Understand low-quality evidence: learn from food chains

Bioaccumulation refers to the accumulation process of toxic substances through food chains. Species on top of food chains are the most severely affected ones. Such a principle is also applicable in evidence-based medicine. Evidence from systematic reviews is considered the top of the pyramid in evidence-based medicine. However, the quality of the evidence from systematic reviews relies on the quality of evidence from the primary randomized controlled trials (RCTs). Bias and other concerns in the primary RCTs, as well as the evidence synthesis process, will affect the validity of results in subsequent systematic reviews.

In this issue of Fertility & Sterility, Samy and colleagues (1) reported a systematic review and network meta-analysis comparing different interventions to reduce blood loss during open and minimal invasive myomectomy. The authors included 26 RCTs reporting on 1,627 women undergoing open, laparoscopic or hysteroscopic myomectomy. They acknowledged that the overall quality of evidence was low and that therefore these results should be interpreted with caution. Still, the authors concluded that oxytocin, ornipressin, misoprostol, bupivacaine plus epinephrine, and vasopressin are effective in reducing blood loss during minimal invasive myomectomy, while vasopressin plus misoprostol, oxytocin, tranexamic acid, and misoprostol are effective in reducing blood loss during open myomectomy. Though the authors wrote a nice review for which they should be commended, in this commentary we argue that the foundation of the results might be too weak, and the differences found might be too small to justify this conclusion.

Some benefits and concerns of network meta-analysis

To reduce hemorrhage during myomectomy procedures, multiple treatment options are available, including uterotonic such as misoprostol and oxytocin, antifibrinolytic agents such as tranexamic acid, vasopressin, epinephrine and mechanical methods. Pairwise meta-analysis of RCTs is not optimal in this case as it only allows the comparison of two interventions at a time. Furthermore, not all these interventions have been compared directly with each other in RCTs. Network meta-analysis extends existing evidence synthesis methods to a new level by allowing the comparisons of multiple interventions simultaneously with the use of both direct and indirect evidence (2). Therefore, it is considered the next generation evidence synthesis toolkit which, when properly applied, can serve decision-making better (3). But for any meta-analysis the quality of the result will depend on the input.

When looking into RCTs in our field, we find that the quality is often just not good enough and that the effect is sometimes beyond what is expected as a chance finding (4). Meta-analyses intend to provide better estimates of the true difference and are used by policy makers. Clearly, we should be worried when the output is based on low-quality RCTs. In a direct comparison, the impact of low quality RCTs is much easier to study and understand than that in a network meta-analysis. How valid will indirect evidence be for interventions that have not been directly compared when this is deduced from low quality RCTs? As indirect evidence is not generated from RCTs, it is essentially based on observational data (3). One way to evaluate the validity of indirect evidence is to check the transitivity assumption, by comparing important factors that may affect the relative effects (i.e. effect modifiers) in different sets of RCTs included in a network analysis (5). In the network meta-analysis by Samy and co-authors (1), the size of myoma and other baseline factors are likely to be effect modifiers. However, the data was not available for all RCTs such that the assumption of transitivity could not be assessed. Another way to evaluate the indirect evidence is to check for inconsistency (incoherence) by comparing indirect evidence with direct evidence (3). Inconsistency refers to a conflict between direct and indirect evidence for a certain comparison. In the present network meta-analysis, inconsistency could not be evaluated as there was no closed loop in the network plots. This means that for each comparison the source of evidence was either from direct comparison or indirect comparison and therefore it was impossible to judge the inconsistency.

Overall the authors considered the quality of the evidence low on basis of the Grading of Recommendations Assessment, Development and Evaluation (GRADE) criteria. The authors presented high heterogeneity for open laparoscopy network but did not consider it in the GRADE assessment. Accounting for this, should the overall certainty of evidence maybe have been very low?

Is the evidence useful in clinical practice?

Neglecting the previous text and assuming that the estimates do largely represent the truth, then what is the clinical value of the observed differences? According to the forest plot for blood loss for women undergoing minimal invasive surgery, the claimed “effective” interventions, i.e., oxytocin, ornipressin, misoprostol, bupivacaine plus epinephrine and vasopressin, seem to be based on the statistical significances rather than clinical important differences. Compared to placebo or no treatment, the mean differences of blood loss varied from -32 ml to -175 ml. It is unclear whether these differences should be considered clinically relevant. For any meta-analysis it is important to translate the findings into clinically relevant effects. Without taking clinically relevant differences in to consideration, the value of statistically significance could be meaningless.

How can future research improve?

We started our argument with the bioaccumulation of toxic substances analogy and what we can learn from that? The key step to prevent bioaccumulation in food chains is avoiding the use of toxic substances. Similarly, improving the quality of the primary RCTs will help and prevent us from making potential misleading recommendation on the basis of meta-
analyses. We can understand how to improve future research by learning from the existing low-quality evidence in this network meta-analysis. The first step is to locate the low-quality evidence at a comparison level in network plots. By doing so, we can easily identify the direct comparisons with weak evidence and the comparisons that have not been compared directly in the past. These comparisons should be the focus of future trials, especially when the interventions are the top-ranking interventions in the network meta-analysis. The second step is to understand the reasons for low quality. In the case of treatments to reduce blood loss following myomectomy, risk of selection bias and imprecision were the key limitations. To reduce the risk of bias in future RCTs, appropriate randomization and allocation concealment methods should be used, ideally in a placebo-controlled design. To increase precision, sample sizes need to be justified based on a clear and valid hypothesis in future trials. Lastly, to make the trial feasible and to improve the generalizability of the results, a multicentric design should be preferred.

Rui Wang, M.D., Ph.D.a
Madelon van Wely, Ph.D.b

a Department of Obstetrics and Gynaecology, Monash University, Melbourne, Victoria, Australia; and b Centre for Reproductive Medicine, UMC Amsterdam, University of Amsterdam, Amsterdam, the Netherlands

https://doi.org/10.1016/j.fertnstert.2019.11.026

You can discuss this article with its authors and other readers at https://www.fertstertdialog.com/users/16110-fertility-and-sterility/posts/56485-29351

REFERENCES


