Design and analysis of pilot studies: recommendations for good practice

Gillian A. Lancaster MSc PhD,1 Susanna Dodd MSc2 and Paula R. Williamson PhD3
1Lecturer in Medical Statistics, 2Research Assistant in Medical Statistics 3Senior Lecturer in Medical Statistics,
Department of Mathematical Sciences, University of Liverpool, Liverpool, UK

Correspondence
Dr. Gillian Lancaster
Department of Mathematical Sciences
University of Liverpool
Liverpool L69 3BX
UK
E-mail: g.lancaster@liv.ac.uk

Abstract
Pilot studies play an important role in health research, but they can be misused, mistreated and misrepresented. In this paper we focus on pilot studies that are used specifically to plan a randomized controlled trial (RCT). Citing examples from the literature, we provide a methodological framework in which to work, and discuss reasons why a pilot study might be undertaken. A well-conducted pilot study, giving a clear list of aims and objectives within a formal framework will encourage methodological rigour, ensure that the work is scientifically valid and publishable, and will lead to higher quality RCTs. It will also safeguard against pilot studies being conducted simply because of small numbers of available patients.

Introduction
Pilot studies play an important role in health research, in providing information for the planning and justification of randomized controlled trials (RCTs) (Anderson & Prentice 1999). RCTs are costly and time-consuming, and major funding bodies such as the UK Medical Research Council require this evidence before large amounts of money will be allocated. Pilot studies may lead to changes in study design. A clear list of aims and objectives for a pilot study is therefore very important.

In this paper we concentrate on pilot studies that are used specifically to plan an RCT. For the main part of the paper we focus on external pilot studies, which we define as stand-alone pieces of work planned and carried out independently to the main study. In contrast, an internal pilot study is incorporated into the main study design of the RCT. The aims of the study are to review the use of pilot studies in the literature, to propose suitable objectives for conducting an external pilot study and to provide suggestions for the analysis of a pilot study, in order to promote scientific rigour.

What constitutes a pilot study?
In a comprehensive literature search using Medline and the Web of Science we could find no formal methodological guidance as to what constitutes a pilot study. In the statistical literature, papers relating to pilot studies dealt mainly with internal pilots (see below). Limiting the search to the years 2000–01 we reviewed four general clinical journals, British Medical Journal (BMJ), Lancet, Journal of the American Medical Association (JAMA), and New England Journal of Medicine (NEJM), and three subject-specific journals, the British Journal of Cancer (BJC), British Journal of Obstetrics and Gynaecology (BJOG) and the British Journal of Surgery (BJS). We searched for the words ‘pilot’ or ‘feasibility’ in the title or abstract for Medline, and in the title, abstract or keyword for the Web of Science.

We initially retrieved a total of 115 hits for the years 2000–2001, but 25 were disregarded because they were either letters (five), authors’ replies (four), duplicated hits (four), review/summary/discussion (nine), or were indirectly referring to previous pilot work (three). This resulted in a final total of 90 stud-
ies (2%) out of 4449 research papers published in the journals during this time (Table 1). Just over half of the papers indicated that further study was needed, but only four of the papers specifically stated that the pilot study was in preparation for an RCT (Ross-McGill et al. 2000; Stevinson & Ernst 2000; Bauhofer et al. 2001; Burrows et al. 2001).

We also contacted the editors of the journals asking about publication policy regarding pilot studies. Four of the journal editors said that they had no publication policy, that each submitted manuscript was considered on its own merit, one subject-specific journal editor said that they did not publish pilot studies, and two did not reply.

Objectives of an external pilot study

A clear list of objectives will add methodological rigour to a pilot study. Examples of the methodological aspects of a pilot study are taken from the four articles retrieved (Ross-McGill et al. 2000; Stevinson & Ernst 2000; Bauhofer et al. 2001; Burrows et al. 2001), together with two studies that we have been involved with, either directly as an investigator (Carfoot et al. 2002), or indirectly as an auditor (Bunn et al. 1998). Abstracts of the latter two studies are in conference proceedings, and are readily available from the authors upon request.

Sample size calculation

A major reason for conducting a pilot study is to determine initial data for the primary outcome measure, in order to perform a sample size calculation for a larger trial (Ross-McGill et al. 2000; Stevinson & Ernst 2000). This can be in the form of an estimate of the location (mean) and variability (standard deviation) of measurements for those in the control group for a continuous outcome measure, or an estimate of the proportion on the standard treatment for a categorical outcome measure. The number of patients to be included in the pilot study will depend on the parameter(s) to be estimated. A general rule of thumb is to take 30 patients or greater to estimate a parameter (Browne 1995). A conservative approach has been suggested when estimating a standard deviation, using at least an 80% upper one-sided confidence limit rather than the estimate itself (Browne 1995).

Integrity of study protocol

In preparation for a large, possibly multicentred, trial, a randomized pilot study can be treated as a dummy run (Ross-McGill et al. 2000; Burrows et al. 2001). This will enable all procedures to be put in place, including inclusion/exclusion criteria, drug

Table 1 Pilot studies published in 2000–01 in selected journals*

<table>
<thead>
<tr>
<th>Pilot study</th>
<th>BMJ</th>
<th>Lancet</th>
<th>JAMA</th>
<th>NEJM</th>
<th>BJC</th>
<th>BJOG</th>
<th>BJS</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pilot in preparation for a trial</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>7</td>
<td>4 (4)</td>
</tr>
<tr>
<td>Piloting new treatment, technique, phase I/II trials</td>
<td>5 (3)</td>
<td>11 (8)</td>
<td>4 (1)</td>
<td>3</td>
<td>28 (25)</td>
<td>5 (1)</td>
<td>7 (1)</td>
<td>63 (39)</td>
</tr>
<tr>
<td>Piloting screening programme</td>
<td>1</td>
<td>3 (2)</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>5 (2)</td>
</tr>
<tr>
<td>Piloting guidelines, educational package, patient care strategy</td>
<td>5 (1)</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>11 (1)</td>
<td></td>
</tr>
<tr>
<td>Laboratory testing of activity of compounds, e.g. in vivo or in vitro assays</td>
<td>0</td>
<td>2 (1)</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>7 (1)</td>
<td></td>
</tr>
<tr>
<td>Total pilot studies</td>
<td>11 (4)</td>
<td>17 (11)</td>
<td>7 (1)</td>
<td>3</td>
<td>33 (25)</td>
<td>10 (4)</td>
<td>9 (2)</td>
<td>90 (47)</td>
</tr>
<tr>
<td>Total number of research papers†</td>
<td>372</td>
<td>1115</td>
<td>619</td>
<td>434</td>
<td>1132</td>
<td>381</td>
<td>396</td>
<td>4449</td>
</tr>
</tbody>
</table>

*Numbers in parentheses refer to the number of studies that mentioned the need for further study as a result of the findings of the pilot study. This was not applied to the total number of research papers.

†This is an approximate total, referring to a search of the total number of journal articles containing an abstract, excluding reviews, using PubMed (National Centre for Biotechnology Information 2002).
preparation (if applicable), storage and testing of equipment and materials, training of staff in administration and assessment of the intervention. This approach was taken in preparation for a single centre midwifery trial (BEST) (Carfoot et al. 2002), enabling the investigators to confirm the enrolment procedure and pilot data collection, and determine the number of research assistants necessary to provide 24-h on-call cover.

Testing of data collection forms or questionnaires
Whether incorporated into the above or as a stand-alone study the piloting of data collection forms or questionnaires (Carfoot et al. 2002) is particularly important, especially when the patient has to self-complete the form or when several different assessors will be collecting data. This will ensure the form is comprehensible and appropriate, and that questions are well defined, clearly understood and presented in a consistent manner. Other forms such as patient information documents and consent forms could also be tested. It is important to note that testing out the administration of a questionnaire is not the same as validating the instrument or showing it to be reliable, which would require more rigorous methods (Streiner & Norman 1989).

Randomization procedure
In a pilot study it is possible to test how the randomization procedure is to work (Ross-McGill et al. 2000; Burrows et al. 2001). If sealed envelopes are to be used, then provision can be made as to how they will be prepared, where they will be stored, and how they will be administered. For example, they could be held in a hospital pharmacy and each envelope could be requested and signed for at the pharmacy window to maintain objectivity. Alternatively, a specialist clinical trials unit could be approached to provide a 24-h randomization service, or to provide a service with phone coverage from 0900 h to 1700 h. It would also enable the acceptability of the concept of randomization to the patient to be determined (Bunn et al. 1998), and the best way of providing a suitable explanation and eliciting informed consent.

Recruitment and consent
It is important to determine what the consent rate will be for patients entering the trial (Ross-McGill et al. 2000; Burrows et al. 2001; Carfoot et al. 2002), as this will have a direct impact on planning how long it will take to recruit enough patients into the trial, with consequent implications for funding. Barriers in the recruitment of both clinicians and participants to RCTs have already been highlighted, with recommendations that this aspect of the trial should be carefully planned and piloted (Ross et al. 1999). Failure to recruit sufficient numbers in a trial will reduce the statistical power, and is one of the main reasons for abandoning trials early (Ross et al. 1999).

Acceptability of intervention
In certain situations an intervention may not appeal to all patients, and it would then be wise to determine acceptability in a pilot sample. For example, the intervention may have known side-effects or be difficult to administer (Ross-McGill et al. 2000), or it may be a complementary therapy (Stevinson & Ernst 2000). This would be of particular benefit in a paediatric population when drugs may be licensed and tested in adults but not necessarily in children. The investigators of the CALICO trial (Bunn et al. 1998) adopted this approach before embarking upon a multicentre trial to introduce new oral calorie supplements to children with cystic fibrosis. They needed to be sure that the children would take the supplements, were willing to be randomized and would stick to the dietary regime for a reasonable length of time.

Selection of most appropriate primary outcome measure
The choice of the most appropriate primary outcome measure can be difficult. Issues to be considered include the reliability of the outcome and the feasibility of measurement. It may be that two or more outcomes are proposed (Bunn et al. 1998; Bauhofer et al. 2001), one of which may be deemed more clinically relevant but which may be influenced by factors other than the intervention. It would then be
prudent to examine potential outcome measures in a pilot study (Bunn et al. 1998; Bauhofer et al. 2001), possibly in conjunction with obtaining initial estimates for a sample size calculation. If several primary outcome measures were selected, then each would require estimates for a sample size calculation.

Internal pilot studies

There is considerable discussion in the statistical literature regarding the design of internal pilot studies (Wittes & Brittain 1990; Birkett & Day 1994; Denne & Jennison 1999; Wittes et al. 1999; Zucker et al. 1999; Kieser & Friede 2000). In this type of design the main study is planned and a sample size calculation is performed on the best available data. The study is started and an internal pilot is carried out on the first pre-specified number of patients entering the trial. The sample size is then recalculated from the estimates obtained from the pilot stage, and if the trial requires a larger number of patients than first thought, then this becomes the new target figure. If, on the other hand, the calculation points to less patients being needed, then the original number is maintained; that is, the number can not be less than the original estimate. The advantage of this type of design is that it allows more accurate sample size calculations without increasing the time required for conducting the full trial. An internal pilot study does not, however, allow for pre-testing the feasibility of other factors relating to a trial, as it is part of the main trial. In addition the type I error rate will be slightly inflated as the pilot and main phase of the study are being treated as if they were independent of each other, when in fact the patients from the pilot phase are included in the final analysis. An internal pilot should always be stipulated in the study protocol.

Analysis of a pilot study

The analysis of any type of pilot study should be mainly descriptive (Bunn et al. 1998; Bauhofer et al. 2001; Carfoot et al. 2002) or should focus on confidence interval estimation (Burrows et al. 2001), depending upon the objectives of the study. An external pilot is treated as a stand-alone study, and there is a question as to whether it should be analysed using hypothesis testing (Ross-McGill et al. 2000; Stevinson & Ernst 2000). Such an approach should be taken with extreme caution. It would not be appropriate to place undue significance on results from hypothesis tests, as no formal power calculations have been carried out. With such small numbers there is likely to be imbalance in pre-randomization covariates, which would need adjustment in the analysis. Moreover, the confidence interval is likely to be imprecise even when there are significant differences. Results from hypothesis testing should therefore be treated as preliminary and interpreted with caution when writing a manuscript. Most of all the temptation not to proceed with the main study when significant differences are found should be avoided.

Ensuring methodological rigour

In this paper we have provided a framework for pilot studies related to RCTs, and have suggested several methodological reasons why a pilot study might be conducted. Two of the authors (GL, PW) have served on various local hospital research and development committees and an ethics committee where pilot studies are often proposed, although specific objectives are rarely formally presented. A more methodologically rigorous framework would safeguard against pilot studies being conducted simply because of small numbers of available patients (Williamson et al. 2000), when a multicentre trial may be more appropriate. In some situations patients involved in an external pilot are later included in the main study, to make savings in recruitment. We suggest that this is not a good idea. If these patients were to be included in the larger study, then the decision to proceed with the main study would not be made independently of the results of the pilot study, so there could be selection bias, as well as an inflated type I error rate.

Randomized pilot studies appear in meta-analyses as primary studies in Cochrane systematic reviews (Hazell et al. 2002; Horn & Limburg 2002). Work on systematic reviews has emphasized the problem of publication bias (Begg & Berlin 1988). We postulate that this will probably also apply to pilot studies, resulting in work with important negative findings never reaching publication. In 1995 Chalmers et al. (1995) called for all trials to be registered on a national health register. The UK National Research
Register (Update Software 2002) was subsequently extended and now contains over 84,000 research projects, including many pilot studies, collected from a variety of sources. It would be interesting to audit how many of the pilot studies do eventually reach publication. A methodologically rigorous study would stand more chance of publication. There is also a strong ethical argument for methodological rigour. Clear aims and objectives in a well-conducted pilot study would promote the necessity of the study, establish a framework within which to work and therefore ensure that it is scientifically valid.

**Recommendations**

- Pilot studies should have a well-defined set of aims and objectives to ensure methodological rigour and scientific validity.
- Participants in an external pilot should not later be included in the main study to make savings in recruitment, because then the decision to proceed with the main study would not be made independently of the results of the pilot study.
- The analysis of a pilot study should be mainly descriptive or should focus on confidence interval estimation.
- Results from hypothesis testing should be treated as preliminary and interpreted with caution, as no formal power calculations have been carried out.
- The temptation not to proceed with the main study when significant differences are found should be avoided.

**References**


http://www.update-software.com/Cochrane/default.htm


http://www.update-software.com/national/


