

Unidimensionality of the Strengths and Vulnerabilities Scales in the Short-Term Assessment of Risk and Treatability (START)

Richard Whittington, Charlotte Pollak, Alice Keski-Valkama, Andrew Brown, Alina Haines-Delmont, Jesper Bak, Jacob Hvidhjelm, Roger Almvik & Tom Palmstierna

To cite this article: Richard Whittington, Charlotte Pollak, Alice Keski-Valkama, Andrew Brown, Alina Haines-Delmont, Jesper Bak, Jacob Hvidhjelm, Roger Almvik & Tom Palmstierna (2021): Unidimensionality of the Strengths and Vulnerabilities Scales in the Short-Term Assessment of Risk and Treatability (START), International Journal of Forensic Mental Health, DOI: [10.1080/14999013.2021.1953193](https://doi.org/10.1080/14999013.2021.1953193)

To link to this article: <https://doi.org/10.1080/14999013.2021.1953193>



© 2021 The Author(s). Published with license by Taylor and Francis Group, LLC



[View supplementary material](#)



Published online: 02 Aug 2021.



[Submit your article to this journal](#)



[View related articles](#)



[View Crossmark data](#)

Unidimensionality of the Strengths and Vulnerabilities Scales in the Short-Term Assessment of Risk and Treatability (START)

Richard Whittington^{a,b,c}, Charlotte Pollak^{d,e}, Alice Keski-Valkama^f, Andrew Brown^{c,g}, Alina Haines-Delmont^{c,h,i}, Jesper Bak^j, Jacob Hvidhjelm^j, Roger Almvik^{a,b}, and Tom Palmstierna^{a,b,d}

^aForensic Department Brøset, St. Olav's University Hospital, Trondheim, Norway; ^bDepartment of Mental Health, Norwegian University of Science and Technology, Trondheim, Norway; ^cDepartment of Health Services Research, University of Liverpool, Liverpool, UK; ^dDepartment of Clinical Neuroscience, Centre for Psychiatry Research, Karolinska Institutet, Solna, Sweden; ^eStockholm Forensic Psychiatric Clinic, Region Stockholm, Sweden; ^fVanha Vaasa Hospital, Vaasa, Finland; ^gRowan View, Mersey Care NHS Foundation Trust, Liverpool, UK; ^hDepartment of Nursing, Manchester Metropolitan University, Manchester, UK; ⁱCentre for Perfect Care, Mersey Care NHS Foundation Trust, Liverpool, UK; ^jMental Health Centre Sct Hans, Clinical Mental Health and Nursing Research Unit, Roskilde, Denmark

ABSTRACT

The Short-Term Assessment of Risk and Treatability (START) is a 20-item structured professional judgment instrument for assessing dynamic risk in mental health services. Much of the START research literature examines the relationship between Strengths and Vulnerabilities sub-scale total scores and various adverse outcomes including violence. This assumes that the two sub-scales have the psychometric property of unidimensionality i.e. all the items cluster together as a measure of a single construct. Such assumed unidimensionality is a necessary condition for any analyses based on scale "total score" and the widespread use of scores summated in this way in research studies may obscure more specific clusters of items within each sub-scale. This multinational study examined START assessments ($n = 685$) conducted in four forensic services in Scandinavia and the UK using principal component analysis. It was found that all but three Strengths items (Substance Use, Social Support and Material Resources) and all but four Vulnerabilities items (Substance Use, Social Support, Material Resources and Self care) loaded >0.5 on the expected component. This indicates a unidimensional structure underlying the START and provides empirical support from a large multinational sample for the widespread use of summated Strengths and Vulnerabilities scores in forensic psychiatric risk research.

KEYWORDS


Risk assessment; START; psychometrics; dimensionality

Introduction

The Short Term Assessment of Risk and Treatability (START) (Webster et al., 2006) is a structured professional judgment (SPJ) instrument developed in the Canadian forensic mental health system in the early 2000s which has since been widely implemented in forensic and general mental health services in many countries (O'Shea & Dickens, 2014; Ramesh et al., 2018). It is concerned with the improvement of medium term risk management (i.e. over weeks to months) and is usually completed by clinicians based on interactions with the patient, consultation with colleagues and case note review. In some situations co-completion with the person being assessed has been implemented.

The START consists of 20 items covering a broad range of domains considered pertinent to mental health and risk including, for example, social support, treatment adherence and substance use. It stands out in the crowded field of risk assessment instruments for two particular reasons. Firstly, whilst there are many tools which guide decision making on violence and, to a lesser extent, self-harm/suicide (Carter et al., 2017; Viljoen et al., 2018), the START purports to provide information relating to four other negative outcomes beyond these i.e. substance misuse, self-neglect, unauthorized leave and victimization (Marriott et al., 2017). Secondly, each item is rated in terms of its relevance as both a strength and a vulnerability in the patient's profile. This emphasis on the importance

CONTACT Richard Whittington  richard.whittington@ntnu.no  Forensic Department Brøset, St. Olav's University Hospital, Trondheim 7446, Norway.

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/14999013.2021.1953193>.

© 2021 The Author(s). Published with license by Taylor and Francis Group, LLC

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

of individual strengths alongside vulnerabilities has contributed to a widespread acceptance that an overall risk estimate is more valid if it takes positive, protective factors into account (Robb  et al., 2013). These two features, alongside the comprehensiveness of the domains covered by the twenty items, has led to the popularity of the START in a large number of services (Nielsen et al. 2015; Singh et al., 2016).

Since its introduction in 2004 a solid research evaluation literature has developed supporting the use of the START in clinical services. A number of studies have examined its successful implementation (Kroppan et al., 2017) and established that it has adequate psychometric properties in terms of internal consistency and inter-rater reliability (Nicholls et al., 2006; Timmins et al., 2018). There is also evidence of good predictive validity in relation to violence for the separate Strengths and Vulnerabilities scales (Chu et al., 2011; O'Shea & Dickens, 2016). As a tool which consists entirely of dynamic risk factors, it has been used to examine associations between changes in risk and changes in violent outcomes (Whittington et al., 2014) and it has also been tested as an active intervention in a randomized controlled trial involving forensic outpatients (Troquete et al., 2013).

Most studies of predictive validity test the association of the total scores on the two START scales with violent outcomes but some more concise variations of the START have also been tested in this way including the accuracy of individual items (O'Shea & Dickens, 2015; Paetsch et al., 2019). In particular, Braithwaite et al. (2010) examined the predictive validity of the overall START (all twenty items) and compared this overall predictive validity with that for various shortened "optimized scales" related to violence and the other outcomes. They report better performance in this method by the optimized scales than the overall instrument. For example, four vulnerability items on their own (Mental state, Impulse control, External triggers and Conduct) were more highly associated with subsequent violence (OR 1.23) than the full scale of twenty items (OR 1.05). Equally six vulnerability items on their own (Emotional State, Impulse Control, External triggers, Attitudes, Rule Adherence, and Conduct) were more highly associated with subsequent victimization (OR 1.26) than the full scale of twenty items (OR 1.05) and three strength items (Impulse Control, Rule Adherence and Conduct) were better predictors (OR 0.72) of the avoidance of victimization (OR 0.97 for the full scale). This finding raises the possibility that one or more shortened versions of the START could be developed

which are easier to administer and might have improved validity for predicting the likelihood of the various outcomes when this is required. This in turn raises the theoretical question of whether the broad domain of risk captured by the START masks a number of underlying clusters which meaningfully constitute different components of the overall risk construct. These clusters would be sub-groups of items which associate with each other and dissociate from other sub-groups empirically.

This evidence of potential item clusters within the START highlights the absence of empirical evidence on the dimensionality of the START. Whilst the START manual advises against calculating total scores for clinical purposes (Webster et al., 2009), many, if not most, research studies on the START construct total Strengths and Vulnerabilities scores by summing across the twenty ratings in each domain and then conduct analyses on these total scores (e.g. Abidin et al., 2013; Hogan & Olver, 2018; Wilson et al., 2010). However, such an approach assumes that the START is a unidimensional scale and that it is meaningful to "add up" across all the contributing items in this way. Unidimensionality is an important attribute of a measurement instrument because, in psychometric terms, it indicates that there is a single latent trait (e.g. in this case "general risk") underlying the responses (Hattie, 1985) rather than several such traits which are conceptually incompatible and potentially irrelevant to the key trait. This assumption is wrong if the START Strengths and Vulnerabilities subscales are actually made up of item clusters which are distinct from each other and which relate to different aspects of a person's risk level. Unidimensionality should be examined in any psychological measurement tool to ensure the soundness of the assessments being made about the overall concept under consideration (Ziegler & Hagemann, 2015).

Unidimensionality is tested using factor analysis or related techniques and, in psychometric terms, only a finding that "all the items have substantial loadings on a single factor can be used to justify adding the item scores together to generate a single scale score" (Gardner, 1995). For comparison, it should be noted that the assumption of unidimensionality in various widely-used depression scales was unsupported in a recent investigation indicating that summated scores on such scales should not be interpreted theoretically as reflecting a single construct of depression (Fried et al., 2016). Measurement of the theoretical construct of "total risk" reflected by summated Strengths and

Vulnerabilities scores from the START would benefit from similar interrogation and clarification.

To the best of our knowledge there are no previous published studies examining the factor structure of the START and thus this commonly used unidimensional approach remains an assumption without evidence. There is an argument that structured professional judgment (SPJ) instruments such as the START are not psychometric tools at all and only the latter are specifically designed with the aim of measuring a single underlying theoretical construct using multiple items (Fayers & Hand, 2002). As an SPJ instrument, it is argued that the START more closely resembles a clinimetric (rather than a psychometric) tool. Such tools have a more practical purpose as “an index that is ‘clinically sensible’ and has desirable properties for prognosis or prediction” (Machin & Fayers, 2016, p. 53) in which case the abstract concept of unidimensionality is not relevant.

However, SPJ tool development relies heavily on a wide range of psychometric techniques to establish credibility in terms of, for example, inter-rater reliability, internal consistency and convergent validity (Nonstad et al., 2010; O’Shea & Dickens, 2014). Summated Strength and Vulnerability scores are regularly presented and discussed as if the items can be meaningfully combined presumably to represent high or low levels of an unobserved construct sometimes called “risk.” The dimensionality or factor structure of SPJ tools is therefore important regardless of whether such instruments have primarily a psychometric or clinimetric rationale. This is confirmed by, for example, a recent factor analysis of the Structured Assessment of Protective Factors for Violence (SAPROF) which indicated a 4 factor structure in contrast to the 3 subscales rationally derived by the instrument authors (Abbiati et al., 2020).

It is also true that a number of studies have demonstrated that START subscales have high internal consistency with estimates ranging from 0.80 to 0.95 for Strengths and from 0.76 to 0.95 for Vulnerabilities in O’Shea & Dickens’ review (2014). Whilst Hattie (1985) considers the main measure of internal consistency (Cronbach’s alpha) to be “suspect” as a measure of unidimensionality and O’Shea and Dickens (2014) note that it is not a direct metric of unidimensionality, the latter do argue that repeated internal consistency values at this level (≥ 0.80) from several studies are a good proxy measure of it. Nevertheless the average sample size of studies in this review was 60 indicating the need for further direct examination of unidimensionality in a large combined sample as

reported below. This study therefore sought to test the dimensionality of the START instrument in a large multinational forensic sample in order to establish whether the common research practice of summing total Strengths and Vulnerabilities scores is justifiable in psychometric terms.

Materials and methods

The START instrument (Webster et al., 2009)

The START consists of 20 items each of which is rated on an ordinal scale with values of 0 (not present), 1 (present to some extent) and 2 (fully present) according to the degree to which a factor is considered a feature of a specific patient’s current clinical profile. Each item is considered and scored in terms of its potential as both a risk factor (vulnerability) and as a protective factor (strength) in relation to the propensity to engage in 7 different types of adverse behavior (violence, self-harm, suicide, unauthorized leave, substance abuse, self-neglect, and exposure to victimization). It is designed for completion through consensus discussion amongst a clinical team but can be meaningfully completed by an individual practitioner. In many countries it is primarily completed by nursing staff based on personal knowledge of the patient amongst team members and appraisal of case notes and relates to the period since the last assessment. Assessment at least every 12 weeks and at most every week is recommended. Raters are required to possess a qualification in one of the recognized mental health professions and ideally should attend a training course. The instrument Manual provides extensive guidance on item descriptors and scoring anchors.

The EuroSTART dataset

This is an integrated standardized dataset of START assessments conducted as part of routine clinical practice in mental health services in Scandinavia and the UK. It has been constructed through collaboration between forensic mental health services in five countries with the aim of pooling data to increase statistical power and enable cross-national comparisons. Four high-security forensic mental health in-patient services have contributed all START ratings conducted as part of routine clinical practice over a specified time period: Forensic Psychiatric Clinic of Stockholm County, (FS), Sweden (168 beds, ~55 admissions per year); Sct. Hans Mental Health Center (SH), Roskilde, Denmark (104 beds, ~40 admissions per year); the Scott Clinic (SC), Merseyside, UK (66 beds, ~29

admissions per year); and Vanha Vaasa Hospital (VV), Vaasa, Finland (152 beds, ~91 admissions per year).

Three of these services (FS, SH and SC) are regional in scope serving catchment areas with a population of approximately 2 million people. The fourth service (VV) is a national hospital covering the whole of Finland with a population of 5.5 million. The four samples largely reflect the overall demographic and clinical profile of each service and are broadly comparable. The average admission duration (years) was as follows: VV: 7.0; FS: 4.9; SC: 2.3; SH: 2.2. The percentage of each sample that was male and the mean/median age was as follows: VV: 79%, 41 years; FS: 84%, 35 years; SC: 98%, 33 years; SH: 94%, 40 years. The most common diagnosis was schizophrenia or psychosis in all services and the most common legal decision governing compulsory treatment in all services was diminished responsibility, not guilty by reason of insanity or equivalent. The median number of days between admission to the service and the first START assessment varied substantially being 8 months in FS and 57 months in SH. Further information on the demographic and clinical characteristics of the samples is not available as such information was required to be removed to obtain ethical approval. The project is co-ordinated by the Brøset Center for Research and Education in Forensic Psychiatry, Trondheim, Norway.

The START had been implemented in each service for a number of years prior to data integration. The median number of assessments per patient were VV: 4; FS: 2; SC: 7; SH: 5. Assessments at VV were conducted at fixed time points every six months regardless of clinical condition but assessments in the other services were conducted as required and/or when staff resources were available. In SC and SH, some patients were assessed many times with a maximum of 29 assessments on one patient at SC and 22 assessments at SH. One fifth of patients at FS (22.6%) and SH (20.8%) had a single assessment compared to less than a tenth (8.6%) of patients at SC. Completion of the START in all cases was conducted by staff who had training based on the instrument Manual and who followed the protocol set out in that document as far as possible. Assessments were based on clinical documentation, multidisciplinary team consultation and, where possible, collaboration with the patient. If recorded on paper they were subsequently loaded in a digital format to a centralized secure drive run by the relevant service.

Each service obtained appropriate local ethical and research governance approval and exported anonymized START ratings into a standardized Excel spreadsheet. The four datasets were then merged and exported into SPSS v25 for analysis.

The overall dataset consists of 2890 START assessments but only one rating per patient was included in this analysis to avoid confounding through repeated assessments (Tabachnik & Fidell, 2007). The selected rating was usually the first assessment conducted on the patient during the study period. When the assessment date was unknown the first assessment listed in the dataset for that patient was selected. This may reflect the first assessment in time but may also be the result of how the data were loaded and sorted during data preparation. This created a sample of 685 patients with full ($n=593$) or partial ($n=92$) item completion on either the Strength or Vulnerabilities scales (VV: $n=112$, 16% of the overall sample; FS: $n=327$, 48%; SC: $n=112$, 16%; SH: $n=134$, 20%).

Statistical analysis

Categorical principal components analysis (CATPCA) (Linting et al., 2007) using variable principle normalization in the SPSS Dimension Reduction menu was used to examine the dimensionality of the START Strengths and Vulnerabilities ratings. This approach was chosen because of the ordinal nature of the three category response format for each item. Whilst factor analysis is routinely used with ordinal data in applied social sciences it can be unsuitable as it can generate erroneous factors (Dolan, 1994). The ordinal scale was selected as optimal for the SPSS procedure and all ratings were recoded from 0/1/2 to 1/2/3 as values of 0 are treated as missing by SPSS in this procedure (IBM Support, 2020). A stringent factor loading cut off >0.5 was set (Tabachnik & Fidell, 2007). There was a small amount of missing data (1.12% of observations). For all item-level analyses these were treated passively (Linting et al., 2007) in that the missing observation on a variable did not contribute to the analysis on that variable only. Subscale scores were not calculated for a case when missing data was present.

Results

Mean scores on individual items and the summed Strength and Vulnerabilities scale for each country are given in Table 1.

Variations between countries were highly statistically significant (chi-squared test, $p < .005$) for every

Table 1. Mean scores on individual items and the summed strength and vulnerabilities scale for each country.

	Denmark (SH: n = 134)		Finland (VW: n = 112)		Sweden (FS: n = 327)		UK (SC: n = 112)		Total (n = 685)	
	Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation	Mean	Std. Deviation
Strengths										
Social skills	0.89	0.57	1.08	0.59	0.90	0.55	1.28	0.54	0.99	0.58
Relationships	0.69	0.64	0.86	0.66	0.60	0.57	1.14	0.55	0.75	0.63
Occupational	0.64	0.72	0.88	0.75	0.45	0.60	1.13	0.67	0.67	0.71
Recreational	0.72	0.70	1.08	0.70	0.71	0.62	1.22	0.64	0.86	0.68
Self-care	1.02	0.74	1.23	0.68	1.22	0.66	1.35	0.61	1.20	0.68
Mental state	0.47	0.59	0.64	0.67	0.69	0.56	1.20	0.61	0.72	0.63
Emotional state	0.72	0.58	0.88	0.52	0.83	0.55	1.17	0.63	0.87	0.58
Substance use	1.09	0.84	1.23	0.88	1.04	0.86	1.34	0.75	1.13	0.85
Impulse control	0.84	0.76	0.95	0.71	0.85	0.64	1.17	0.67	0.91	0.69
External triggers	0.83	0.53	0.90	0.67	0.63	0.59	1.01	0.67	0.78	0.62
Social support	1.05	0.58	1.12	0.64	1.05	0.48	1.28	0.58	1.10	0.55
Material resources	1.16	0.63	1.14	0.77	0.94	0.65	1.30	0.63	1.07	0.68
Attitudes	0.65	0.59	0.80	0.63	0.63	0.59	1.16	0.63	0.75	0.63
Medical adherence	1.08	0.78	1.11	0.75	0.97	0.71	1.54	0.59	1.11	0.74
Rule adherence	1.07	0.75	1.31	0.68	1.23	0.61	1.28	0.74	1.22	0.68
Conduct	0.78	0.68	1.03	0.68	0.87	0.62	1.25	0.67	0.94	0.66
Insight	0.50	0.62	0.53	0.66	0.41	0.52	1.01	0.69	0.54	0.63
Plans	0.79	0.70	0.67	0.70	0.52	0.55	1.11	0.67	0.69	0.66
Coping	0.55	0.57	0.56	0.57	0.54	0.55	1.13	0.59	0.64	0.60
Treatability	0.77	0.66	0.89	0.70	0.70	0.54	1.24	0.60	0.83	0.63
Strengths subscale total*	15.97	8.58	18.95	9.36	15.72	6.73	24.92	7.46	17.55	8.32
Vulnerabilities										
Social skills	1.20	0.62	1.09	0.64	1.12	0.57	1.02	0.67	1.11	0.61
Relationships	1.05	0.67	0.96	0.64	1.07	0.64	1.26	0.67	1.08	0.66
Occupational	1.23	0.76	0.80	0.78	1.32	0.70	1.06	0.74	1.17	0.76
Recreational	1.23	0.75	0.88	0.67	1.16	0.64	0.95	0.69	1.10	0.69
Self-care	0.91	0.74	0.79	0.72	0.66	0.68	0.93	0.73	0.77	0.71
Mental state	1.31	0.68	1.22	0.71	1.11	0.62	1.21	0.72	1.18	0.67
Emotional state	1.06	0.67	1.20	0.60	1.14	0.58	1.14	0.66	1.13	0.61
Substance use	0.81	0.84	0.66	0.84	0.90	0.86	0.85	0.86	0.83	0.86
Impulse control	1.00	0.77	0.97	0.72	1.12	0.69	1.07	0.72	1.06	0.72
External triggers	1.14	0.74	0.96	0.70	1.20	0.66	1.13	0.70	1.14	0.69
Social support	1.07	0.68	1.06	0.66	1.18	0.62	1.12	0.69	1.13	0.65
Material resources	0.89	0.68	0.78	0.74	1.01	0.65	0.75	0.74	0.90	0.69
Attitudes	1.09	0.72	1.01	0.69	1.07	0.63	1.08	0.72	1.07	0.67
Medical adherence	0.68	0.77	0.50	0.67	0.71	0.71	0.79	0.79	0.68	0.73
Rule adherence	0.85	0.75	0.56	0.67	0.69	0.64	0.85	0.81	0.73	0.70
Conduct	0.70	0.77	0.54	0.66	0.67	0.65	0.81	0.72	0.67	0.69
Insight	1.41	0.65	1.46	0.64	1.54	0.54	1.26	0.68	1.46	0.61
Plans	1.09	0.72	1.20	0.78	1.37	0.64	1.12	0.67	1.25	0.70
Coping	1.40	0.60	1.43	0.60	1.37	0.58	1.19	0.73	1.39	0.62
Treatability	1.10	0.74	0.96	0.70	1.25	0.61	1.08	0.73	1.14	0.68
Vulnerabilities subscale total*	21.41	9.03	19.00	8.42	21.84	7.31	20.66	8.86	21.07	8.14

*Total n = 593 due to incomplete subscale item completion.

Strength item. The distributions of scores were more similar across the four countries for the Vulnerabilities scale but variation in all cross-national comparisons was statistically significant ($p < .05$) for all items apart from Emotional State, Substance Use, Impulse Control, Social Support and Attitudes. Both scale total scores also varied significantly between countries (Strengths: $F = 31.07$, $p < .001$; Vulnerabilities: $F = 3.43$, $p = .017$; $df = 3, 589$) with Strengths rated particularly highly in the UK sample. However, for Vulnerabilities, despite the statistical significance of the variation, no national sample varied by more than 10% away from the overall mean.

The results of the PCA are presented in [Table 2](#) for the combined sample from all four countries and in [Supplementary Table 1](#) for each country separately. In the overall sample, there was strong evidence of unidimensionality for both Strengths and Vulnerabilities. All Strengths items loaded >0.3 on Component 1 and all but three items (Substance Use, Social Support and Material Resources) loaded on this component at the required cutoff (>0.5). Two of the Component 1 non-loading items (Substance use and Material resources) loaded instead onto Component 2 and the third (Social support) loaded onto Component 3. The eigenvalue for Component 1 indicates that it explained 43% of the overall variance and the high Cronbach's alpha value indicates that Component 1 had high internal consistency. With regard to Vulnerabilities, again all items loaded >0.3 on Component 1 and all but four items loaded >0.5 . Three of these items were the same as the non-loading Strengths items but the distribution of loadings on the other components was slightly different for two of these three items. Social support did not load onto any component for Vulnerabilities and Material resources loaded onto Component 4 (Vulnerabilities) rather than Component 2 (Strengths). The eigenvalue and Cronbach's alpha for Component 1 Vulnerabilities was slightly lower (7.33, 0.91) than that for Strengths (8.60, 0.93) though remained high overall.

The same analysis conducted for each country individually was largely consistent with this unidimensional pattern ([Supplementary Table 1](#)). Unidimensionality was somewhat weaker in the Swedish sample especially for Vulnerabilities. Material resources and Substance use were the least well-fitted items and did not load on more than half of the individual country analyses.

Discussion

This is the first study to examine the unidimensionality of the START Strengths and Vulnerabilities scales

in a multinational sample in order to test the empirical justification for summed total scores. There is strong evidence here that both scales are indeed unidimensional with all but three or four items loading strongly onto a single component in both cases. This is preliminary evidence that a total Strengths or Vulnerabilities score as used by researchers is a meaningful entity. Thus the rational process used to select items for inclusion in the START as a clinimetric tool in the development stage has successfully produced a psychometrically robust pair of scales. Given this evidence of unidimensionality it appears that those making the ratings view risk or protective factors as a unified concept with few separate domains. A set of items here consistently did not fit with this unidimensional pattern for either Strengths or Vulnerabilities. Substance use in particular cross-loaded onto other potential components with one or two other items. This suggests perhaps the unique contribution of substance use to perceptions of risk, cutting across other risk domains as a general factor exacerbating the potential for poor outcomes in violence, self-harm and the other aspects to which the START purports to relate. The other two divergent items, Material resources and Social support, both clearly relate to the patient's external environment and factors which are relatively beyond their sphere of personal control. As such, they may be perceived by raters as contributing a special set of challenges for the patient which is independent of their internal world.

The unidimensionality demonstrated here does not in itself provide empirical support for the theoretical concept of risk sub-domains embedded in the overall START item list. However, this does not undermine the potential usefulness of the optimized scales proposed by Braithwaite et al. (2010). Again, the lack of evidence here for multiple dimensionality is a psychometric issue but the relatively high level of predictive validity demonstrated for the optimized scales in that study is evidence of the START's effectiveness as a clinimetric tool.

Whilst there is support here for the process of summation, a number of aspects of the clinical context must be considered. Firstly, despite the availability of quite specific rating guidelines in the manual, the complexity of the behavior being assessed still leaves much room for subjectivity in evaluations. For example the Substance Use item could be rated as a Strength (1 or 2) in a variety of ways: the patient has never used substances; the patient has used substances before hospitalization and has a substance use diagnosis, but cannot access drugs in the hospital because of

Table 2. Component loadings for strengths and vulnerabilities items (>0.5 highlighted in grey).

Dimension	Strengths				Vulnerabilities			
	1	2	3	4	1	2	3	4
Social skills	0.68	-0.29	-0.06	0.23	0.65	-0.36	-0.16	0.09
Relationships	0.73	-0.25	0.17	0.02	0.69	0.10	-0.15	0.02
Occupational	0.67	0.05	0.11	0.25	0.67	-0.19	0.16	0.35
Recreational	0.64	-0.06	-0.02	0.42	0.56	-0.45	0.08	0.36
Self-care	0.51	-0.14	-0.42	0.44	0.48	-0.30	-0.09	0.21
Mental state	0.68	-0.26	0.06	0.01	0.57	-0.30	-0.20	-0.23
Emotional state	0.66	-0.24	0.09	0.06	0.56	-0.18	-0.46	0.01
Substance use	0.43	0.65	-0.07	-0.06	0.38	0.60	0.14	0.27
Impulse control	0.73	0.13	-0.15	-0.13	0.72	0.22	-0.32	-0.08
External triggers	0.65	0.39	0.19	0.03	0.60	0.30	-0.29	0.13
Social support	0.46	0.02	0.66	-0.03	0.49	-0.02	0.34	0.14
Material resources	0.40	0.59	0.11	0.31	0.39	0.29	0.16	0.53
Attitudes	0.74	0.04	-0.07	-0.19	0.65	0.32	-0.03	-0.15
Medication adherence	0.66	0.07	-0.14	-0.26	0.60	0.14	0.21	-0.30
Rule adherence	0.68	0.14	-0.40	-0.23	0.73	0.25	0.10	-0.11
Conduct	0.78	0.04	-0.21	-0.09	0.69	0.22	-0.26	-0.15
Insight	0.71	-0.20	0.09	-0.32	0.61	-0.06	0.43	-0.34
Plans	0.70	-0.01	0.20	0.04	0.64	-0.17	0.34	-0.04
Coping	0.71	-0.07	-0.02	0.05	0.61	-0.16	-0.20	-0.06
Treatability	0.75	-0.11	0.02	-0.23	0.70	-0.14	0.34	-0.21
Cronbach's alpha*	0.93	0.25	0.03	-0.01	0.91	0.37	0.11	0.01
VAF (Eigenvalue)**	8.60	1.30	1.03	0.93	7.33	1.54	1.11	1.01

*Total: 0.96 Strengths, 0.96 Vulnerabilities.

**Total: 11.86 Strengths, 10.99 Vulnerabilities.

restrictive conditions; or the patient admits his or her urges, but does not use substances. All of these are valid responses but reflect very diverse clinical situations. This raises the question of consistent communication within and between clinical teams over the precise meaning of individual items and overall risk estimates. One way to enhance such communication is to provide comprehensive training for staff in making the risk formulation including risk-scenarios with related actions described together with the patients. In addition it should be noted that simple summation assumes equal weighting across all items when it is quite possible that some items are more important than others and so different weights should be attached to some items.

The focus throughout this report has been on the psychometric approach to the START. It has been noted that clinical staff are advised in the START manual to avoid summation in practice and to use the individual items to guide individualized planning and interventions with patients. They are encouraged however to make a specific risk estimate (low, medium or high) for each of the adverse outcomes (aggression etc.) and space for recording such estimates are prominent on the START summary sheet. As with all SPJ tools, these risk estimates are explicitly not intended to be direct numerical translations of the summated score but instead are expected to be more sophisticated reflective interpretations of the person's overall profile. Such formulations will take into account specific key risks or strengths particular to the person

being assessed and awareness of this individual profile plays a bigger role in guiding an estimate of specific risk than a total score. However, it is interesting to consider the degree to which a risk estimate of "high" in clinical practice relates to a high numerical score on the two subscales especially as there is little guidance on what constitutes high, medium or low categories in risk estimates. It is likely that in practice, the process of conducting an assessment which yields consistent scores of "2" across all START Vulnerabilities items and "0" on Strength items will create a presumption that a high risk estimate is appropriate. Indeed, high total scores have been found to be significantly associated with specific risk estimates for violence in several studies (O'Shea & Dickens, 2014). In that sense, the summated score is at least likely to be an influence on the risk estimate and the psychometric issue of unidimensionality is thus relevant in the clinical context as well as the research context.

The main strength of this study is the overall sample size which is more than double the requirement for meaningful PCA (Guadagnoli & Velicer, 1988; Tabachnik & Fidell, 2007). The reliance here on secondary data derived from routine clinical information has both positive and negative implications. The approach has high ecological validity as it clearly reflects the "real world" usage of the START instrument by clinicians as they go about their business of managing dangerous individuals. At the same time the reliability and validity of individual ratings is

questionable given the large number of different raters in diverse countries and the constraints imposed upon them when conducting risk assessments as part of a busy routine and heavy workload. The strength of the overall findings is also limited by identification of some variations in unidimensionality between the contributing countries and generalizability is restricted due to the use of the instrument here with an overwhelmingly male sample. The individual country samples are variable in size with some countries contributing fewer than 150 cases and one country contributing a much larger proportion than others. Conclusions about individual countries should therefore be made with extra caution and those drawn about the overall sample must be made with an awareness of the large contingent from Sweden. There was also substantial variation between countries in terms of the length of time between admission and the START assessment in the study. This should be noted when considering the results but the variation itself is not fundamentally relevant to the issue of the internal structure of the scale being addressed here.

This analysis of the EuroSTART dataset represents the first step in a potential research programme examining the psychometric properties of the START instrument. A number of additional analyses will be considered based on the current dataset and additional variables when they can be added. Exploratory and then confirmatory factor analysis may be conducted to enable comparison between this unidimensional PCA model and any multidimensional models which may be identified by EFA. Also item response theory (IRT) models may be used to examine the question of whether additional weightings should be allocated to one or more items when calculating the total score. It is desirable to add supplementary variables to the dataset when this is feasible. These variables include the specific risk estimates from the START assessments and the relevant outcomes in terms of adverse behavior. Whilst these aspects are universally available in the various contributing clinical services contributing to the project there are major challenges when integrating these into a combined dataset. The START itself is always recorded in a standardized structure which make it relatively straightforward to integrate but outcome data in particular is recorded very differently across services even within the same country. Addressing this inconsistency is a priority but will require some time to achieve.

The START evidence base continues to expand and the findings from this study suggest a number of

avenues for future research. In particular the formulation process by which assessors move from rating of individual items to a specific overall risk estimate as the key construct underpinning the prioritization of clinical decisions could be examined further. Tighter protocols for rating and replication of the approach adopted here with larger female samples and non-forensic samples would also be worthwhile.

In conclusion, the evidence here supports the use of summated START Strengths and Vulnerabilities scores for research purposes. It is therefore meaningful to conduct analyses using such total scores. However, notwithstanding this evidence for a single “general risk” construct underpinning START assessments, there may still be other clinimetric reasons for developing sub-scales or identifying clusters of items if these produce clinically useful findings.

Acknowledgements

Thanks to Dr. Ghitta Weizmann-Henelius for contributing to the early stages of this project and to Dr. Nutmeg Hallett for comments on a draft of the paper.

Conflict of interest

The authors have no conflicts of interest to report.

References

- Abbiati, M., Golay, P., Gasser, J., & Moulin, V. (2020). Protective factor assessments: What are we measuring?—An investigation of the internal validity of the structured assessment of protective factors for violence risk. *Criminal Justice and Behavior*, 47(4), 383–398. <https://doi.org/10.1177/0093854819901157>
- Abidin, Z., Davoren, M., Naughton, L., Gibbons, O., Nulty, A., & Kennedy, H. (2013). Susceptibility (risk and protective) factors for in-patient violence and self-harm: Prospective study of structured professional judgement instruments START and SAPROF, DUNDRUM-3 and DUNDRUM-4 in forensic mental health services. *BMC Psychiatry*, 13(1), 197. <https://doi.org/10.1186/1471-244X-13-197>
- Braithwaite, E., Charette, Y., Crocker, A. G., & Reyes, A. (2010). The predictive validity of clinical ratings of the short-term assessment of risk and treatability (START). *International Journal of Forensic Mental Health*, 9(4), 271–281. <https://doi.org/10.1080/14999013.2010.534378>
- Carter, G., Milner, A., McGill, K., Pirkis, J., Kapur, N., & Spittal, M. J. (2017). Predicting suicidal behaviours using clinical instruments: Systematic review and meta-analysis of positive predictive values for risk scales. *The British Journal of Psychiatry: The Journal of Mental Science*, 210(6), 387–395. <https://doi.org/10.1192/bjp.bp.116.182717>
- Chu, C. M., Thomas, S. D. M., Ogloff, J. R. P., & Daffern, M. (2011). The predictive validity of the short-term

- assessment of risk and treatability (START) in a secure forensic hospital: Risk factors and strengths. *International Journal of Forensic Mental Health*, 10(4), 337–345. <https://doi.org/10.1080/14999013.2011.629715>
- Dolan, C. V. (1994). Factor analysis of variables with 2, 3, 5 and 7 response categories: A comparison of categorical variable estimators using simulated data. *British Journal of Mathematical and Statistical Psychology*, 47(2), 309–326. <https://doi.org/10.1111/j.2044-8317.1994.tb01039.x>
- Fayers, P., & Hand, D. (2002). Causal variables, indicator variables and measurement scales: An example from quality of life. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 165(2), 233–261. <https://doi.org/10.1111/1467-985X.02020>
- Fried, E. I., van Borkulo, C. D., Epskamp, S., Schoevers, R. A., Tuerlinckx, F., & Borsboom, D. (2016). Measuring depression over time . . . Or not? Lack of unidimensionality and longitudinal measurement invariance in four common rating scales of depression. *Psychological Assessment*, 28(11), 1354–1367. <https://doi.org/10.1037/pas0000275>
- Gardner, P. L. (1995). Measuring attitudes to science: Unidimensionality and internal consistency revisited. *Research in Science Education*, 25(3), 283–289. <https://doi.org/10.1007/BF02357402>
- Guadagnoli, E., & Velicer, W. F. (1988). Relation of sample size to the stability of component patterns. *Psychological Bulletin*, 103(2), 265–275. <https://doi.org/10.1037/0033-2909.103.2.265>
- Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement*, 9(2), 139–164. <https://doi.org/10.1177/014662168500900204>
- Hogan, N. R., & Olver, M. E. (2018). A prospective examination of the predictive validity of five structured instruments for inpatient violence in a secure forensic hospital. *International Journal of Forensic Mental Health*, 17(2), 122–132. <https://doi.org/10.1080/14999013.2018.1431339>
- IBM Support. (2020). *CATPCA error message states some variables have 0 variance*. <https://www.ibm.com/support/pages/catpca-error-message-states-some-variables-have-0-variance>
- Kroppan, E., Nonstad, K., Iversen, R. B., & Sondenaa, E. (2017). Implementation of the short-term assessment of risk and treatability over two phases. *Journal of Multidisciplinary Healthcare*, 10, 321–326. <https://doi.org/10.2147/JMDH.S133514>
- Linting, M., Meulman, J. J., Groenen, P. J. F., & van der Kooij, A. J. (2007). Nonlinear principal components analysis: Introduction and application. *Psychological Methods*, 12(3), 336–358. <https://doi.org/10.1037/1082-989X.12.3.336>
- Machin, D., & Fayers, P. (2016). *Quality of life: The assessment, analysis and reporting of patient-reported outcomes* (3rd ed.). John Wiley.
- Marriott, R., O'Shea, L. E., Picchioni, M. M., & Dickens, G. L. (2017). Predictive validity of the short-term assessment of risk and treatability (START) for multiple adverse outcomes: The effect of diagnosis. *Psychiatry Research*, 256, 435–443. <https://doi.org/10.1016/j.psychres.2017.07.009>
- Nicholls, T., Brink, J., Desmarais, S., Webster, C., & Martin, M. (2006). The short-term assessment of risk and treatability (START). A prospective validation study in a forensic psychiatric sample. *Assessment*, 13(3), 313–327. <https://doi.org/10.1177/1073191106290559>
- Nielsen, L. H., Mastrigt, S. v., Otto, R. K., Seewald, K., Ruiter, C. d., Rettenberger, M., Reeves, K. A., Rebocho, M. F., Pham, T. H., Mei Yee Ho, R., Grann, M., Godoy-Cervera, V., Folino, J. O., Doyle, M., Desmarais, S. L., Condemarin, C., Arbach-Lucioni, K., & Singh, J. P. (2015). Violence risk assessment practices in Denmark: A multidisciplinary national survey. *Scandinavian Journal of Forensic Science*, 21(2), 103–110. <https://doi.org/10.1515/sjfs-2015-0003>
- Nonstad, K., Nettet, M. B., Kroppan, E., Pedersen, T. W., Nøttestad, J. A., Almvik, R., & Palmstierna, T. (2010). Predictive validity and other psychometric properties of the short-term assessment of risk and treatability (START) in a Norwegian high secure hospital. *International Journal of Forensic Mental Health*, 9(4), 294–299. <https://doi.org/10.1080/14999013.2010.534958>
- O'Shea, L. E., & Dickens, G. L. (2014). Short-term assessment of risk and treatability (START): Systematic review and meta-analysis. *Psychological Assessment*, 26(3), 990–1002. <https://doi.org/10.1037/a0036794>
- O'Shea, L. E., & Dickens, G. L. (2015). Predictive validity of the short-term assessment of risk and treatability (START) for aggression and self-harm in a secure mental health service: Gender differences. *International Journal of Forensic Mental Health*, 14(2), 132–146. <https://doi.org/10.1080/14999013.2015.1033112>
- O'Shea, L. E., & Dickens, G. L. (2016). Performance of protective factors assessment in risk prediction for adults: Systematic review and meta-analysis. *Clinical Psychology: Science and Practice*, 23(2), 126–138. <https://doi.org/10.1111/cpsp.12146>
- Paetsch, A., van Os, T. W. D. P., Troquete, N. A. C., & van den Brink, R. H. S. (2019). Single-item predictive validity of the short-term assessment of risk and treatability (START) for violent behaviour in outpatient forensic psychiatry. *The Journal of Forensic Psychiatry & Psychology*, 30(4), 630–641. <https://doi.org/10.1080/14789949.2019.1581831>
- Ramesh, T., Igoumenou, A., Vazquez Montes, M., & Fazel, S. (2018). Use of risk assessment instruments to predict violence in forensic psychiatric hospitals: A systematic review and meta-analysis. *European Psychiatry: The Journal of the Association of European Psychiatrists*, 52, 47–53. <https://doi.org/10.1016/j.eurpsy.2018.02.007>
- Robb e, M. d. V., Vogel, V. d., & Douglas, K. S. (2013). Risk factors and protective factors: A two-sided dynamic approach to violence risk assessment. *Journal of Forensic Psychiatry & Psychology*, 24(4), 440–457. <https://doi.org/10.1080/14789949.2013.818162>
- Singh, J., Desmarais, S. L., Otto, R., Nicholls, T., Petersen, K., & Protchard, M. (2016). The international risk survey: Use and perceived utility of structured violence risk assessment tools in 44 countries. In J. Singh, S. Bjorkly, & S. Fazel (Eds.), *International perspectives on violence risk assessment* (pp. 101–126). Oxford University Press.
- Tabachnik, B., & Fidell, L. (2007). *Using multivariate statistics* (5th ed.). Pearson.
- Timmins, K. L. E., Evans, L., & Tully, R. J. (2018). Interrater reliability of the short-term assessment of risk and treatability (START). *The Journal of Forensic Psychiatry*

- Psychology*, 29(6), 968–988. <https://doi.org/10.1080/14789949.2018.1523945>
- Troquete, N. A. C., van den Brink, R. H. S., Beintema, H., Mulder, T., van Os, T. W. D. P., Schoevers, R. A., & Wiersma, D. (2013). Risk assessment and shared care planning in out-patient forensic psychiatry: Cluster randomised controlled trial. *British Journal of Psychiatry*, 202(5), 365–371. <https://doi.org/10.1192/bjp.bp.112.113043>
- Viljoen, J. L., Cochrane, D. M., & Jonnson, M. R. (2018). Do risk assessment tools help manage and reduce risk of violence and reoffending? A systematic review. *Law and Human Behavior*, 42(3), 181–214. <https://doi.org/10.1037/lhb0000280>
- Webster, C., Nicholls, T., Martin, M.-L., Desmarais, S., & Brink, J. (2006). Short-term assessment of risk and treatability (START): The case for a new structured professional judgment scheme. *Behavioral Sciences & the Law*, 24(6), 747–766. <https://doi.org/10.1002/bsl.737>
- Webster, C. D., Martin, M., Brink, J., Nicholls, T. L., & Desmarais, S. L. (2009). *Manual for the short-term assessment of risk and treatability (START)* (version 1.1).
- Whittington, R., Bjørngaard, J. H., Brown, A., Nathan, R., Noblett, S., & Quinn, B. (2014). Dynamic relationship between multiple START assessments and violent incidents over time: A prospective cohort study. *BMC Psychiatry*, 14, 323. <https://doi.org/10.1186/s12888-014-0323-7>
- Wilson, C. M., Desmarais, S. L., Nicholls, T. L., & Brink, J. (2010). The role of client strengths in assessments of violence risk using the short-term assessment of risk and treatability (START). *International Journal of Forensic Mental Health*, 9(4), 282–293. <https://doi.org/10.1080/14999013.2010.534694>
- Ziegler, M., & Hagemann, D. (2015). Testing the unidimensionality of items. *European Journal of Psychological Assessment*, 31(4), 231–237. <https://doi.org/10.1027/1015-5759/a000309>