Improving reporting of research reports

By asking how many top fertility journals report both relative and absolute effect size measures and their precision, Glujovsky et al. (1) call attention to an important detail in the reporting and interpreting of research reports. They urge the use of internationally well-recognized standardized checklists for reviewers (2). They illustrate the concept that if enough subjects are studied, statistical significance can always be reached; however, this does not assure that the proper effect size of clinical importance is found. Their cross-sectional study is transparent and well reported yet in a sense is limited because only one database was used, only randomized clinical trials (RCTs) are included, and only top fertility journals are assessed. As clinicians we rely on RCTs to evaluate treatment effects. We put RCTs higher on the hierarchy of study designs, perhaps because RCTs can be very helpful to deal with selection bias … a common problem. The message the authors are trying to convey is just as relevant to other comparative study designs. As readers we need to be familiar with the strengths and weaknesses of all study designs and their relative hierarchy to practice based on best evidence. Unfortunately, the majority of evidence we need to assess will never be studied by RCT design for ethical, cost, and/or other considerations. RCTs are most useful if properly designed to answer burning questions, and they should have enough power to answer whether a minimal clinically significant difference can be found before they are undertaken. In relative terms, they will remain uncommon. We also need to improve reporting of other comparative studies to practice based on best evidence.

Every author should provide a structured abstract for their submitted manuscript and should mention whether the difference between study arms was clinically significant as well as statistically significant. In the abstract and in the results sections, the absolute differences, including confidence intervals (CIs), should be reported. As clinicians we need to be able to look at abstracts quickly. Medical information continues to explode exponentially. We have limited time to read research reports. Electronic tools continue to improve. We can more easily capture specific information relevant to well-framed answerable clinical questions if this information is easily found. We need concise reporting of key elements to help us better focus our precious time and attention to find individualized clinically meaningful relevant information.

The authors looked for reporting of a minimal clinically important difference (MCID) or its proxy, which is the difference needed to achieve adequate power in the planning of the study. MCID is a statistical term calculated using one of several methods, and there is no consensus on the optimal technique to obtain this measure. It is defined as the smallest change in an outcome that a patient would identify as important. MCID offers a threshold above which outcomes are experienced as relevant by the patient; this avoids the problem of mere statistical significance. The MCID varies according to diseases and outcome instruments, but it does not depend on treatment methods. Two different treatments for a similar disease can be compared using the same MCID if the outcome measurement instrument is the same (3). The authors' purpose is to plead for better reporting of clinical relevance. The optimal statistical methods to do this will invite controversy; however, the overall message should not.

As reviewers we should be asking authors to reveal the smallest clinically meaningful difference they are using. Readers have the option to determine, based on [1] other elements of the study, [2] understanding of the topic, [3] other literature, and [4] knowledge of other alternatives, whether that clinically meaningful difference is meaningful. Absolute difference or risk difference (difference in calculated risk for each arm) is what the clinician needs to know. In evidence-based practice, we divide that difference into 1 to obtain the number needed to benefit or to harm (number needed to treat) (4). Many clinicians believe this is a better way to understand the magnitude of the effect size in clinical terms. Central to the authors' point, however, is the necessity to report CIs to draw conclusions about both clinical and statistical significance (4). From a clinician's perspective, the CI should be a very useful tool to help understand the clinical relevance of the comparison. A *P* value by itself suggests the difference is not because of chance alone. Strength of association, direction, or precision of the measure is not conveyed when only the *P* value is reported. An arbitrary comfort level of $<.05$ has evolved as statistically significant. By definition, if the comparison is not statistically significant, it cannot be clinically significant. It may be inconclusive. Smaller trials are more prone to alpha or type I errors (incorrectly concluding a difference exists) and to a type II or beta error (no difference found when there really is a difference). Without the CI and without understanding what an MCID is, interpreting the *P* value by itself does not convey adequate clinically meaningful information. All readers should know how to use and interpret CIs. When the limits of the interval cross the null (0 is a null for differences and 1 is the null for ratios), the findings are nonsignificant when the arms of the trial are compared. The *P* value is not needed for this inference. Odds ratios, relative risks, hazard ratios, risk differences, and MCID should be part of our lexicon. Our lexicon should also include how to interpret and use CIs. Whether we look for differences or ratios, understanding whether or not the lower or upper limit has clinical importance helps us better understand the precision of the estimates (closer with greater absolute number of events or larger samples and therefore more precision). Knowing if the point estimate and CI are all on the same side of the null gives direction so we can understand the superiority or noninferiority of the comparative arms of the study.

I agree with the authors that an in-depth reading of the methods and results could help readers avoid the subjectivity that authors place in the rest of the manuscript. Optimal

results reporting helps readers reach a correct interpretation. Conclusions, based on the results, to answer the objectives are a must for better reporting of clinical research. Their absence can mislead. However, the abstract is often the only segment many readers read! Transparent reporting to include CIs for the minimally clinically relevant difference and statistical differences included in the abstract helps provide a better opportunity to practice based on best evidence. The authors are politely reminding us to up our game as peer reviewers. This will lead to more helpful clinical interpretation of the reports we review. I congratulate them for letting us know how and where we can improve!

Robert A. Wild, M.D., M.P.H., Ph.D.
Section Reproductive Endocrinology and Infertility,
Department of Obstetrics and Gynecology,
Oklahoma University Health Sciences Center,
Oklahoma City, Oklahoma

http://dx.doi.org/10.1016/j.fertnstert.2016.01.010

You can discuss this article with its authors and with other ASRM members at

http://fertstertforum.com/wildr-improving-reporting-research-reports/

Use your smartphone to scan this QR code and connect to the discussion forum for this article now.*

* Download a free QR code scanner by searching for "QR scanner" in your smartphone's app store or app marketplace.

REFERENCES

1. Glujovsky D, Sueldo CE, Borghi C, Nicotra P, Andreucci S, Ciapponi A. Misleading reporting and interpretation of results in major infertility journals. Fertil Steril 2016;105:1301–6.
2. Equator Network. Enhancing the quality and transparency of health research. Available at: http://www.equator-network.org/reporting-guidelines. Accessed February 1, 2016.
3. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials 1989;10:407–15.
4. Guyatt G, Meade M, Rennie D, Cook D. Users' guides to the medical literature: a manual for evidence-based clinical practice. 3rd ed. New York: McGraw Hill Education; 2015.