Accuracy of Baseline Prevalence Estimates for Sample-size Calculations in Randomized Controlled Trials

ANDREW AULTMAN, BS; KELLY C. DITTER, MD; HECTOR MENDEZ-FIGUEROA, MD; MEGHA GUPTA, MD, MSc; MICHAL FISHEL BARTAL, MD; SUNEET P. CHAUHAN, MD, HON DSc; STEPHEN WAGNER, MD

OBJECTIVE

Sample-size calculations are an underpinning to designing a randomized controlled trial (RCT).[1] If the prevalence of a disease process is overestimated during the design phase of an RCT, the trial will recruit fewer participants and appear more feasible from a time and resource perspective. However, this overestimation increases the risk of type II error, creating a noninformative trial.[2] As a result, studies with inaccurate sample-size calculations may lack clinical utility and may even be considered unethical due to the risks undertaken by participants.[3,4] Data in the obstetrical literature suggests that the majority of RCTs inaccurately estimate the prevalence of the disease process being investigated.[5] The objective of this study was to ascertain whether the estimated prevalence of the outcome used for sample-size calculations in RCTs in all specialties is similar to the actual reported prevalence(as identified in the frequency of primary outcome in the placebo arm).

METHODS

During a one-year study period (2021) all RCTs were manually identified and abstracted from three major journals (*JAMA, NEJM*, and the *Lancet*). In each trial, the disease prevalence for the primary outcome that was used for the sample-size calculation was identified and compared to the actual disease prevalence detected in the completed RCT. If the prevalence of the primary outcome was below the assumed rate in the study design, the trial was considered to have an inaccurate assessment of prevalence. Studies with underestimated disease prevalence rates were not considered to have inaccurate assessments, as this variation would still allow the study to detect a true association. Chi-square was used for all categorical variables, and P <0.05 was considered significant.

RESULTS

Of the 265 identified RCTs, 101 were included in the analysis. The most common reasons for trial exclusion were a primary outcome that did not include disease prevalence (e.g., time) and non-inferiority trial design. Forty-six trials (45%) had an estimated disease prevalence rate equal to or greater than the disease prevalence in the study, and 55 (55%) trials had an inaccurate assessment of the prevalence of the disease (**Table 1**). There was no difference in the estimated sample sizes, number of centers, or geographic location between trials. Trials with an accurate baseline rate were more likely to anticipate a change in the baseline rate of > 30% for sample size. There was no difference in the rate of positive trials between those with an accurate and inaccurate estimate of disease prevalence (62.2% vs. 67.2%, p=0.60). Eighty percent (80%) of trials had a disease prevalence estimate within 42% of the disease prevalence in the study, and 95% of all trials were within 83%.

DISCUSSION

Over half of the RCTs examined here overestimated the prevalence of the disease process in the primary outcome and thus were at increased risk of a type II error. When designing RCTs researchers should consider factoring in an underestimation of disease prevalence (by as much as 40%) into their sample-size calculations, to minimize the risk of type II error. Overestimation of treatment effect, which was not examined in this paper, could also result in a type II error and examination of estimated effect size could be a focus of a future investigation. Further research to identify methods of improving disease estimates is warranted.

References

1. Biau DJ, Kerneis S, Porcher R. Statistics in brief: the importance of sample size in the planning and interpretation of medical research. Clin Orthop Relat Res 2008;466(9):2282-8. DOI: 10.1007/s11999-008-0346-9.
2. Moher D, Dulberg CS, Wells GA. Statistical power, sample size, and their reporting in randomized controlled trials. Jama 1994;272(2):122-4. (In eng).
3. Bacchetti P, Wolf LE, Segal MR, McCulloch CE. Ethics and sample size. Am J Epidemiol 2005;161(2):105-10. DOI: 10.1093/aje/kwi014.
4. Speich B, von Niederhäusern B, Schur N, et al. Systematic review on costs and resource use of randomized clinical trials shows a lack of transparent and comprehensive data. J Clin Epidemiol 2018;96:1-11. (In eng). DOI: 10.1016/j.jclinepi.2017.12.018.
5. Ditter KC, Crowe EH, Mendez-Figueroa H, et al. Accuracy of baseline prevalence estimates for sample size calculations in obstetrical randomized controlled trials. Am J Obstet Gynecol 2022;226(6):855-856. DOI: 10.1016/j.ajog.2022.01.032.

Table 1. Characteristics of randomized controlled trials based on accuracy of estimation of baseline primary outcome rates

Data presented as N (%) or median (quartiles). An inaccurate estimate was defined if the prevalence of the primary outcome was below the assumed rate in the trial.

	Sample size had accurate assessment of prevalence of disease (N=46)	Sample size had inaccurate assessment of prevalence of disease (N=55)	P
Journals, N (%) JAMA Lancet NEJM	 11 (23.9) 14 (30.4) 21 (45.6)	 19 (34.6) 11 (20.0) 25 (45.5)	0.36
Centers, N (%) Single Multiple	 5 (11.1) 40 (88.9)	 7 (7.4) 50 (92.6)	0.52
Country, N (%) USA Alone USA and Other Countries Other Countries	 10 (12.7) 17 (37.0) 19 (41.3)	 7 (12.7) 32 (58.2) 16 (29.1)	0.10
Baseline Rate Derived From, N (%) Reference Institutional Rate No reference	 20 (44.4) 23 (51.1) 2 (4.4)	 29 (52.7) 22 (40.0) 4 (7.3)	0.51
Presumed Change in Baseline Rate, N (%) < 30% ≥ 30%	 29 (63.0) 17 (37.0)	 23 (41.8) 32 (58.2)	**0.03**
Power, N (%) 80–89% 90–99%	 20 (46.5) 23 (53.5)	 32 (59.3) 22 (40.7)	0.21
Positive trials, N (%)	28 (62.2)	37 (67.2)	0.60
Sample Needed, Median (IQR)	2372 (350–4000)	2378 (260–1170)	0.13
Expected Lost to Follow-Up %, Median (IQR)	4.10 (0–9)	4.65 (0–9)	0.77
Assess for Eligibility for Trial	16691.1 (735–5187)	10477.6 (621–6112)	0.61
Ineligible, Median (IQR)	1537.3 (160–1656)	7013.0 (9–2264)	0.98
Declined Participation, Median (IQR)	121.5 (0–131.5)	270.1 (5–226)	0.08
Lost to Follow-Up, Median (IQR)	4.1 (0–9)	4.6 (0–9)	0.77

Authors

Andrew Aultman, BS, Department of Obstetrics and Gynecology, Alpert Medical School, Brown University, Providence, RI

Kelly C. Ditter, MD, Department of Obstetrics and Gynecology, Methodist Hospital, Houston, TX

Hector Mendez-Figueroa, MD, Department of Obstetrics, Gynecology, and Reproductive Sciences, McGovern Medical School, The University of Texas Health Science Center at Houston, Houston, TX

Megha Gupta, MD, MSC, Department of Obstetrics and Gynecology, Beth Israel Deaconess Medical Center, Department of Obstetrics, Gynecology and Reproductive Biology, Harvard Medical School, Boston, MA

Michal Fishel Bartal, MD, Department of Obstetrics, Gynecology, and Reproductive Sciences, McGovern Medical School, The University of Texas Health Science Center at Houston, Houston, TX

Suneet P. Chauhan, MD, Hon DSc, Department of Obstetrics, Gynecology, and Reproductive Sciences, McGovern Medical School, The University of Texas Health Science Center at Houston, Houston, TX

Stephen Wagner, MD, Department of Obstetrics and Gynecology, Beth Israel Deaconess Medical Center, Department of Obstetrics, Gynecology and Reproductive Biology, Harvard Medical School, Boston, MA

Disclosures

The authors report no conflicts of interest.

There was no funding received for conducting this project.

Correspondence

Stephen Wagner, MD
Department of Obstetrics and Gynecology
Harvard Medical School
330 Brookline Ave, Boston, MA 02445
717-283-7796
Fax 401-453-7599
smw5120@gmail.com