Clinical Epidemiology

Dovepress
open access to scientific and medical research

Open Access Full Text Article

METHODOLOGY

Guidance for using pilot studies to inform the design of intervention trials with continuous outcomes

This article was published in the following Dove Press journal:
Clinical Epidemiology

Melanie L Bell[1]
Amy L Whitehead[2]
Steven A Julious[2]

[1]Department of Epidemiology and Biostatistics, Mel and Enid Zuckerman College of Public Health, University of Arizona, Tucson, AZ, USA; [2]Medical Statistics Group, Design, Trials and Statistics, School of Health and Related Research (ScHARR), University of Sheffield, Sheffield, UK

Background: A pilot study can be an important step in the assessment of an intervention by providing information to design the future definitive trial. Pilot studies can be used to estimate the recruitment and retention rates and population variance and to provide preliminary evidence of efficacy potential. However, estimation is poor because pilot studies are small, so sensitivity analyses for the main trial's sample size calculations should be undertaken.

Methods: We demonstrate how to carry out easy-to-perform sensitivity analysis for designing trials based on pilot data using an example. Furthermore, we introduce rules of thumb for the size of the pilot study so that the overall sample size, for both pilot and main trials, is minimized.

Results: The example illustrates how sample size estimates for the main trial can alter dramatically by plausibly varying assumptions. Required sample size for 90% power varied from 392 to 692 depending on assumptions. Some scenarios were not feasible based on the pilot study recruitment and retention rates.

Conclusion: Pilot studies can be used to help design the main trial, but caution should be exercised. We recommend the use of sensitivity analyses to assess the robustness of the design assumptions for a main trial.

Keywords: pilot, feasibility, sample size, power, randomized controlled trial, sensitivity analysis

Introduction

Prior to a definitive intervention trial, a pilot study may be undertaken. Pilot trials are often small versions of the main trial, undertaken to test trial methods and procedures.[1,2] The overall aim of pilot studies is to demonstrate that a future trial can be undertaken. To address this aim, there are a number of objectives for a pilot study including assessing recruitment and retention rates, obtaining estimates of parameters required for sample size calculation, and providing preliminary evidence of efficacy potential.[3–6]

We illustrate how to use pilot studies to inform the design of future randomized controlled trials (RCTs) so that the likelihood of answering the research question is high. We show how pilot studies can address each of the objectives listed earlier, how to optimally design a pilot trial, and how to perform sample size sensitivity analysis. Our example uses a continuous outcome, but most of the content can be applied to pilot studies in general.

Considerations for trial design

When designing a definitive trial, one must consider

- The target effect size, such as the difference in means for continuous outcomes;
- The variance about the estimates for continuous outcomes, which is used to give a range of responses for individuals in the trial;

Correspondence: Melanie L Bell
Department of Epidemiology and Biostatistics, Mel and Enid Zuckerman College of Public Health, University of Arizona, 1295 North Martin Avenue, PO Box 2945211, Tucson, AR 85724, USA
Tel +1 520 626 2795
Email melaniebell@email.arizona.edu

submit your manuscript | www.dovepress.com
Dovepress
http://dx.doi.org/10.2147/CLEP.S146397

- Feasibility, including referral, recruitment, and retention rates.

Pilot trial results can inform each of these elements. Factors such as type I error and power are set independent of the pilot and are discussed in detail elsewhere.[7] We focus on external pilot studies, where the trial is run before the main trial, and results are not combined.[8]

Feasibility

The first consideration is feasibility: will the researchers be able to recruit the required number of participants within the study timeframe and retain them in the main trial? While review of clinical records can be used to give some indication of potential participant pool, pilot studies provide estimates of the number of participants that actually enroll and consent to randomization, and these estimates should be included in the manuscripts that report the pilot study results.[9] Many trials struggle to reach their sample size goal, which can result in trial extensions or failure to recruit to their prespecified sample size.[10] Failure to recruit is a major issue in UK publicly funded trials, where 45% fail to reach target sample size.[10] Along with review of prior trials at the same centers in similar populations, pilot studies can also give estimates of retention rates and adherence rates.[11] Missing data and dropouts are issues in most RCTs[12] and need to be considered at each step of the research process,[13] including design, reporting,[9] and progression to a larger definitive trial.[11,14]

Target effect size and potential efficacy

Hislop et al[15] undertook a systematic review to identify seven approaches for determining the target effect size for an RCT and classified them as clinically important and/or realistic. A specific type of clinically important difference is the smallest value that would make a difference to patients or that could change care, a quantity referred to as the minimum important difference (MID), or sometimes minimally clinical important difference. The MID can be difficult to determine, particularly as it can change with patient population. However, researchers in various fields have investigated MID estimation and provide guidance on estimation.[16,17] In the absence of a known MID for continuous outcomes, particularly patient-reported outcomes, a standardized effect size[15] between 0.3 and 0.5 has been recommended.[17,18] Expert opinion is also used to specify important differences.[15] Although some researchers use the pilot effect size to power the definitive trial, this is a practice that should be avoided in general, as estimation is poor due to the small sample size, and is likely to mislead.[19]

The target effect size must also be realistic, and the estimated effect size and confidence interval (CI) from the pilot can give some evidence here, ie, whether there is any indication that the intervention is effective and important differences might be obtained in the main trial.[5] The small sample size of a pilot makes estimation uncertain, so caution must be exercised.[19,20] One approach for handling this uncertainty is to use significance levels other than the "traditional" 5% to provide preliminary evidence for efficacy, with corresponding CIs such as 85 and 75% in addition to 95% CIs.[21] A figure showing these CIs, the MID, and the null value can be a helpful way of displaying pilot results, by facilitating an assessment of both statistical significance and the potential for clinical significance.[31] While some authors argue against carrying out hypothesis tests and assessing efficacy from pilots, even potential efficacy, most pilot studies do undertake hypothesis tests.[6] We strongly stress that preliminary efficacy evidence from a pilot study should not be overstated, and researchers should avoid temptation to forgo the main trial.[20,22]

Estimating the standard deviation (SD)

The population SD is another key element of sample size estimation for continuous outcomes, and its estimation is one of the objectives for conducting a pilot study. However, similar to the effect size, the SD can be imprecisely estimated due to the pilot's small sample size. Using a pilot study's SD to design a future sample size has been shown to often result in an underpowered study.[23,24] Thus, sensitivity analyses should be undertaken.

Sensitivity analysis for sample size

Sensitivity analyses are important to assess the robustness of study results to the assumptions made in the primary analysis.[25] Sensitivity analyses should also be performed in the design stage[26] and can take the form of accounting for the uncertainty in estimation by calculating sample sizes based on a range of plausible SDs and retention/dropout rates. Browne[23] suggested using the pilot study's upper limit of the 80% CI for the SD to calculate sample size in the subsequent trial. One may also consider SDs from the literature.

Pilot study sample size

In order to have the best chance of answering the research question, researchers should carefully consider the size of not only the definitive trial but also the pilot as well. Although traditional power calculations are inappropriate for pilot studies (since the primary aim of a pilot study is not to test

superiority of one treatment over the other), a sample size justification is important. While there are several rules of thumb for the size of a pilot study, ranging from 12 to 35 individuals per arm,[5,27] none of these guidelines account for the likely size of the future trial.

Whitehead et al[27] showed how, if you know the main trial's target effect size, you can estimate the pilot study's optimum sample size, minimizing the number of patients recruited across the two studies. From this work, they proposed stepped rules of thumb for pilot studies based on the target effect size and the size of the future trial. These rules are summarized in Table 1. For example, if the future trial will be designed around a small effect, then the number of patients per arm for the pilot study should be 25 for 90% power. Using these rules increases the likelihood of appropriate power for the future trial. Cocks and Torgerson[5] also recommend basing the pilot study size on the future trial's size, if the SD is known.

Example

Suppose a research team is planning a pilot in the anticipation of designing a definitive trial. The main trial will be a two-arm RCT comparing a new supportive care regimen for cancer patients to usual care, with assessments at baseline, 6 weeks, and 3 months. Their primary outcome is the quality of life at 3 months as measured by the Functional Assessment of Cancer Therapy-General (FACT-G), a 27-item questionnaire covering aspects of physical, social, family, emotional, and functional well-being.[28]

Pilot study sample size

To use the stepped rules of thumb for pilot sample size, the researchers must consider the target effect size and SD for the main trial in order to calculate the standardized difference (effect size). They find that the estimated FACT-G MID is between three and six points[29] and an SD estimate from the literature[30] is 14 in similar populations. Using an MID estimate of four points, and an SD of 14, the standardized effect size is 4/14=0.29. For a 90% powered main trial, they should use a sample size of 25 per arm for the pilot (Table 1).

Pilot study results

Suppose now the researchers undertake the pilot study of 50 participants with recruitment over 2 months. Of the 100 potential participants, 70 participants were referred by their oncologist, 60 participants met eligibility criteria, and 50 participants agreed to participate. This indicates a recruitment rate of 50% of eligible patients, at 25 recruitments per month. Of the 50 participants, 40 participants completed all three assessments; retention is 80%. These rates will aid in estimating the main trial duration.

The difference in the quality of life between the arms at 3 months is estimated at 3.1 points, with 95% CI −1.8 to 8.0, and SD =11.2. Figure 1 shows several CIs demonstrating

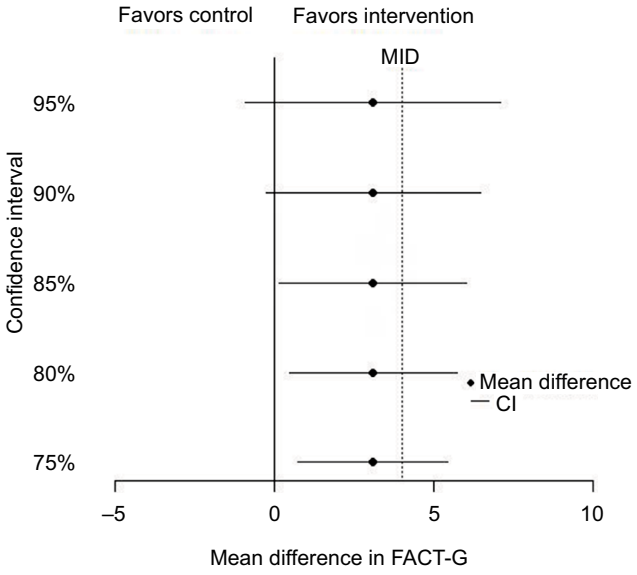

Figure 1 Mean difference in FACT-G scores between pilot study intervention and control arms with confidence intervals.
Abbreviations: FACT-G, Functional Assessment of Cancer Therapy-Genera; MID, minimum important difference.

Table 1 Stepped rules of thumb for pilot study sample size per arm, as a function of the target effect size (standardized difference) and power of the main trial

Standardized difference, d[a]	80% powered main trial		90% powered main trial	
	Pilot N per arm	Main trial N per arm	Pilot N per arm	Main trial N per arm
Extra small (d<0.1)	50	>1571	75	>2103
Small (0.1≤d<0.3)	20	176–1571	25	235–2103
Medium (0.3≤d<0.7)	10	34–176	15	44–235
Large (d≥0.7)	10	≤34	10	≤44

Notes: [a] $d = \frac{\bar{X}_{int} - \bar{X}_{ctl}}{SD_{pooled}}$, where $SD_{pooled} = \sqrt{\frac{(n_{int}-1)s_{int}^2 + (n_{ctl}-1)s_{ctl}^2}{n_{int}+n_{ctl}-2}}$. The corresponding likely size of the main trial is also shown. int is the intervention arm, and ctl is the control arm.

Table 2 A range of sample sizes varying dropout, recruitment rate, and estimated SD assuming an effect size of four points

SD justification	SD	Dropout (%)	90% powered main trial			80% powered main trial		
			N (total)	Required recruitment rate per month[a]	Feasible[b]	N (total)	Required recruitment rate per month[a]	Feasible[b]
Pilot study SD	11.2	15	392	22	Yes	296	16	Yes
		20	416	23	Yes	314	17	Yes
		25	444	25	Yes	334	19	Yes
Upper 80%	13.2	15	542	30	No	406	23	Yes
confidence limit		20	576	32	No	430	24	Yes
from pilot		25	614	34	No	460	26	No
Literature	14.0	15	610	34	No	458	26	No
		20	648	36	No	486	27	No
		25	692	38	No	518	29	No

Notes: [a]Based on 1.5 years of recruitment. [b]Based on the pilot study recruitment rate of 25 participants per month.
Abbreviation: SD, standard deviation.

that the intervention is promising, as each CI contains the MID of 4. Thus, the objective of the pilot study to provide preliminary evidence of efficacy has been met.

Sample size calculations and sensitivity analyses

Table 2 shows sample sizes based on the pilot study's SD, its upper 80% CI limit (taken as the square root of the CI for the variance), and the original estimate from the literature. Sample sizes are also given for the observed dropout rate (20%) and for >5 and <5%. For 90% power, sample size ranges from 392 to 692. For 80% power, sample sizes range from 296 to 518. Note that the sensitivity analysis is quantified in terms of the effect of assumptions on the sample size. An alternative approach is to fix the sample size (at 392 say) and observe how power varies based on assumptions.

Feasibility of the main trial

We now consider feasibility. Specifically, are the researchers likely to be able to recruit the required number of participants within the study timeframe? Based on the funding and the follow-up time of 3 months, recruitment can take 1.5 years. If the pilot recruitment rate of 25 participants per month is a good estimate, then the study will be able to recruit and enroll 450 participants. This falls below several of the estimates in Table 2. Further consideration may be needed how to expand the pool of participants.

Conclusion

We have illustrated how pilot studies can aid in the design of future trials with continuous outcomes by providing estimates of population SD, evidence of potential for intervention effectiveness, and quantification of feasibility in the form of recruitment and retention rates. We have introduced guidelines on pilot study sample size and demonstrated sample size sensitivity analysis. The example demonstrated how main trial sample size estimates can vary dramatically by plausibly altering assumptions.

The decision to progress from a pilot trial to a main trial is generally made using feasibility estimates, as well as issues such as protocol nonadherence. For more information on progression, refer to Avery et al,[11] and for information on the context of internal pilots, refer to Hampson et al.[14] Whether researchers decide to progress to a definitive trial or not, results of pilot studies should be published. A CONSORT extension for reporting results of pilot and feasibility studies gives detailed guidelines.[9]

Acknowledgments

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. The current address of ALW is Southampton Clinical Trials Unit, University of Southampton, Southampton, UK.

Disclosure

Professor MLB is supported by the University of Arizona Cancer Center, through NCI grant P30CA023074. Professor SAJ is funded by the University of Sheffield. Dr ALW was funded by a University of Sheffield studentship. The authors report no other conflicts of interest in this work.

References

1. Thabane L, Ma J, Chu R, et al. A tutorial on pilot studies: the what, why and how. *BMC Med Res Methodol*. 2010;10:1.
2. Arain M, Campbell MJ, Cooper CL, Lancaster GA. What is a pilot or feasibility study? A review of current practice and editorial policy. *BMC Med Res Methodol*. 2010;10:67.

3. Craig P, Dieppe P, Macintyre S, et al. Developing and evaluating complex interventions: the new Medical Research Council guidance. *BMJ*. 2008;337:a1655.
4. Lancaster G, Campbell M, Eldridge S, et al. Trials in primary care: statistical issues in the design, conduct and evaluation of complex interventions. *Stat Methods Med Res*. 2010;19(4):349–377.
5. Cocks K, Torgerson DJ. Sample size calculations for pilot randomized trials: a confidence interval approach. *J Clin Epidemiol*. 2013;66(2):197–201.
6. Shanyinde M, Pickering RM, Weatherall M. Questions asked and answered in pilot and feasibility randomized controlled trials. *BMC Med Res Methodol*. 2011;11(1):117.
7. Julious SA. Sample sizes for clinical trials with normal data. *Stat Med*. 2004;23(12):1921–1986.
8. Wittes J, Brittain E. The role of internal pilot studies in increasing the efficiency of clinical trials. *Stat Med*. 1990;9(1–2):65–72.
9. Eldridge SM, Chan CL, Campbell MJ, et al. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. *BMJ*. 2016;355:i5239.
10. Sully B, Julious SA, Nicholl J. A reinvestigation of recruitment to randomised, controlled, multicenter trials: a review of trials funded by two UK funding agencies. *Trials*. 2013;14:166.
11. Avery KNL, Williamson PR, Gamble C, et al. Informing efficient randomised controlled trials: exploration of challenges in developing progression criteria for internal pilot studies. *BMJ Open*. 2017;7(2):e013537.
12. Bell ML, Fiero M, Horton NJ, Hsu CH. Handling missing data in RCTs; a review of the top medical journals. *BMC Med Res Methodol*. 2014;14(1):118.
13. Bell ML, Fairclough DL. Practical and statistical issues in missing data for longitudinal patient-reported outcomes. *Stat Methods Med Res*. 2014;23(5):440–459.
14. Hampson LV, Williamson PR, Wilby MJ, Jaki T. A framework for prospectively defining progression rules for internal pilot studies monitoring recruitment. *Stat Methods Med Res*. Epub 2017 Jan 01.
15. Hislop J, Adewuyi TE, Vale LD, et al. Methods for specifying the target difference in a randomised controlled trial: the Difference ELicitation in TriAls (DELTA) systematic review. *PLoS Med*. 2014;11(5):e1001645.
16. Revicki DA, Cella D, Hays RD, Sloan JA, Lenderking WR, Aaronson NK. Responsiveness and minimal important differences for patient reported outcomes. *Health Qual Life Outcomes*. 2006;4:70.
17. King M. A point of minimal important difference (MID): a critique of terminology and methods. *Expert Rev Pharmacoecon Outcomes Res*. 2011;11(2):171–184.
18. Norman GR, Sloan JA, Wyrwich KW. The truly remarkable universality of half a standard deviation: confirmation through another look. *Expert Rev Pharmacoecon Outcomes Res*. 2004;4(5):581–585.
19. Kraemer HC, Mintz J, Noda A, Tinklenberg J, Yesavage JA. Caution regarding the use of pilot studies to guide power calculations for study proposals. *Arch Gen Psychiatry*. 2006;63(5):484–489.
20. Loscalzo J. Pilot trials in clinical research: of what value are they? *Circulation*. 2009;119(13):1694–1696.
21. Lee EC, Whitehead AL, Jacques RM, Julious SA. The statistical interpretation of pilot trials: should significance thresholds be reconsidered? *BMC Med Res Methodol*. 2014;14(1):41.
22. Lancaster GA, Dodd S, Williamson PR. Design and analysis of pilot studies: recommendations for good practice. *J Eval Clin Pract*. 2004;10(2):307–312.
23. Browne RH. On the use of a pilot sample for sample size determination. *Stat Med*. 1995;14(17):1933–1940.
24. Vickers AJ. Underpowering in randomized trials reporting a sample size calculation. *J Clin Epidemiol*. 2003;56(8):717–720.
25. Thabane L, Mbuagbaw L, Zhang S, et al. A tutorial on sensitivity analyses in clinical trials: the what, why, when and how. *BMC Med Res Methodol*. 2013;13(1):92.
26. Lewis JA. Statistical principles for clinical trials (ICH E9): an introductory note on an international guideline. *Stat Med*. 1999; 18(15):1903–1942.
27. Whitehead A, Julious S, Cooper C, Campbell MJ. Estimating the sample size for a pilot randomised trial to minimise the overall trial sample size for the external pilot and main trial for a continuous outcome variable. *Stat Methods Med Res*. 2016;25(3):1057–1073.
28. Cella DF, Tulsky DS, Gray G, et al. The functional assessment of cancer therapy scale: development and validation of the general measure. *J Clin Oncol*. 1993;11(3):570–579.
29. Webster K, Cella D, Yost K. The functional assessment of chronic illness therapy (FACIT) measurement system: properties, applications, and interpretation. *Health Qual Life Outcomes*. 2003;1:79.
30. Bell ML, McKenzie JE. Designing psycho-oncology randomised trials and cluster randomised trials: variance components and intra-cluster correlation of commonly used psychosocial measures. *Psychooncology*. 2013;22(8):1738–1747.
31. Bell, ML, Fiero MH, Dhillon HM, Bray VJ and Vardy JL. Statistical controversies in cancer research: using standardized effect size graphs to enhance interpretability of cancer-related clinical trials with patient-reported outcomes. *Ann Oncol*. 2017;28(8):1730–1733.