (p. 310). Those board members who received funnel charts
117 recommended significantly fewer investigations ($M_s = 0.5, 1.0, 0.2$) than those board members
118 who received league tables ($M_s = 0.9, 4.5, 1.6$). **Thus, randomized controlled trial evidence
119 exists to say that hospital decision-makers' calls for investigative action are affected (in this
120 case restrained) by presenting the data in a funnel chart rather than a league table.**

121 More recently, Rakow, Wright, Spiegelhalter and Bull (2014) examined lay-people's
122 interpretation of funnel charts displaying different hospitals' mortality or survival rates. To
123 determine if people understood the basic information presented in funnel charts, they were first
124 asked questions that did not require them to consider the role of chance (e.g., which hospital has

125 the highest survival rate). After this people were asked to imagine they were going for treatment
126 and to say which of two designated hospitals they preferred. Their preferences were sensitive to
127 sample-size, suggesting that funnel charts might help people appreciate the role of chance.

128 While Rakow et al.'s findings are encouraging and inform the current study, two
129 limitations should be noted. First, the two designated hospitals people were asked to choose
130 between both often fell within the control lines. Therefore there was usually no statistically
131 outlying reason to prefer either hospital. The present experiments addresses this concern by using
132 hypothetical data, so that the two designated hospitals have equal performance measures but due
133 to different sample-sizes one data point falls within and the other outside of the control lines. The
134 second limitation is that Rakow's study did not assess participants' use of non-control charts, and
135 so inferences that funnel charts facilitate statistical reasoning better than non-control chart
136 methods might be premature. The current study addresses this second concern by assessing lay-
137 people's use of funnel charts compared to league tables.

138 *Time-series comparisons.* The current study also compares people's use of run charts
139 with and without control lines. Previous work suggests that run charts with control lines can
140 improve healthcare providers' management of many variables, such as asthma attacks,
141 infections, medical errors, and so on. (Alemi & Neuhauser, 2004; Carey & Teeters, 1995;
142 Curran, Benneyan, & Hood, 2002). However, a limitation of many such studies is that they are
143 often repeated measures designs without control groups.

144 At least one randomised controlled field trial has been performed by Curran, et al. (2008).
145 Different hospital wards were or were not provided with control charts displaying their infection
146 performance over several months. Those wards that received control charts decreased their
147 infection rates more than those that did not, but not significantly so. Plausibly, significant

148 differences were not obtained because staff in wards that received control charts interacted with
149 staff in wards that did not, thereby bolstering the latter's performance. The present experiments
150 follow this work but use the alternative approach of a highly controlled laboratory design.

151 Hypotheses.

152 The design of control charts causes statistically irregular data to pop-out (e.g., Tresiman
153 & Gelade, 1980). Therefore, control charts should empower even people untrained in control
154 charts use to identify special-cause data. Accordingly our first hypothesis is that control charts
155 have a strong advantage over non-control charts:

156 H1: People are more likely to accurately identify a special-cause datum when provided
157 with a control chart than a non-control chart.

158 In contrast, control charts do not immediately tell people how to make more difficult
159 decisions, which require further inferences and the applications of further principles. Therefore
160 control charts are likely less able to help people untrained in control chart use to make treatment
161 or investigative choices based on solid statistical reasoning. Regarding treatment choices, lay-
162 people are asked to choose which hospital is more likely to see them within two weeks, may
163 substitute **the intended statistical question with largely non-statistical questions.**

164 **Specifically, instead of focusing on the variability presented between-hospitals, these non-**
165 **statistical questions, will likely focus on whether the hospital is 'small' or 'large', for**
166 **example: *Would I feel more comfortable in small or large hospitals?, or Based on my previous***
167 ***beliefs, are small or large hospitals faster?*** (Kahneman, 2003). Regarding investigative choices,
168 hospital staff asked to monitor data over time may choose to rely on intuitive judgment to
169 determine when investigations are warranted instead of statistical reasoning. Thus, our second
170 hypothesis is as follows:

171 H2: People's responses to more difficult choices (i.e., which hospital to be treated at or
172 whether to investigate) may not differ when provided with a control chart than with a
173 non-control chart.

174

175 **Experiment 1. Lay-people**

176 **Methods**

177 We compare lay-people's use of control and non-control charts with regard to two types
178 of decisions: deciding between different hospitals (i.e., the between-hospitals comparisons), and
179 monitoring performance over time (i.e., the time-series comparisons).

180 **Participants.** One-hundred and seventy participants from the University of Warwick
181 completed an on-line survey (59% Female, $M_{\text{age}} = 21.3$ years, $SD = 3.5$). They were from the
182 Science (N = 57), Business (N = 55) Social Science (N = 46), and Arts Schools (N = 8), or did
183 not say which school (N = 4). No participants said they were part of the Medical School. The
184 majority of these participants (66%) had taken at least one statistics course, but few (16%) knew
185 of control charts. Following completion of the survey, participants were entered into a lottery
186 draw for a chance to win an Amazon gift voucher.

187 **Experimental design.** The experiment compared the responses of participants who
188 examined a control chart with those who examined a non-control chart. Each participant
189 sequentially examined and answered a set of questions about two randomly selected graphs. The
190 first set of questions asked them to examine either a funnel chart or a league table (a between-
191 hospitals comparison), and the second set of questions asked them to examine a run chart that
192 either had or did not have control lines (a time-series comparison).

193 Materials. The survey was created and delivered online using *Qualtrics*©2012.

194 Hypothetical data were used to create the charts in Excel.

195 *Graphs presenting between-hospitals comparisons.* The four graphs presenting between-
196 hospitals comparisons contained data representing 15 hospitals' compliance with the maximum
197 two week wait from receipt of an urgent GP referral for suspected cancer to the date patients are
198 first seen. The original data set was transformed, flipped around its **grand proportion**
199 **maintaining whole numbers**, to create a second data set and both were presented as funnel
200 charts and league tables.³ For an example of such transformation see Appendix A which contains
201 all the graphs. Creating exact transformation of the data was not possible because of the grand
202 proportion (i.e., the centre line) changes when the data are flipped. Within each data set, two
203 hospitals had equal performance (i.e., 99% or 91%) obtained from different sample-sizes (i.e.,
204 201 vs 623). Figure 1A presents one of the funnel charts, and Figure 2A presents its
205 complimentary league table.

206 *Graphs presenting time-series comparisons.* The eight graphs presenting time-series
207 comparisons contained data representing the percentage of patients waiting more than four hours
208 to be seen in a single hospital's emergency department across 12 months. Within each graph the
209 last datum was placed either within or outside of the control line. The original two data sets (one
210 with the last data point within and the other outside of the control lines) were transformed,
211 **flipped around its grand proportion maintaining whole numbers**, to create a second data set,

³ As we wanted the percentages displayed in the charts to convey something that could actually happen, and so whole numbers were needed for the numerator of each hospital's performance. To do this we first found the flipped percentages. From the original data set, we deducted each hospital's percent compliance from its grand mean. Then we deducted these differences from the grand mean. Then, second, to find the new numerator, we multiplied each hospital's denominator by its flipped percentage, and rounded to the nearest whole number. The second funnel chart contains each hospital's percent compliance based on the rounded numerator and the original denominator.

212 and both were presented as run charts with and without control lines.⁴ An exact transformation
213 was not possible, because whether the data is in- or out-of-control depends on the average value
214 which was different for increasing and decreasing data; so adjustments were made to preserve
215 the in- and out-of-control states of the last datum. Figure 1B presents one of the run charts with
216 control lines, and Figure 2B shows the complimentary run chart without control lines. An array
217 of these time-series graphs can be viewed in appendix B.

218 Procedure. **The survey was set up to randomly present the selected graphs (but**
219 **evenly to ensure similar group sizes).** For the first trial, the program selected one of the four
220 available graphs displaying between-hospitals comparisons. For the second trial, the program
221 selected one of the eight available graphs displaying time-series comparisons. Now we describe
222 the questions asked for each graph.

223 *Between-hospitals comparisons.* Participants were asked the following questions about
224 the graphs presenting between-hospitals comparisons; note the first four questions do not require
225 participants to consider chance variation:

226 (Q1) Which hospital (or hospitals) has the lowest percent compliance?

227 (Q2) Which hospital (or hospitals) has the highest percent compliance?

228 (Q3) Which hospital (or hospitals) surveyed the smallest number of patients?

229 (Q4) Which hospital (or hospitals) surveyed the largest number of patients?

230 (Q5) Which hospital's (or hospitals') surveyed percent compliance lies beyond what is
231 likely due to chance variation, three standard deviations from the centre line?

232 Participants responded to each question by selecting the relevant hospital code(s) A-O. Question
233 5 contained two additional choice responses: None and I can't tell.

⁴ As we wanted the percentages displayed in the charts to convey something that could actually happen, we followed the same procedure as that described in footnote 3 to flip the run chart data.

234 (Q6) For this question participants were told to imagine that their GP suspected they
235 might have cancer and would refer them to one of two hospitals with equal performance
236 measures (i.e., either both were 99% or both were 91%). If the graph displayed was a funnel
237 chart, the hospitals were B and M. If the graph displayed was a league table the hospitals were A
238 and B or N and O. Participants were asked to use the information provided in the graph to
239 determine which hospital would most likely see them within two weeks. The four response
240 options included the following: (A) select hospital [insert relevant hospital code], (B) select
241 hospital [insert relevant hospital code], (C) I am equally likely to be seen in two weeks or sooner
242 at either hospital, and (D) I do not understand the question. The correct answer when the graph
243 presented two equally high data points was M for the funnel chart and B for the league table, and
244 the correct answer when the graph presented two equally low data points was B for the funnel
245 chart and O for the league table.

246 *Time-series comparisons.* Participants were asked the following questions about the
247 graphs presenting time-series comparisons; note that the first two questions do not require
248 participants to consider chance variation:

249 (Q7) Which month (or months) has the highest percent of patients waiting over four
250 hours?

251 (Q8) Which month (or months) has the lowest percent of patients waiting over four
252 hours?

253 (Q9) Which month's (or months') percent lies beyond what is likely due to chance
254 variation, three standard deviations from the centre line?

255 Participants responded to each question by selecting the relevant month(s), March - February.
256 Question 8 contained two additional choice responses: None and I can't tell.

257 (Q10) For this question participants were told to imagine they were a hospital manager
258 who based on the information provided in the graph needed to decide whether to investigate the
259 change that occurred in February. They were cautioned that while investigations can improve
260 future performance measures, such investigations are costly and they should not investigate data
261 which are likely the result of chance variation. The four response options included the following:
262 (A) Yes, (B) No, (C) I have no preference, and (D) I do not understand this question.

263 Analyses. To organise our results, we first describe participants' responses to the graphs
264 presenting between-hospitals and then time-series comparisons. We first present descriptive
265 statistics of responses to each of the questions. Next, the participants' responses are
266 dichotomously categorised as either correctly interpreting the information contained in the graph
267 or not (e.g., identifying the statistically designated special-cause datum or not, making the
268 statistically informed decision or not). With these dichotomised responses we use a Chi-square
269 test to compare the performance of those participants who saw a control chart with those who
270 saw a non-control chart. The results for the more difficult questions (i.e., hospital preference or
271 investigative choice) are shown in Figure 2.

272 **Experiment 1. Results**

273 Between-hospitals comparisons.

274 Similar numbers of participants received each graph type; 81 saw a funnel chart and 86
275 saw a league table.

276 *Identifying the Most Extreme data points.* No matter which chart participants viewed,
277 their responses were largely accurate. Of participants who saw a funnel chart 77% correctly
278 identified the lowest and 70% the highest performing hospital; 85% correctly identified the
279 smallest and 82% the largest sized hospital. Of those who saw a league table 77% correctly

280 identified the lowest and 84% the highest performing hospital; 93% correctly identified the
281 smallest and 87% the largest sized hospital. The Chi-square tests only found a difference for the
282 question asking participants to identify the highest performing hospital, favouring league tables
283 ($\chi^2(1, N = 170) = 4.37, p = 0.04, \phi = 0.16$). None of the other questions yielded statistically
284 different results ($ps > 0.08$).

285 *Identifying the special-cause datum.* Funnel charts increased participants' ability to
286 identify the special-cause datum. Given a funnel chart, 51% correctly identified the special-cause
287 datum; 24% selected one or more incorrect options, and 25% said they could not tell. For the
288 league table only 1% correctly identified the special-cause datum; 48% selected one or more
289 incorrect options, and 51% said they could not tell. A Chi-square test determined that
290 participants presented with a funnel chart responded more accurately than those presented with a
291 league table, $\chi^2(1, N = 170) = 55.44, p < 0.001, \phi = 0.57$.

292 *Hospital Choice.* Participants struggled to identify the hospital that would most likely see
293 them within two weeks, no matter which graph they saw. Given a funnel chart 43% chose the
294 statistically recommended hospital, 31% chose the other hospital, 24% said they were equal and
295 2.4% reported that they did not know. For the league table 49% chose the statistically
296 recommended hospital, 27% chose the other hospital, and 24% said they were equal. The Chi-
297 square test did not find a significant difference ($p = 0.43$).

298 Time-series comparisons.

299 Similar numbers of participants received each time-series graph type; 87 received a run
300 chart with control lines and 84 a run chart without control lines.

301 *Identifying the Most Extreme data points.* No matter which chart participants viewed,
302 their responses were largely accurate. Of participants who saw a run chart with control lines,

303 88% correctly identified the lowest and 93% the highest performing hospital. Of those who saw a
304 run chart without control lines, 96% correctly identified the lowest and 100% the highest
305 performing hospital. The run chart without control lines encouraged more accurate responses for
306 both these questions as analysed using the Chi-square tests, $\chi^2_s(1, N = 170) > 3.90$, $ps < 0.05$, ϕ
307 = 0.15.

308 *Identifying the special-cause datum.* Run charts with control lines increased participants'
309 ability to identify the presence or absence of the special-cause datum. Of participants who saw a
310 run chart with control lines, 64% correctly identified the presence or absence of the special-cause
311 datum, 16% were incorrect and 20% said they could not tell. Of participants who saw a run chart
312 without control lines, 13% correctly identified the special-cause datum, 36% selected incorrect
313 data and 51% said they could not tell. A Chi-square test found that those participants presented
314 with a run chart with control lines responded more accurately than those presented with a run
315 chart without control lines, $\chi^2(1, N = 170) = 48.7$, $p < 0.005$, $\phi = 0.54$.

316 *Investigative Choice.* When asked whether they would call for an investigation,
317 participants struggled no matter which graph they saw. Of participants who saw a run chart with
318 control lines, 56% chose the statistically recommended course of action, 40% chose the other
319 course of action, 4% had no preference and 1% did not understand the question. Of participants
320 who saw a run chart without control lines, 48% chose the statistically recommended course of
321 action, 45% chose the other course of action, 6% had no preference, and 1% indicated they did
322 not understand the question. A Chi-square test did not find a significant difference, ($p = 0.29$).

323 **Experiment 1. Discussion**

324 This experiment compared lay-people's use of control charts to non-control charts.
325 Confirming our hypotheses, control charts helped participants identify the special-cause datum;

326 but did not help participants answer more difficult questions (i.e., hospital preference or
327 investigative choice). This suggests training people to use control charts should not focus on
328 helping them identify special-cause data; rather training should focus on how people should
329 apply such information to support more difficult decisions.

330 Regarding hospital choice, one may note that most people do not consider quantitative
331 healthcare data when choosing which hospital to attend. Rather people are more likely to rely on
332 their GP's advice or the experiences of their friends (Department of Health 2009; Dixon, et al.,
333 2010). However, this should not be taken to mean people do not want more reliable information.
334 People report that health websites allowing them to compare hospitals are a useful source of
335 information, which they believe should be more widely available (Boyce, et al., 2010). To
336 support people's desire for such information, general practitioners could inform them that such
337 websites are available when patients are most likely to want this information. Our general
338 discussion provides an idea for how to make the information provided in control charts easier to
339 interpret thus allowing people to better account for the role of chance in their decisions.

340 The next section discusses Experiment 2. This experiment was conducted to account for
341 two concerns presented by Experiment 1 about the results for graphs presenting time-series
342 comparisons. The first is that our participants were inexperienced with investigative **hospital**
343 questions and so our results may not generalise to **hospital decision-makers**. The second
344 concern is that participants in Experiment 1 were only asked whether they would call for an
345 investigation based on the data provided. Such a question conflates participants' ability to
346 interpret the statistical recommendations contained in the graphs with the desire to actually
347 investigate. However, as lay-people have no experience with actually calling for an investigation
348 there may be no meaningful difference between these questions for them. In contrast, there is

349 more likely a difference between the questions for experienced hospital decision-makers. To
350 address these concerns, in Experiment 2 we surveyed hospital decision-makers who are
351 experienced in these choices, and replaced the investigative choice question with questions about
352 the graphs' investigative recommendations and their actual investigative choices.

353

354 **Experiment 2. Hospital Decision-Makers**

355

Methods

356 Participants. A lead consultant invited high level hospital decision-makers within one
357 United Kingdom NHS trust to voluntarily complete this survey. In total, 47 participants (45%
358 female) completed the survey (2.1% under 31 years old; 68.1% = 31-50 years old; 29.8% 51-70
359 years old). These participants included consultants (60%), managers (25%), doctors (9%), nurses
360 (4%), and 1 clinical audit adviser (2%). Most (87%) had two or more years of work experience
361 in a decision-making capacity. Nearly three quarters of the participants recalled completing a
362 formal statistics course (70%) and more than half knew of control charts (60%). Following
363 completion of the survey, participants were entered into a lottery draw for a chance to win an
364 Amazon gift voucher.

365 Materials. The survey was created and delivered on-line using *Qualtrics*©2012. The
366 time-series graphs of Experiment 1 were used but with one alteration. This alteration was to add
367 the sample-size aside each month on the horizontal axis. Although sample-size information is
368 rarely included in charts, such information affects where the control lines are set, which could
369 plausibly be considered useful, and so we wanted it to be available for the hospital decision-
370 makers.

371 Procedure. Participants were presented with two graphs presenting time-series
372 comparisons. The same questions used in Experiment 1 were used to assess participants' ability
373 to interpret the graphs and identify the presence or absence of special-cause data (Q7-Q9). To
374 address the concerns raised for Experiment 1, the previously used investigative choice question
375 was replaced with two questions designed to distinguish participants' abilities to interpret the
376 graphs' investigative recommendations and their actual investigative choices.

377 (Q11) The question requiring participants to interpret the graphs' investigative
378 recommendations asked: "Regardless of what you would actually do, does the information
379 provided in the chart above suggest you should call for an investigation to learn more about the
380 increase (or decrease) in February?" The four response options included the following: (A) Yes,
381 (B) No, (C) This chart makes no recommendations about what I should do and (D) I do not
382 understand this question.

383 (Q12) The question asking for participants' actual investigative choices asked: "In reality,
384 if these performance data were from your hospital would you call for an investigation to learn
385 more about the increase (or decrease) in February?" The four response options included the
386 following: (A) Yes, (B) No, (C) I don't know and (D) I do not understand this question. After
387 responding, participants were asked to further explain why they would or would not actually
388 investigate based on data like those presented in the graph.

389 **Experiment 2. Results**

390 The results are reported in a similar manner as Experiment 1. However, due to the small
391 expected values in Experiment 2 (a result of a smaller sample size), **Fisher's** exact test was
392 applied rather than the Chi-squared test. Similar numbers of participants received each time-

393 series graph type; 24 received a run chart with control lines and 23 received a run chart without
394 control lines.

395 *Identifying the Most Extreme data points.* No matter which chart participants viewed,
396 their responses were largely accurate. Of participants who saw a run chart with control lines,
397 100% correctly identified the lowest and 100% the highest performing hospital. Of those who
398 saw a run chart without control lines, 96% correctly identified the lowest and 91% the highest
399 performing hospital. Neither question statistically differed when analysed using Fisher's test (p
400 > 0.23).

401 *Identifying the special-cause datum.* The control charts increased participants' ability to
402 identify the presence or absence of the special-cause datum. Of participants who saw a run chart
403 with control lines, 63% correctly identified the presence or absence of the special-cause datum,
404 13% were incorrect and 25% said they could not tell. Of participants who saw a run chart
405 without control lines, only 4% correctly identified the special-cause datum, 17% selected
406 incorrect data and 78% said they could not tell. Fischer's test showed that those presented with a
407 run chart with control lines responded more accurately than those presented with a run chart
408 without control lines ($p < 0.001$).

409 *Investigative recommendation.* The statistical recommendations given by control charts
410 were accurately interpreted by most participants. Of the participants who saw a run chart with
411 control lines, 79% correctly identified the course of action the control chart recommended, 8%
412 chose the other course of action, 13% said the chart made no recommendations. Of the
413 participants who saw a run chart without control lines, 9% chose the recommended course of
414 action, 26% choosing the other course of action, 61% said the chart made no recommendations
415 and 4% did not understand the question. A Fischer's test showed that those presented with a run

416 chart with control lines responded more accurately than those presented with a run chart without
417 control lines ($p < 0.001$).

418 *Investigative choice.* The recommendations made by the control charts were largely
419 followed by our participants. Of the participants who saw a run chart with control lines, 68% said
420 they would follow the statistically recommended course of action, 29% chose the other course of
421 action, 4% did not know. Participants who saw a run chart without control lines were more
422 evenly split; 39% would follow the statistically recommended course of action, 22% chose the
423 other course of action, 40% said they did not know. A Fisher's test showed that those presented
424 with a run chart with control lines responded more accurately than those presented with a run
425 chart without control lines ($p < 0.001$).

426 **Experiment 2. Discussion**

427 Experiment 2 compared hospital decision-makers' use of run charts with control lines to
428 run charts without control lines. Notably, when provided with a run chart with control lines, as
429 opposed to one without, hospital decision-makers' were better able to identify the special-cause
430 datum, interpret the recommendations made by the chart and apply those recommendations to
431 their investigative choice. These results support our first, but not second hypothesis.

432 **General Discussion**

433 Control charts facilitate statistically informed decision-making better than non-control
434 charts. This is true for both lay-people and hospital decision-makers. The advantage of using
435 control charts is more easily observed with simple questions (is there any special-cause data?),
436 than more difficult questions (which hospital is more likely to see you within two weeks, or
437 would you call for an investigation?). In the following discussion we first compare and contrast

438 the responses of lay-people and hospital decision-makers. Lastly, we close with
439 recommendations to help people effectively employ control charts.

440 **Limitations**

441 It should be noted that our lay-people were recruited from a university and many had
442 statistical training. While such people are part of the general population, further work with non-
443 university samples would be a welcomed addition to this literature. **Our sample of lay-people**
444 **was a convenient sample with which we could initially explore how control charts influence**
445 **people's decisions.**

446 Another limitation of the current study is that we did not assess hospital decision-makers
447 use of funnel charts and league tables, and so cannot compare lay-people and hospital decision-
448 makers' use of them. We did not assess hospital decision-makers use of funnel charts and league
449 tables for at least two reasons. First, a randomized controlled trial of this comparison was already
450 performed in 2004 (Marshall et al.). Second, the hospital requested we keep the survey brief.

451 Comparing lay-people and hospital decision-makers

452 Descriptively our lay-people and hospital decision-makers exhibited some similarities in
453 their choices. For Q7, when using a control chart 63% of hospital decision-makers and 64% lay-
454 people correctly identified the presence or absence of the special-cause datum. Hospital decision-
455 makers were more cautious. Provided with a run chart without control lines, approximately three
456 in every four hospital decision-makers said they could not tell whether any data were special-
457 cause, fewer (approximately two in every four) lay-people did. This difference may reflect
458 hospital decision-makers being better acquainted with the costs of making a mistake, and their
459 need to be more certain before answering statistical questions. Control charts provide such
460 certainty.

461 A major difference between populations was their ability to use the control lines to decide
462 whether to investigate. Lay-people's investigative choices were not affected by the control lines,
463 but hospital decision-makers' choices were. This difference may suggest that lay-people do not
464 understand or appreciate the thinking behind decisions that policy makers and hospital decision-
465 makers make. For example, they may not understand how costly investigations may become,
466 how often data appear aberrant or how systems changes need not include investigations of
467 particular data.

468 The current study shows that control charts are indeed useful. Hospital decision-makers
469 should use them more often. While the use of control charts in healthcare is increasing, they are
470 still very much underused (Koetsier, et al., 2012; Taylor, et al., 2014). In organizations where
471 control charts are not presently used, a prevailing social norm may prevent their introduction
472 (For a demonstration of control charts being introduced into a plausibly resistant setting, see de
473 Leval et al., 1994). Our work demonstrates empirically some benefits that control charts offer
474 and therefore may provide the spark that organizations need to start using them.

475 **Helping people effectively employ control charts.**

476 People may have problems interpreting graphs in general. To help these people, Rakow,
477 et al., 2014 note that carefully constructing the graphs so that the axes provide the right
478 information (**e.g.**, mortality vs survival) is an important step to take. Control charts however offer
479 an additional challenge, in that if people do not know what control lines represent they will likely
480 ignore them. Indeed, Zikmund-Fisher's (2007) research found that people are largely unfamiliar
481 with control charts and experience difficulty interpreting them. This is not a problem with
482 control charts, but rather the users' ability to interpret them. This barrier may be overcome with
483 educational interventions. **Curran, et al. (2008) used a very light** educational intervention,

484 wherein their control charts were accompanied by one or two sentences describing the observed
485 variation and any out-of-control episodes or trends.

486 Statistically more astute decision-makers may also experience difficulties using control
487 charts. Wheeler (2011) and Balestracci (2011) offer more complete discussions of such concerns.
488 Based on their work, we provide a brief description of two of the concerns below.

489 *Concern 1.* Many statistically **astute** people may think that if control charts data are not
490 normally distributed then you cannot use a control chart. This premise is wrong. The control
491 charts' use is largely insensitive to the data's distribution, which is particularly true for
492 individual control charts (Wheeler, 1995). When organizations measure a process over time, they
493 often do not know whether the process is stable enough for the data to be treated as if it had
494 come from a single population (with a well-defined distribution); control charts help determine
495 whether this is the case. Even if the data to be plotted are not normally distributed, control charts
496 still have a reasonable false positive rate and any rule breaks warn that the data may be coming
497 from different processes.

498 For quality improvement efforts in hospital organisations, a further distinction should be
499 made between two phases of control chart use. In Phase one control charts are used in an
500 explorative and iterative fashion to eradicate factors that cause worrisome variations and in so
501 doing create an in-control or stable process. Ensuring process stability in itself often improves
502 quality performance, but such improvements may not reach the desired specifications. Phase two
503 can then be used to promote further quality improvements. Similar to hypothesis testing, in Phase
504 two control charts are used to detect deviations from an expected distribution (aka outliers in the
505 hypothesis testing literature). Phase one is a necessary step before advancing to Phase two

506 because if existing special-causes are not yet understood, they will complicate experimental
507 attempts to improve the process (for more information see: Woodall, 2000).

508 *Concern 2.* Another concern raised by the statistical astute regards where the control lines
509 are set. Determining, precisely where the control lines should be set on charts for different
510 measures should reflect the cost of investigation and the cost of not investigating, in terms of
511 financial costs, quality, and harm. This is a question of judgement and cannot be resolved
512 statistically, i.e., that the lines on a control charts should always be set a number of deviations
513 from the central tendency. When more variation is acceptable, then wider set control lines (**lower**
514 **sensitivity**) may be appropriate. **However the precise position of the control lines must arise**
515 **from the data themselves, as opposed to the precise position of a target lines which may be**
516 **based on a precise external standard (e.g., 95% of patients attending an Accident and**
517 **Emergency department are seen within four hours of arrival).** Additionally, the control lines
518 need not reflect a normal distribution; for example, a binomial distribution is often used to set
519 control limits for proportion data.

520 To briefly demonstrate a consequence of setting the control lines more or less
521 conservatively the following case is offered. Given a normal distribution, setting the control lines
522 at three standard deviations is very conservative. At three standard deviations, the control lines
523 will encompass about 99.5% of the data in a stable process, creating a small false alarm rate
524 0.25%. But here the astute statistician notes that such a conservative setting also increases the
525 number of times truly concerning data is overlooked, i.e., the miss rate. If indeed the distribution
526 is normal one may be more comfortable setting the control lines at two standard deviations, so
527 that about 95% of the data in a stable process are encompassed, the false alarm rate is still only
528 5%, and the number of misses is reduced.

529 However, data are not always normally distributed. As stated in the introduction, if the
530 data come from an arbitrary distribution, 25% of data can be located beyond two standard
531 deviations depending on the shape of the underlying distribution (Chebyshev's inequality), and
532 so the chance for false alarms is quite high (Kvanli, Pavur & Keeling, 2006). Put another way, if
533 hospital decision-makers set the control limit at two standard deviations, this could prompt them
534 to investigate up to 1 in every 4 data; in most cases this would be infeasible. In contrast, if the
535 control lines are set at three standard deviations, this could prompt them to investigate 1 in every
536 10 data. Thus setting the control lines at three standard deviations may not be as conservative as
537 first thought.

538 Ultimately, no probabilistic detector (including intuition) is 100% accurate, but
539 investigating every data point is infeasible. Compared to intuitions, control charts allow one to
540 better understand variation in performance measures and in so doing better focus quality
541 improvement efforts. Without control lines variation is often ignored, so at the very least control
542 lines prompt crucial discussions about the costs and benefits of investigative action. Specifically,
543 when data are statistically irregular, they are more likely to be the result of a special-cause that
544 an investigation can identify to guide quality improvement efforts. When data are statistically
545 regular, i.e., common-cause data, there is nothing an investigation can find and so investigations
546 absorb resources that could be better spent elsewhere. Where common-cause data are
547 unacceptable quality improvement efforts should focus on modifying processes that are common
548 to all the data.

549

550 **References**

- 551 Alemi, F., & Neuhauser, D. (2004). Time-between control charts for monitoring asthma
552 attacks. *Joint Commission Journal on Quality Improvement and Patient Safety*, 30, 95-
553 102.
- 554 Amin, S. G. (2001). Control charts 101: A guide to health care applications. *Quality*
555 *Management in Health Care*, 9(3), 1-27.
- 556 Balestracci, D. (2011). Four control chart myths from foolish experts: Don't teach people
557 statistics—teach them to solve problems. *Quality Digest*, Retrieved March 23, 2015,
558 from: [http://www.qualitydigest.com/inside/quality-insider-article/four-control-chart-](http://www.qualitydigest.com/inside/quality-insider-article/four-control-chart-myths-foolish-experts.html)
559 [myths-foolish-experts.html](http://www.qualitydigest.com/inside/quality-insider-article/four-control-chart-myths-foolish-experts.html)
- 560 Boyce, T., Dixon, A., Fasolo, B., et al. (2010). Choosing a high quality hospital: the role of
561 nudges, scorecard design and information. The Kings Fund: London.
- 562 Carey, R. G., & Teeters, J. L. (1995). CQI case-study—reducing medication errors. *Joint*
563 *Commission Journal on Quality Improvement*, 21, 232-237.
- 564 Chance, B. L. (2002). Components of Statistical Thinking and Implications for Instruction and
565 Assessment, *Journal of Statistics Education [Online]*, 10(3)
566 www.amstat.org/publications/jse/v10n3/chance.html.
- 567 Curran, E. T., Benneyan, J. C., & Hood, J. (2002). Controlling Methicillin-Resistant
568 *Staphylococcus Aureus*: A feedback approach using annotated statistical process control
569 charts. *Infection Control and Hospital Epidemiology*, 23, 13-18.
- 570 Curran, E., Harper, P., Loveday, H., Gilmour, H., Jones, S., Benneyan, J., Hood, J., & Pratt, R.
571 (2008). Results of a multicentre randomised controlled trial of statistical process control
572 charts and structured diagnostic tools to reduce ward-acquired meticillin-resistant

- 573 Staphylococcus aureus: the CHART Project. *Journal of Hospital Infection*, 70(2), 127-35.
574 doi: 10.1016/j.jhin.2008.06.013
- 575 de Leval, M.R., Francois, K., Bull, C., Brawn, W. & Spiegelhalter, D. (1994). Analysis of a
576 cluster of surgical failures: application to a series of neonatal arterial switch operations.
577 *Journal of Thoracic and Cardiovascular Surgery*,
578 107, 914-924.
- 579 Deming, W. E. (1975). On probability as a basis for action, *The American Statistician*, 29(4),
580 146-152. doi: 10.1080/00031305.1975.10477402
- 581 Department of Health. (2009). Report on the National Patient Choice Survey – March 2009
582 England [online]. Available at:
583 ww.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/documents/digitalasset/dh_1036
584 81.pdf (accessed on 27 September 2015).
- 585 Dixon, A., Robertson, R., Appleby, J., Burge, P., Devlin, N., Magee, H. (2010). Patient choice:
586 How patients choose and how providers respond. London: The King's Fund. Available
587 at: www.kingsfund.org.uk/publications/patient_choice.html. (accessed on 28 September
588 2015).
- 589 Goldstein, H., & Spiegelhalter, D. J. (1996). League tables and their limitations: Statistical issues
590 in comparisons of institutional performance. *Journal of the Royal Statistical Society:*
591 *Series A*, 159, 385-443.
- 592 Hildon, Z., Allwood, D., & Black, N. (2012). Making data more meaningful: Patients' views of
593 the format and content of quality indicators comparing health care providers. *Patient*
594 *Education and Counseling*, 88(2), 298-304. doi: 10.1016/j.pec.2012.02.006
- 595 Kahneman, D. (2003). A perspective on judgement and choice. *American Psychologist*, 58, 697-

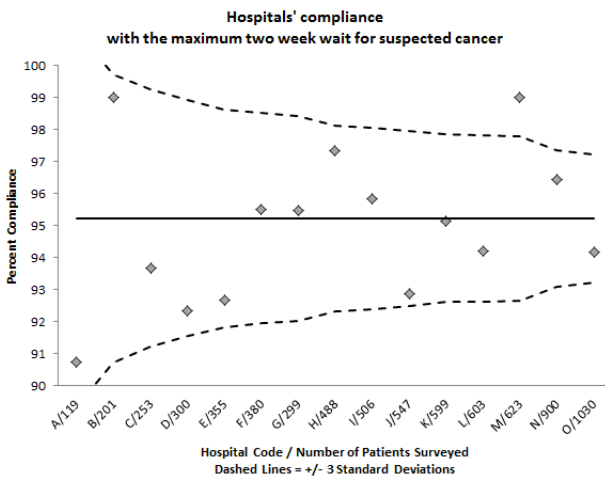
- 596 720. doi:10.1037/0003-066x.58.9.697
- 597 Koetsier, A., van der Veer, S. N, Jager, K. J., Peek, N., & de Keizer, N. F. (2012). Control charts
598 in healthcare quality improvement: A systematic review on adherence to methodological
599 criteria. *Methods of Information in Medicine*. 51(3), 189-198.
- 600 Kvanli, A. H., Pavur, R. J., & Keeling, K. B. (2006). *Concise Managerial Statistics*. cEngage
601 *Learning*. 81-82. ISBN 9780324223880
- 602 Marshall T, Mohammed MA, & Rouse A. A (2004). A randomized controlled trial of league
603 tables and control charts as aids to health service decision-making. *International Journal*
604 *for Quality in Health Care*, 16(4), 309-315.
- 605 Mohammed, M., Worthington, P., & Woodall, W. H. (2008). Plotting basic control charts:
606 tutorial notes for healthcare practitioners. *Quality and Safety in Health Care*, 17, 137-
607 145. doi: 10.1136/qshc.2004.012047
- 608 Perla, R. J., Provost, L. P., & Murray, S. K. (2011). The run chart: a simple analytical tool for
609 learning from variation in healthcare processes. *BMJ Quality and Safety*, 20(1), 46-51.
610 doi: 10.1136/bmjqs.2009.037895
- 611 Peymané, A., Rouse, A. M, Mohammed, M. A, Marshall T. (2002). Performance league tables:
612 The NHS deserves better. *BMJ*, 324(7329), 95–98.
- 613 Polit, D. F. & Chaboyer, W. (2012). Statistical process control in nursing research. *Research in*
614 *Nursing and Health*, 35, 82-93.
- 615 Rakow, T., Wright, R. J., Spiegelhalter, D. J., & Bull, C. (2014). The pros and cons of funnel
616 plots as an aid to risk communication and patient decision making. *British Journal of*
617 *Psychology*, 106(2), 327-348.
- 618 Shewhart, W. A. (1939). *Statistical Method from the viewpoint of Quality Control*. Graduate

- 619 School of the Department of Agriculture, Washington, D.C.
- 620 Speekenbrink, M., Twyman, M., Harvey, N. (2012). Change detection under
621 autocorrelation. CogSci 2012 Sapporo, Japan. Proceedings of the 34th Annual
622 Conference of the Cognitive Science Society.
- 623 Spiegelhalter, D. J. (2005). Funnel plots for comparing institutional performance. *Statistics in
624 Medicine*, 24(8), 1185-1202. doi: 10.1002/sim.1970
- 625 Sunstein, C. R. (2005). Moral Heuristics. *Behavioral and Brain Sciences*, 28(4), 531-542. doi:
626 10.1017/S0140525X05000099
- 627 Taylor, M. J., McNicholas, C., Nicolay, C., Darzi, A., Bell, D., & Reed, J. E. (2014). Systematic
628 review of the application. *BMJ Quality & Safety*, 23(4), 290-298. doi: 10.1136/bmjqs-
629 2013-001862.
- 630 Thor, J., Lundberg, J., Ask, J., Olsson, J., Carli, C., Harenstam, K. P., & Brommels, M. (2007).
631 Application of statistical process control in healthcare improvement: Systematic review.
632 *Quality Safety in Health Care*, 16, 387-399. doi: 10.1136/qshc.2006.022194.
- 633 Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases.
634 *Science*, 185(4157), 1124-1131. doi:10.1126/science.185.4157.1124.
- 635 Treisman, A., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive
636 Psychology*, 12, 97-136.
- 637 Wainer, H. (2013). *Medical Illuminations: Using Evidence, Visualization and Statistical
638 Thinking to Improve Healthcare*. Oxford University Press. ISBN 0199668795.
- 639 Wheeler, D. J. (1995). *Advanced Topics in Statistical Process Control*. Statistical Process
640 Controls, Inc., Knoxville, TN.
- 641 Wheeler, D. J. (2011). Myths about process behavior charts: How to avoid some common

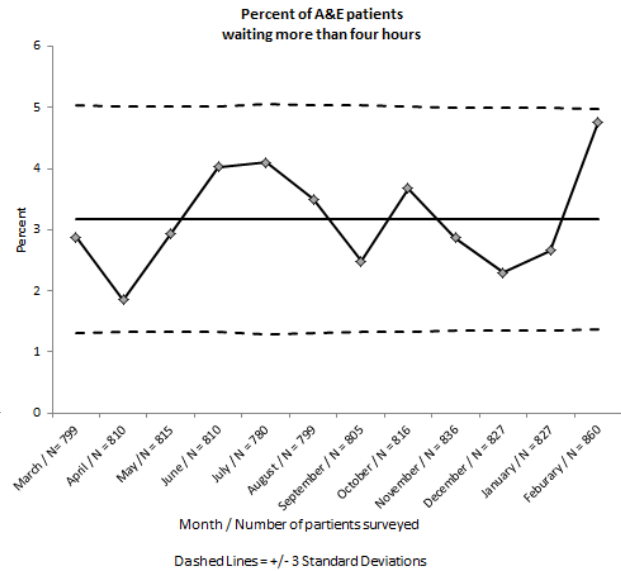
- 642 obstacles to good practice. *Quality Digest*, retrieved March 23, 2015, from:
643 [http://www.qualitydigest.com/inside/quality-insider-article/myths-about-process-
645 behavior-charts.html#](http://www.qualitydigest.com/inside/quality-insider-article/myths-about-process-
644 behavior-charts.html#)
- 645 Woodall, W. H. (2000). Controversies and contradictions in statistical process control. *Journal of*
646 *Quality Technology*, 32(4), 341-350.
- 647 Woodall WH. (2006). The use of control charts in health-care and public-health surveillance
648 (with discussion). *Journal of Quality Technology*, 38, 104-105.
- 649 Zikmund-Fisher, B. J., Smith, D. M , Ubel, P. A., & Fagerlin, A. (2007). Validation of the
650 subjective numeracy scale: Effects of low numeracy on comprehension of risk
651 communications and utility elicitations. *Medical Decision Making*, 27, 663-671.
- 652

Control charts

A. Funnel chart

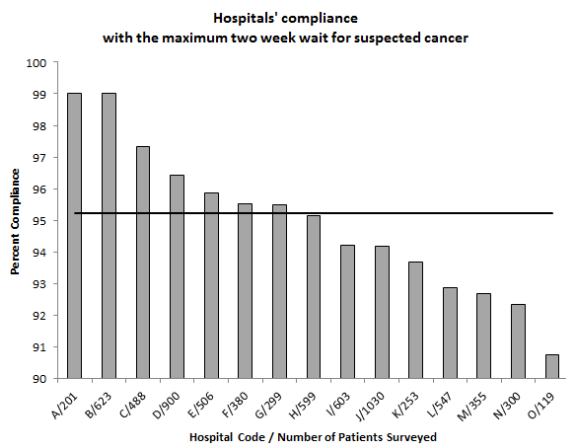


B. Run chart with control lines

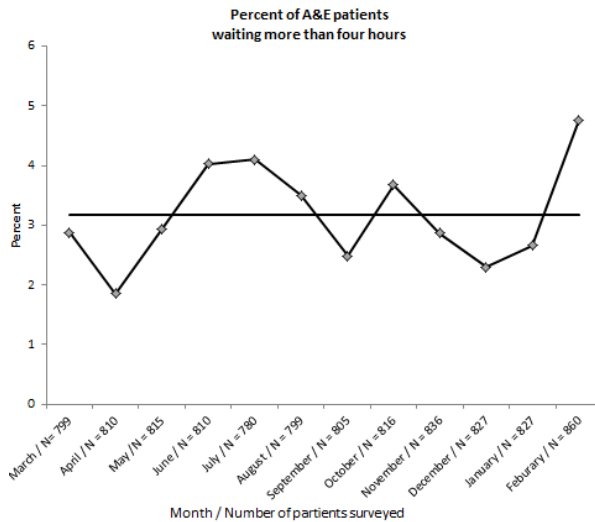


Non-control charts

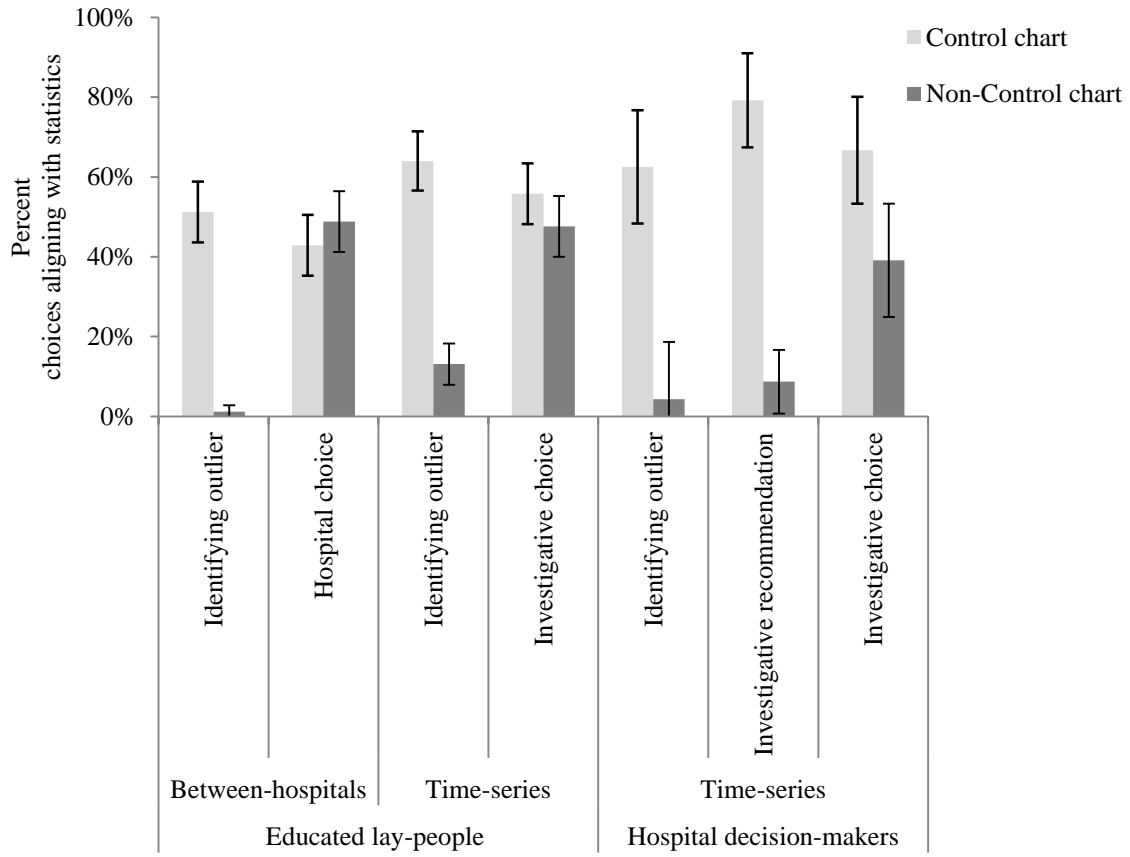
C. League table



D. Run chart without control lines



653 Figure 1. Example control charts and non-control charts presenting between-groups (funnel
654 charts and league tables) and time-series comparisons (run charts).



655

656 *Figure 2.* A graphical representation of the Chi-square tests. Error bars represent 2 SEs.

657

658 Figure Captions

659 *Figure 1.* Example control charts and non-control charts presenting between-groups (funnel
660 charts and league tables) and time-series comparisons (run charts).

661 *Figure 2.* A graphical representation of the Chi-square tests. Error bars represent 2 SEs.

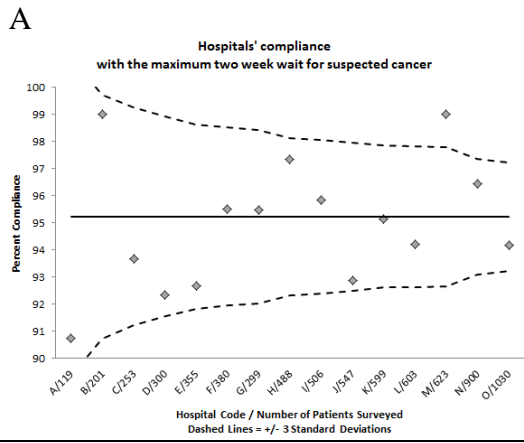
662

663 Appendix A. Graphs presenting between subjects comparisons.

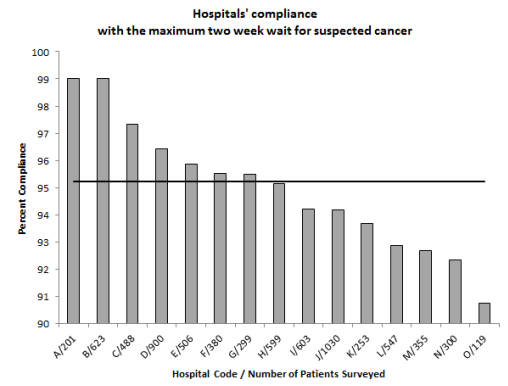
Funnel Chart

League Table

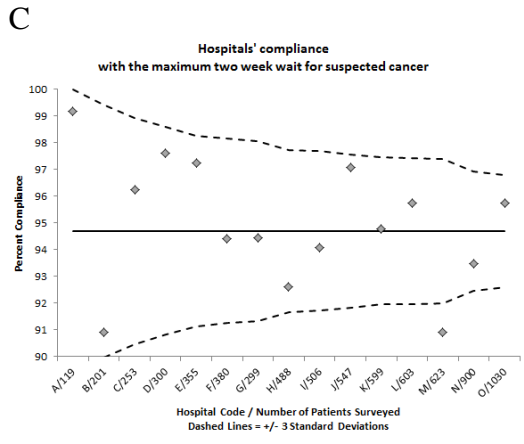
Data set with two equally high (potentially outlying) data points



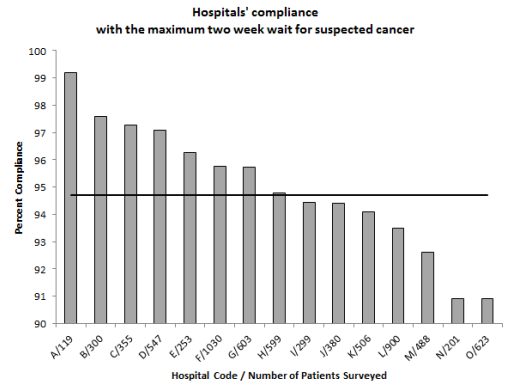
B



Data set with two equally low (potentially outlying) data points



D



664

665 Appendix B. Graphs presenting time-series comparisons.

