EDITORIAL

The clinician's guide to randomized trials: interpretation

INTRODUCTION

Ophthalmologists rely on scientific evidence from randomized controlled trials (RCTs) to inform clinical decisions. When designed and executed optimally, large RCTs balance both the known and unknown factors that may affect the outcome of interest (e.g. visual acuity, intraocular pressure) resulting, theoretically, in an observed effect solely driven by the intervention/exposure (e.g. drug or surgery). However, an understanding of the fundamental elements of the RCT is essential if clinicians are to accurately interpret the results of RCTs; not all RCTs are designed, conducted, and reported with the same methodological rigour [1, 2].

We will focus on four key elements to assess when interpreting RCTs—risk of bias, statistical power, treatment effect, and applicability.

ASSESSING THE RISK OF BIAS

Bias in trials is systematic deviation from the truth, which can significantly affect the observed treatment effect in an RCT [3, 4]. Clinicians interpreting RCTs should therefore place a high level of scrutiny on the potential sources of bias in a trial to weight their confidence in the results appropriately. The Cochrane Collaboration's tool for assessing risk of bias [5] provides a modern tool for assessing methodological quality of an RCT. We extensively describe the types of bias a clinician should be aware of when interpreting an RCT in our previous editorial on risk of bias measurement [6].

ASSESSING STATISTICAL POWER

The sample size calculation determines how many participants are required for a specific trial—considering recruitment, randomization, administration of interventions, follow-up for outcomes, and analysis—to detect a minimum important difference (MID) in the primary outcome of interest. Thus, the sample size calculation is a critical component of the RCT design affecting the capability of the study to correctly answer the primary study question. A clinician needs be aware of how many participants were required, and whether this target was met during the recruitment phase as well as at the subsequent follow-up time points. An illustration of an adequately powered sample is the Protocol W trial investigating the efficacy of intravitreous aflibercept injections for nonproliferative diabetic retinopathy [7], which reported an enrolment target of 386 eyes to provide 89% power to detect an MID of 15% in the primary outcome of centre-involved diabetic macular oedema development with vision loss or proliferative diabetic retinopathy between groups at 2 years follow-up: this clearly defines the rationale behind sample calculation and provides the necessary details for the assessment of the statistical power. It is important to note that RCTs that were adequately powered at the outset of the study may lose the required statistical power (ability to detect the MID) if there is high patient dropout, crossover, or loss to follow-up. Underpowered studies are difficult to interpret, especially if the result is not statistically significant. Conversely, overpowered studies can yield results that may not be clinically important (e.g. a large RCT on diabetic macular oedema may detect a 20 µm of difference in OCT central macular thickness for a new drug, but such a difference, while statistically different, may not be clinically important).

ASSESSING THE TREATMENT EFFECT

The heart of any RCT is the "results" section and the observed treatment effect. When interpreting the treatment effect, a clinician ought to consider two key aspects: the precision, and the importance of relevance of the estimate. The precision of the treatment-effect estimate is identified best through an examination of the confidence interval (typically 95%). The 95% confidence interval represents the range of values which the clinician can be 95% confident the true treatment effect lies within. A precise estimate is observable in the form of tight confidence intervals; large confidence intervals should be interpreted with caution, as the true treatment effect could be anywhere within the wide range of values, making conclusions on treatment effect unclear. For example, the PANORAMA trial [8] demonstrated a significantly higher proportion of eyes with an improvement in the primary outcome of diabetic retinopathy severity scale (DRSS) (2 levels or more) in the combined aflibercept groups compared to the control group at 24 weeks, with a difference of 52.3% [95% CI 45.2%, 59.5%, $p < 0.001$]. Given the high precision of the effect estimate, we can be confident that intravitreal aflibercept injections are effective in improving the severity of diabetic retinopathy by 45.2–59.5%. Moreover, the clinical importance of the observed treatment effect is of great interest, as clinical importance translates into meaningful improvement for the patient. A statistically significant result does not necessarily translate into a clinically meaningful difference between groups, and conversely, it is possible that clinically important differences between groups are observed but not statistically significant, particularly when the sample size is small and estimate is imprecise. Clinicians should apply a combination of clinical acumen, experience, and established MID values in the literature to determine whether the observed treatment effect is clinically important.

ASSESSING THE APPLICABILITY

Applicability of study findings is critical for observed treatment benefits to translate into clinical benefits for patients in the real-world settings. When assessing the applicability of the study results, there are several factors to consider. Firstly, the study population used in the trial ought to be comparable to the clinician's patient population. If inclusion/exclusion criteria used in

the trial are too strict, it is possible that the observed treatment effect may not translate to the clinician's general patient population. For example, a RCT on a new minimally invasive glaucoma surgery (MIGS) on patients who are treatment naive may not be applicable to patients who are on two or three glaucoma drops in the clinic. Secondly, the feasibility of the intervention delivery and expertise of the health-care provider are important factors affecting applicability. Particularly in surgical interventions, the level of expertise can have significant effects on treatment outcomes; [9, 10] a clinician must be able to adequately deliver the treatment to achieve an optimal treatment effect. Additionally, the intervention must also be reasonable with regards to compliance demands on the patient in the "real-world" setting. A treatment may be effective in a controlled environment, but if it creates unreasonable demands on the patient or health-care provider, its effectiveness may be different in a real-world setting. A classic example were the initial RCTs on intravitreal anti-VEGF therapy for age-related macular degeneration which required fixed monthly therapy over 24 months [11]. Results of the RCTs could not be replicated in real-world settings as adherence to monthly intravitreal anti-VEGF therapy was not practical for many patients. Thus, applicability is a key consideration for the clinician when deciding whether to adjust his clinical practice based of the results of an RCT.

CONCLUSION
The ability to interpret RCTs through an understanding of the risk of bias, statistical power, treatment effect, and applicability of an RCT is critical for clinicians to make sound decisions in the field.

Alex Thabane [1], Mark R. Phillips [1], Tien Y. Wong[2],
Lehana Thabane [1,3], Mohit Bhandari [1,4],
Varun Chaudhary [1,4 ✉] and for the Retina Evidence Trials
InterNational Alliance (R.E.T.I.N.A.) Study Group*
[1]Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Ontario, Canada. [2]Singapore Eye Research Institute, Singapore National Eye Centre, Singapore, Singapore. [3]Biostatistics Unit, St. Joseph's Healthcare-Hamilton, Hamilton, Ontario, Canada. [4]Department of Surgery, McMaster University, Hamilton, Ontario, Canada. *A list of authors and their affiliations appears at the end of the paper.
✉email: vchaudh@mcmaster.ca

REFERENCES
1. Yao AC, Khajuria A, Camm CF, Edison E, Agha R. The reporting quality of parallel randomised controlled trials in ophthalmic surgery in 2011: a systematic review. Eye. 2014;28:1341–9. https://doi.org/10.1038/eye.2014.206.
2. Lee CF, Cheng ACO, Fong DYT. Eyes or subjects: are ophthalmic randomized controlled trials properly designed and analyzed? Ophthalmology. 2012;119:869–72. https://doi.org/10.1016/j.ophtha.2011.09.025.
3. Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., eds. Cochrane handbook for systematic reviews of interventions. John Wiley & Sons; 2019.
4. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials. Jama. 1995;273:408–12. https://doi.org/10.1001/jama.273.5.408.
5. Sterne JA, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019;366:l4898. https://doi.org/10.1136/bmj.l4898.
6. Phillips MR, Kaiser P, Thabane L, Bhandari M, Chaudhary V. Risk of bias: why measure it, and how? Eye 2021. https://doi.org/10.1038/s41433-021-01759-9.
7. Maturi RK, Glassman AR, Josic K, Antoszyk AN, Blodi BA, Jampol LM, et al. Effect of intravitreous anti–vascular endothelial growth factor vs sham treatment for prevention of vision-threatening complications of diabetic retinopathy: the Protocol W Randomized Clinical Trial. JAMA Ophthalmol. 2021;139:701–12. https://doi.org/10.1001/jamaophthalmol.2021.0606.
8. Brown DM, Wykoff CC, Boyer D, Heier JS, Clark WL, Emanuelli A, et al. Evaluation of intravitreal aflibercept for the treatment of severe nonproliferative diabetic retinopathy: results from the PANORAMA randomized clinical trial. JAMA Ophthalmol. 2021;139:946–55. https://doi.org/10.1001/jamaophthalmol.2021.2809.
9. Birkmeyer JD, Finks JF, O'Reilly A, Oerline M, Carlin AM, Nunn AR, et al. Surgical skill and complication rates after bariatric surgery. N Engl J Med. 2013;369:1434–42. https://doi.org/10.1056/nejmsa1300625.
10. Fecso AB, Szasz P, Kerezov G, Grantcharov TP. The effect of technical performance on patient outcomes in surgery. Ann Surg. 2017;265:492–501. https://doi.org/10.1097/SLA.0000000000001959.
11. Okada M, Mitchell P, Finger RP, Eldem B, Talks SJ, Hirst C, et al. Nonadherence or nonpersistence to intravitreal injection therapy for neovascular age-related macular degeneration: a mixed-methods systematic review. Ophthalmology. 2021;128:234–47. https://doi.org/10.1016/j.ophtha.2020.07.060.

FOR THE RETINA EVIDENCE TRIALS INTERNATIONAL ALLIANCE (R.E.T.I.N.A.) STUDY GROUP
Varun Chaudhary [1,4 ✉], Mohit Bhandari [1,4], Charles C. Wykoff[5,6], Sobha Sivaprasad[7], Lehana Thabane [1,3], Peter Kaiser[8], David Sarraf[9], Sophie Bakri[10], Sunir J. Garg[11], Rishi P. Singh[12,13], Frank G. Holz[14], Tien Y. Wong[2] and Robyn H. Guymer[15,16]

[5]Retina Consultants of Texas (Retina Consultants of America), Houston, Texas, USA. [6]Blanton Eye Institute, Houston Methodist Hospital, Houston, Texas, USA. [7]NIHR Moorfields Biomedical Research Centre, Moorfields Eye Hospital, London, UK. [8]Cole Eye Institute, Cleveland Clinic, Cleveland, OH, USA. [9]Retinal Disorders and Ophthalmic Genetics, Stein Eye Institute, University of California, Los Angeles, CA, USA. [10]Department of Ophthalmology, Mayo Clinic, Rochester, MN, USA. [11]The Retina Service at Wills Eye Hospital, Philadelphia, PA, USA. [12]Center for Ophthalmic Bioinformatics, Cole Eye Institute, Cleveland Clinic, Cleveland, OH, USA. [13]Cleveland Clinic Lerner College of Medicine, Cleveland, OH, USA. [14]Department of Ophthalmology, University of Bonn, Boon, Germany. [15]Centre for Eye Research Australia, Royal Victorian Eye and Ear Hospital, East Melbourne, Australia. [16]Department of Surgery (Ophthalmology), The University of Melbourne, Melbourne, Australia.