The problem of mixing ‘apples and oranges’ in meta-analytic studies

Sandro C. Esteves¹, Ahmad Majzoub², Ashok Agarwal³

¹ANDROFERT, Andrology and Human Reproduction Clinic, Referral Center for Male Reproduction, Campinas, SP, Brazil; ²Department of Urology, Hamad Medical Corporation, Doha, Qatar; ³American Center for Reproductive Medicine, Cleveland Clinic, Cleveland, OH, USA

Correspondence to: Sandro C. Esteves. Director, ANDROFERT—Andrology & Human Reproduction Clinic, and Professor, Division of Urology, UNICAMP, Av. Dr. Heitor Penteado 1464, Campinas, SP, 13075-460, Brazil. Email: s.esteves@androfert.com.br.


Submitted Feb 04, 2017. Accepted for publication Feb 04, 2017.
doi: 10.21037/tau.2017.03.23

View this article at: http://dx.doi.org/10.21037/tau.2017.03.23

Dr. Pool in his commentary (1) responding to the practice recommendations for Sperm DNA Fragmentation (SDF) testing based on clinical scenarios by Agarwal et al. (2) critically appraised the applicability of such testing to couples embarking in Assisted Reproductive Technology (ART).

As noted by the author, conflicting views have emerged regarding the predictive value of SDF for pregnancy according to various meta-analysis. In our reply, we add to the discussion by explaining some possible reasons for these discrepancies.

No one can deny that meta-analyses have benefits over conventional reviews in that all data is combined and presented. In this process, however, differences between individual studies may be lost in an attempt to aggregate enough data for analysis. In fact, any meta-analysis is subjected to the mix-up of ‘oranges and apples’ due to the heterogeneity of the included studies. Therefore, application of rigorous quality control methods is crucial (3). Failure to do so may result in erroneous conclusions presented by the author or drawn by the reader.

Notably, the literature is rich in meta-analyses exploring the clinical utility of SDF testing in ART. The results are conflicting, with studies indicating an association between high SDF rates and low pregnancy rates in IVF, ICSI, or both, and others suggesting a limited capacity of SDF testing to predict the outcomes of ART [reviewed by Agarwal (4)]. To examine this fact, we discuss further the most recent study published (5) on the matter concerned, which was detailed by Pool. In this study, the authors suggested that the current SDF tests have a poor capacity to predict the chance of pregnancy in ART. However, there was high statistical heterogeneity among the included studies even when they were grouped by the same method for SDF assessment, and meta-regression was applied to examine the influence of the type of treatment, i.e., IVF or ICSI (5).

Statistical heterogeneity means that the estimated effects were quite different across studies, thus adding uncertainty to the results. Therefore, authors of meta-analyses with high heterogeneity should perform subgroup and sensitivity analyses in an attempt to determine the source of variation across studies. Subgroup analysis is aimed at assessing whether the effect is similar across specified groups of patients or modified by certain patient characteristics (6). Sensitivity analysis comprises a series of analyses using the dataset to evaluate whether altering any of the assumptions made leads to different results (6). There are several examples of studies where sensitivity and subgroup analyses were utilized (7-9), but unfortunately, their use is very low (6). In the study mentioned above, sensitivity analyses were not performed, thus leaving the reader unaware about the possible reasons for heterogeneity (5). Notably, the authors reported that among the testing methods TUNEL showed the best accuracy to predict pregnancy in ART (0.71; 95% CI, 0.66–0.76). On the contrary, the accuracy of SCSA and SCD was low. Not surprisingly, the heterogeneity seen in the TUNEL meta-analysis was very low ($I^2=0\%$), in contrast to the high heterogeneity ($I^2>50\%$) observed in both SCSA and SCD meta-analyses. It means that when there was low
variation across the included studies—as in TUNEL—the predictive accuracy of SDF was higher. On the contrary, when the variation was high the accuracy was low, thus suggesting that the effect size might have been diluted by heterogeneity among studies.

As stated by Pool, to which we concurred and discussed elsewhere (10,11), it is the nature of the DNA fragmentation process and the inherent characteristics of each SDF testing method that complicate matters. Therefore, any meta-analysis on SDF should take into account the various quality control measures (e.g., publication bias, sensitivity analysis, subgroup analysis) to avoid misinterpretation of data. It is important to remember that readers may not be familiar with the interpretation of the statistical techniques used in meta-analyses. Therefore, extreme caution should be taken when mixing ‘apples and oranges’. The general assumption that meta-analyses provide better evidence than high-quality individual trials may be misleading.

Acknowledgements

None.

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

References


Cite this article as: Esteves SC, Majzoub A, Agarwal A. The problem of mixing ‘apples and oranges’ in meta-analytic studies. Transl Androl Urol 2017;6(Suppl 4):S412-S413. doi: 10.21037/tau.2017.03.23