Letter re: stratification by quality is not recommended in meta-analysis

We read with interest the letter by Page et al. [1] who disagree with the conclusions of our letter, “Stratification by quality induced selection bias in a meta-analysis of clinical trials” [2]. Page et al. [1] critique the directed acyclic graph (DAG) we presented and suggest that conditioning on quality stratum does not imply that the results within strata defined by quality ranking are biased. We believe that bias occurs when studies are selected into strata such that systematic differences between larger and smaller studies emerge (where none initially existed), as shown in the funnel or Doi plot (Figures 2–4 in our published article [2]). The association develops solely because of selection of studies into a quality stratum and does not imply that all the results within strata defined by quality ranking are biased, and this is a situation akin to publication bias [3,4]. Page et al. next state that within stratum bias would only be the case if results in smaller studies were biased and less precise. On the contrary, selection bias would occur through conditioning on a quality stratum if studies selected into the stratum are either biased or less precise (not necessarily both) as in our DAG [2].

The alternative DAG Page et al. [1] present depicts methodological expertise as a common cause (confounder) of study size and quality stratum. As this DAG depicts methodological expertise as a confounding variable, assuming the latter is true, they make the point that conditioning on quality stratum can then remove the association between biased results and precision. However, we note two problems with the DAG. First, although methodological expertise could be a parent of quality stratum, the latter cannot be a parent of biased results, as higher quality studies are not necessarily unbiased, and lower quality studies are not necessarily biased [5]. Moreover, the DAG represents a causal diagram, and an assessment of quality stratum membership cannot be a cause of biased results because such assessments occur ex post facto, and therefore, the cause does not precede the effect in this instance. Furthermore, as depicted in the DAG in Fig. 1A, methodological expertise as a parent of study size is an unlikely link, as lack of expertise could lead to study samples that are too small or too large. We are not aware of any evidence to justify this link. Second, the DAG in Panel C adds a causal link from study size to biased results. This seems improbable; small studies are not necessarily biased but tend to belong to lower quality strata for other reasons [6–8].

Page et al. [1] then move on to state that “Quality and risk of bias are very different concepts, though Stone et al. do not distinguish between them clearly.” When we discuss quality in this article, we are referring to methodological quality, and we do not view reporting items or reporting checklists as quality items or scales, respectively. Quality assessment is the process through which methodological safeguards against bias are assessed in published research, whereas risk of bias assessment is the judgment based on the quality assessment (high, unclear, or low risk of bias). Many creators of quality scales do include reporting items in quality assessment scales for various reasons, one of which may be expertise related, and this is evident in the quality scales included in Jüni et al.’s [9] article. This does not mean that a quality scale ceases to exist or that quality scales automatically are reporting checklists. This also does not mean that quality scales that contain reporting items are not used (albeit incorrectly) to make risk of bias judgments nor that the study by Jüni et al. is not worth replicating to further explore the reasons for their findings. In epidemiology, it is well known that many terms apply to the same concepts, and researchers need to be able to discern from the context of the research what is actually meant by the terms the authors use. We believe it is misleading of Page et al. to refer to reporting checklists as “quality scales” and to refer to “tools” as assessment of “risk of bias.”

Page et al. [1] have stated that we calculated a composite quality score incorporating some reporting quality items across the 25 scales. This statement by Page et al. is incorrect as we clearly stated in our article that “Across all scales, we only included items that focus on the internal validity of the clinical trial rather than the quality of reporting of an article included in this analysis” (p. 53) and have provided the composite scale for readers to view in the supplementary material. In this article, we reproduced the work by Jüni et al. but, in addition, eliminated reporting items (based on the wording used in the article) but stuck to the same scales to retain comparability to Jüni et al. We were not promoting the use of our composite quality scale by readers (see Stone et al. [10] for our empirically supported quality scale) and were instead using the composite scale provided in this article for agreement, which is stated in

DOI of original article: https://doi.org/10.1016/j.jclinepi.2019.06.003.
Conflict of interest: The authors declare no other conflicts of interest.
Funding Statement: Australian National University Higher Degree by Research Scholarship and Erasmus + program (Radboud University) scholarship (J.C.S.).
In our article, “In doing so, the composite scale can be used to judge the agreement between each of the 25 scales and the composite, and to show that this agreement does not influence the estimates of effect” (p. 53). It also follows that the statement by Page et al. that the association between precision and effect size may be because of reporting quality items included within the quality scales is also incorrect, given the information provided previously. In addition, we ran a conditional logistic regression to see if this relationship existed independent of the scale used by conditioning on quality rank, thereby allowing us to focus on quality stratification per se. This was explained in our article in both the methods and results sections.

Based on our findings, we certainly recommend that “The common strategy of assessing the impact of quality in meta-analysis by excluding lower or including higher quality studies should be abandoned” (p. 52) [1]. We are unclear what Page et al. [1] mean by “a single old case study,” and we conclude that the views put forward by Page et al. do not support any change to the position in our article. We reaffirm that our article does question the use of restriction or stratification according to quality assessments in meta-analysis.

Jennifer Stone  
Department of Health Services Research and Policy  
Research School of Population Health  
Australian National University  
Canberra, Australian Capital Territory, Australia

Usha Gurunathan  
Department of Anaesthesia  
The Prince Charles Hospital  
Brisbane, Queensland, Australia

School of Population Health  
University of Queensland  
Brisbane, Queensland, Australia

Kathryn Glass  
National Centre for Epidemiology and Population Health  
Research School of Population Health  
Australian National University  
Canberra, Australian Capital Territory, Australia

Zachary Munn  
The Joanna Briggs Institute  
The University of Adelaide  
Adelaide, South Australia, Australia
References


https://doi.org/10.1016/j.jclinepi.2019.06.004