

Published in final edited form as:

Int J Forensic Ment Health. 2017 ; 16(1): 16–17. doi:10.1080/14999013.2016.1261965.

Response to “The Use of Meta-Analysis to Compare and Select Offender Risk Instruments”

Seena Fazel

Department of Psychiatry, University of Oxford, Oxford, United Kingdom

A complex meta-analysis often addresses more than one principal question, which may be published in two or three separate papers. The Singh 2011 review compares risk assessment instruments and asks the question whether there are differences in their predictive accuracy using multiple performance statistics (Singh, Grann, & Fazel, 2011). It relies on how the research literature has used these instruments, which likely reflects the intended population and practical considerations. However, there were two other key questions that we tried to address when we reviewed risk assessment tools, which required separate publications. In a 2012 publication, using mostly the same data but with updated information, we stratified the performance of these instruments according to whether they focused on violent, sexual, and general criminal offending separately (Fazel, Singh, Doll, & Grann, 2012). So a key criticism of the current paper—that stratification by type of offender was not done—has been addressed to a large extent in the 2012 review. Furthermore, the current article conflates the aims of the 2011 and 2012 reviews. The 2011 review was not primarily intended to examine the overall accuracy of risk assessment tools. Rather that was the main question of the 2012 review according to whether the particular instrument was focused on the assessment of violent, sexual, and criminal recidivism.

One of the novel features of the 2011 Singh review was that it used six performance statistics to examine the comparative accuracy of violence risk assessment instruments (Singh, Grann, & Fazel, 2011). This is in contrast to another review of smaller review of solely head-to-head studies, which included in the range of 6348–7221 individuals from 28 primary studies (Yang, Wong, & Coid, 2010). This compares with 25,980 persons based on 68 studies analyzed in our 2011 review. The Yang review used Cohen’s d and ROC AUC (converted into Cohen’s d) (Yang, Wong, & Coid, 2010). But the problem with the latter approach is that ROC AUC’s are notoriously insensitive to changes in model performance (Pepe, Kerr, Longton, & Wang, 2013), and do not take account of very different numbers of false negative and false positive predictions resulting from different shapes of ROC curve that have the same ROC AUC (Mallett, Halligan, Thompson, Collins, & Altman, 2012). Thus, it is not surprising that ROC AUCs of the risk assessment tools are found to be similar, but this may mask very different consequences of alternative tools due to different

CONTACT Seena Fazel, seena.fazel@psych.ox.ac.uk, Department of Psychiatry, Oxford University, Warneford Hospital, Oxford OX3 7JX, United Kingdom.

Declaration of interest

Dr. Fazel is part of a research team that developed a violence risk assessment tool for released prisoners, OxRec, that was published in 2016 and is currently researching scalable approaches to risk assessment in psychiatric patients.

proportions and absolute numbers of false negative and false positive results, depending on prevalence of reoffending (Mallett, Halligan, Thompson, Collins, & Altman, 2012). Hence, guidelines have recommended for some years that other measures such as sensitivity, specificity, PPV, and NPV are also reported, in addition to measures that investigate model calibration. Thus, the fact that the Singh 2011 review has conclusions that are dissimilar to other reviews is partly because it used a broader and more clinically informative set of prediction statistics. Another reason is that we included considerably more data than previous reviews (much of which was unpublished and received directly from primary study authors). In fact, 61% of the participant data ($n = 15,775$) in our 2011 review was specifically obtained from study authors. Thus, reanalyses of the 2011 review, like the current piece attempts to do, but without access to these data, will be limited and potentially misleading.

The current piece discusses the fact that overlapping confidence intervals and inter-quartile ranges (IQRs) suggest differences between tools are not significant but we clearly explain in our 2011 review that our overall ranking of risk assessment tools was not based on one or two statistics but on the cumulative ranking of 6 performance measures.

The current piece suggests that dichotomisation into risk categories is problematic as it may not be used in practice for some of the included instruments. This may be the case in some jurisdictions at some times, but is not entirely relevant to our review — in order to calculate certain statistics on tool performance, it was necessary to convert scores into categories, which enabled a wider range of statistics to be included in the review. In addition, many structured clinical judgement tools are used in a binary fashion, and a recent violence risk tool for released prisoners (OxRec) provides both categorical and probabilistic scores (Fazel et al., 2016). We have suggested that OxRec provides a model for how to present research findings in violence risk assessment—pre-specifying the protocol, using a large representative sample for derivation and validation, focusing on external validation in the reporting, presenting both measures of discrimination and calibration, translating the findings into a risk calculator that is free to use, online, and requiring no specific training.

The third article arising from our meta-analysis investigated authorship effects in this literature, and found some evidence for such effects in validation studies (Singh, Grann, & Fazel, 2013). We found potential conflict of interests common in the risk assessment literature and little transparency in whether they were declared despite journal policies. In fact, none of the 25 studies where a tool designer or translator was one of the authors of a validity study disclosed a financial or non-financial conflict interest, despite many of the journals in which these studies appeared explicitly requesting such information. We note that the current piece states, “We have never advocated that assessors predict that an offender will or will not recidivate.” We also note that that actuarial tools such as the LSI-R performed worse based on the performance measures of the nine included tools in our review (see Table 6 in the 2011 article) (Singh, Grann, & Fazel, 2011). In terms of research recommendations, we think this is an area that needs addressing in addition to the transparent reporting of a full range of performance statistics (Fazel & Bjorkly, 2016).

Funding

The Wellcome Trust (grant number 095806).

References

- Fazel, S., Bjorkly, S. Methodological considerations in risk assessment research. *International Perspectives on Violence Risk Assessment*. Singh, S., Bjorkly, S., Fazel, S., editors. New York, NY: Oxford University Press; 2016.
- Fazel S, Chang Z, Fanshawe T, Långström N, Lichtenstein P, Larsson H, Mallett S. Prediction of violent reoffending on release from prison: derivation and external validation of a scalable tool. *Lancet Psychiatry*. 2016; 3(6):535–543. DOI: 10.1016/S2215-0366(16)00103-6 [PubMed: 27086134]
- Fazel S, Singh JP, Doll H, Grann M. Use of risk assessment instruments to predict violence and antisocial behaviour in 73 samples involving 24 827 people: systematic review and meta-analysis. *BMJ: British Medical Journal*. 2012; 345:e4692. [PubMed: 22833604]
- Mallett S, Halligan S, Thompson M, Collins GS, Alt-man DG. Interpreting diagnostic accuracy studies for patient care. *BMJ: British Medical Journal*. 2012; 345:e3999. [PubMed: 22750423]
- Pepe MS, Kerr KF, Longton G, Wang Z. Testing for improvement in prediction model performance. *Statistics in Medicine*. 2013; 32(9):1467–1482. [PubMed: 23296397]
- Singh JP, Grann M, Fazel S. A comparative study of violence risk assessment tools: A systematic review and metaregression analysis of 68 studies involving 25,980 participants. *Clinical Psychology Review*. 2011; 31(3):499–513. [PubMed: 21255891]
- Singh JP, Grann M, Fazel S. Authorship bias in violence risk assessment? A systematic review and meta-analysis. *PLoS ONE*. 2013; 8(9):e72484. doi: 10.1371/journal.pone.0072484 [PubMed: 24023744]
- Yang M, Wong SC, Coid J. The efficacy of violence prediction: a meta-analytic comparison of nine risk assessment tools. *Psychological Bulletin*. 2010; 136(5):740–767. [PubMed: 20804235]