Database recruitment: a solution to poor recruitment in randomized trials?

Tracy Stuardi*, Helen Cox and David J Torgerson

Department of Health Sciences, University of York, York, UK.

*Correspondence to Tracy Stuardi, University of York, ARRC Building 2nd Floor, York YO10 5DD, UK; E-mail: tls504@york.ac.uk

Received 12 May 2010; Revised 12 October 2010; Accepted 12 November 2010.

Background. Achieving sample size is imperative to obtaining sufficient power to detect potential effects in health care research, yet many research studies are prone to under-recruitment. Not only does this create problems with power but also it contributes to research extensions, additional costs and delayed results. To combat this problem, one increasingly used technique is database recruitment, a method of searching the electronic medical records system for potential research participants.

Objective. We discuss the advantages and disadvantages of identifying potential research participants using database recruitment with particular reference to primary care.

Conclusion. Database recruitment is a relatively simple and affordable means to recruit large numbers of patients in a timely manner; however, it is not without limitations.

Keywords. Recruitment.

Background

A number of issues make recruitment to randomized controlled trials (RCTs) problematic, including the increasing difficulty in obtaining regulatory approvals, which are largely beyond the researcher’s control. On the other hand, some issues, particularly approaching potential participants, may be solved via trial design. Trials may enlist the services of GPs to recruit patients during the consultation process, known as sequential or opportunistic recruitment; however, a rule of thumb exists whereby 80% of patients are recruited from 20% of sites. For example, a back pain trial looking at the effects of physiotherapy-supervised exercise recruited 60% of its patients from a single GP practice. With >80 GPs involved, the recruitment of three patients each would have been sufficient to meet the target of 300 patients, yet many GPs either recruited no patients or one to two patients. Although the authors do not explain why so many GPs failed to recruit patients, a review by the Cochrane Collaboration indicates that the design and conduct of a trial as well as a GP’s perception of the trial’s effect may affect the GP’s willingness to recruit. Additionally, GPs may emphasize a lack of time as a reason for not recruiting patients.

Several studies reveal that RCTs consistently fail to reach their target sample size within the original timescale. Poor recruitment may result in a variety of undesirable consequences, such as, Type II error, whereby researchers conclude no difference between groups which in fact differ. For example, the Shoulder Acute Pain in Primary Healthcare study (SAPPHIRE) had an original sample size of 500 participants, which was deemed adequate to observe a clinically significant difference at approximately one-third of a standard deviation between groups. However, poor recruitment led to the enrolment of only 200 participants, and although the groups varied by one-third of a standard deviation, it was not statistically significant. Consequently, the effectiveness of that intervention remains uncertain.

A second undesirable consequence of poor recruitment is project extension, which ultimately increases costs. Project extensions may incur both financial costs, in terms of labour expenses, and clinical costs, in terms of the delayed application of study results. Thus, patients may either be denied effective treatment or continue to be exposed to ineffective treatments for an extended period. Finally, poor recruitment may result in the use of inferior evidence by policymakers to make health care decisions. In this paper, we examine an approach that we believe could be more widely used to improve recruitment to RCTs.

Database recruitment

Lists or files of patients have always existed in GP and hospital practices; however, the age of computer-based
databases has created a mechanism by which those lists can be easily searched for potential participants. Quite simply, database recruitment is the identification of those potential participants based on specific electronic medical records (EMR) search/selection criteria derived from the inclusion and exclusion criteria of a study. Such a search can be performed by appropriate National Health Service personnel, on behalf of the relevant clinician. Once identified, potential participants can then be invited to participate in a study.

Table 1 compares recruitment, in a sample of similar trials, using either database or sequential recruitment. Specifically, the Yoga for low back pain study exceeded its original recruitment target as did a study of exercise for low back pain.9,11 Both of these studies conducted database searches for patients experiencing recent (<18 and <6 months, respectively) episodes of low back pain. All potentially eligible patients were sent screening questionnaires to determine if the nature of their low back pain met inclusion criteria. The database searches effectively enabled recruitment of the target samples despite the absence of the minimum low back pain score required for inclusion by the majority (61%10) of participants contacted. In contrast, the UK Back pain exercise and manipulation (BEAM) study and the York back pain study employed GPs to approach patients who were consulting with an ongoing episode of back pain.2,10 Both of these studies struggled to recruit sufficient numbers of patients. Similarly, the Randomized evaluation of calcium or vitamin D (RECORD) trial recruited patients from secondary care fracture clinics to evaluate the use of calcium supplements and struggled to recruit sufficient numbers.12 In comparison, the York primary care trial of calcium and vitamin D piloted database recruitment and found that the recruitment rate was 50% below the expected rate.11 Having this information allowed the researchers to double the number of databases searched.

**Advantages**

The advantages of this type of recruitment may seem fairly obvious. First, studies are able to identify large numbers of potential participants in a matter of minutes and in a relatively simple manner (i.e. computer search). Second, having identified potential participants, study materials can be delivered quickly and easily as in the form of a bulk mail out.1 Third, the process can be scheduled to coincide (or not) with a particular time of year (i.e. holidays) and makes staff needs apparent. The theme here is time; database recruitment provides a high-yield minimal investment strategy. Time savings, in the form of decreased staff hours, may translate into cost savings and may result in recruitment on schedule and error avoidance, which are both common occurrences.1 Additionally, time savings may attract the participation of GPs or surgeries. A study by Embi et al.15 indicates that GPs are receptive to the idea of using EMR systems to facilitate research. Database recruitment has the additional advantage of eliminating subjectivity that may accompany GP referral or nursing recruitment and create bias. Database searches should identify all potentially eligible participants, thereby providing an equal opportunity for anyone to participate.

### Table 1 Method of recruitment comparison

<table>
<thead>
<tr>
<th>Study</th>
<th>Database recruitment</th>
<th>Sequential recruitment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yoga for low back pain9</td>
<td>On time and 18% above target</td>
<td>UK BEAM9</td>
</tr>
<tr>
<td>Manchester low back pain10</td>
<td>On time and 40% above target</td>
<td>Over time and on target</td>
</tr>
<tr>
<td>York calcium and vitamin D trial11</td>
<td>On time and 16% above target</td>
<td>York back pain study2</td>
</tr>
<tr>
<td>ACIBS13</td>
<td>Ahead of time and 6% above target</td>
<td>MRC RECORD12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>38% under target</td>
</tr>
</tbody>
</table>

quick and inexpensive. Using piloted recruitment rates, where possible, is generally the most reliable way of calculating the list size. For example, a pilot study of acupuncture for irritable bowel syndrome (IBS) recruited and randomized 32 participants from a combined list size of 20 300 patients, giving an estimated recruitment rate of 0.158%.14 Using this information and full-scale trial sample size (220), the combined list size necessary was estimated to be 140 000 patients. Where piloted recruitment rates are unavailable, it may be advisable to select a single database (i.e. one GP surgery) from which to generate those numbers. For example, the York primary care trial of calcium and vitamin D piloted database recruitment and found that the recruitment rate was 50% below the expected rate.11 Having this information allowed the researchers to double the number of databases searched.
Disadvantages

A key disadvantage of database recruitment is restriction to application based on condition. For example, database recruitment may be impractical for a trial comparing surgical to non-surgical interventions for humerus fractures. Trials for acute conditions such as this are more likely to benefit from sequential recruitment as patients present to the hospital or fracture clinic. Therefore, database recruitment is relatively limited to identifying patients with chronic conditions. Secondly, the large number of potential patients identified by database recruitment includes a number of patients who do not meet the inclusion criteria and may not be aware of the diagnosis. For example, the calcium and vitamin D study identified all women >70 years. Database recruitment in that age range may include patients with dementia or who are recently deceased and whose families may be upset by receiving study materials. One way to pre-empt the contact of recently deceased or inappropriate patients is to ask the practice manager or GP to review and remove names prior to post. The inevitability that some inappropriate patients will be contacted should be discussed in trial management meetings to ensure that an agreed plan of action is in place.

With regard to preparing a database search strategy, it is important to identify whether practices use different terms such as Read Codes or Systematized Nomenclature of Medicine. Search strategies should be identical where possible to ensure homogeneity among recruited patients; however, the strategies should account for possible differences in software. Considering that the searches may be executed by a different person at each practice, a member of the research team should provide instruction at each site. Additionally, the search strategy should incorporate codes for alternate names and inclusion criteria components. For example, the Acupuncture for Irritable Bowel Syndrome (ACIBS) multicentre trial recruited patients with IBS, which may appear in a database as IBS or IBS subtypes such as IBS-D, IBS-C or IBS-M. A search for IBS would not necessarily identify the IBS subtypes. Careful attention to the search strategy and software will determine whether a single code hierarchically captures all the IBS-related diagnoses or if separate codes are necessary. Failure to pay attention to the search may create heterogeneity among database searches. As mentioned previously, the search may capture patients who fulfil the exclusion criteria; therefore, the codes for exclusion criteria should be included in the search where possible. Exclusion criteria not included in the search (i.e. gastrointestinal surgery in past 6 months) should be included in the eligibility questionnaire to complete the screening process.

Another potential disadvantage is that database recruitment is so rapid that researchers and clinicians may be overwhelmed during the set-up and initial recruitment phases. Consequently, this may cause delayed randomization and/or access to the allocated intervention. This is particularly important for studies that have limited resources/funding. For example, the ACIBS study employed nine acupuncturists to treat 110 patients. If recruitment commenced all at once, there would have been a shortage of treatment slots. Therefore, recruitment was staggered over a 6-month period to ensure treatment availability. Increased workload during the set-up and initial recruitment phases can be accounted for in grant applications by allowing for more staff hours during these phases.

Socio-demographic differences can also affect recruitment rates. We piloted a trial of yoga for low back pain in York using one GP practice, in order to gain information on recruitment rates for a full-scale trial. H. Tilbrook, H. Cox, C. Hewitt, et al. Unpublished data, 2010. Recruitment rates for the full-scale multi-centred trial varied considerably depending of the geographic location. The recruitment rate in York was 0.2% of the practice list size, whereas in Manchester, it was 0.07%. Since the piloting was done in York, we initially under-recruited in Manchester and had to run a ‘second wave’ of recruitment. Ideally, multi-centre trials should pilot recruitment in each area/region; however, this may not be possible due to time and funding constraints.

Ethics and confidentiality

The use of EMRs for database recruitment raises some confidentiality issues. Clinical research depends on the participation of suitable volunteers; yet identifying potential volunteers often requires accessing health care information. By nature, health care information contains a combination of personal demographic details and medical complaints that are confidential. For the purpose of providing care, clinical staff, such as GPs, have access to confidential health care records. On the other hand, researchers are independent from the clinical staff and require a patient’s consent to access confidential information. Since the boundary of confidentiality is maintained by the clinical staff, database searches should be executed by those with appropriate access to EMRs (i.e. practice staff or research nurse) with guidance from the trial team. To maintain confidentiality, only the practice staff or research nurse should have access to the database-generated list throughout the post and invitation process. As mentioned previously, the database-generated list should be reviewed by a GP or practice manager to remove inappropriate patients prior to post. The responsibility of the patient to reply with interest to a postal invitation is characteristic of opt-in research. Depending on the research design, patients
may opt-in by returning a consent form, questionnaire and/or phoning the practice for an appointment. Studies may be more likely to receive ethical approval using an opt-in approach as opposed to an opt-out approach wherein researchers contact patients directly. Papers by Hewison and Haines and Angus et al. discuss the controversy surrounding opt-in versus opt-out designs with particular emphasis on the potential for opt-in study quality to suffer as a result of added barriers to participation and possible response bias. Additionally, decisions about opt-in versus opt-out designs are being made without sufficient knowledge of patients’ preferences and understanding of the research approaches. While the debate continues, ethics applications should clearly describe the recruitment process and highlight the attention to confidentiality and the means by which it is maintained. During the trial design phase, budgetary allowances should be calculated for resources to fund primary care practices and/or research nurses for their work on database recruitment.

Discussion

Poor recruitment is a major threat to RCTs. As a means to overcome this problem, we have described database recruitment and its applications. With regard to applicability, recruitment to studies of medical conditions with acute onsets and/or temporal remission (i.e. sore throat) is not likely to benefit from database recruitment. On the other hand, studies for chronic conditions such as arthritis, mental illness and back pain; smoking cessation and alcohol abuse; or screening and public health interventions are more likely to benefit from database recruitment.

While the advantages of database recruitment make it an attractive strategy and the disadvantages are relatively easy to overcome, there are a few additional items to consider. By nature, database recruitment may select different kinds of patients compared with sequential recruitment. The RECORD trial, for example, recruited patients with a recent fracture who may respond differently to the intervention compared with patients who have general fracture risk factors. Similarly, the UK BEAM trial recruited patients who had an ongoing episode of back pain, while the Manchester back pain trial and the yoga trial likely recruited a higher proportion of chronic patients. While database recruitment is a relatively straightforward procedure in the UK, countries with mixed health care systems may experience difficulty in implementing this strategy. Lastly, ethics rules and regulations regarding data protection change continuously and should be monitored. The current environment seems favourable for primary care settings but may be more difficult for secondary or tertiary care.

Summary points

Database recruitment is becoming a popular technique to overcome recruitment difficulties.

Advantages of database recruitment include the speed and ease with which the procedure is carried out as well as the associated cost savings.

Disadvantages include: variations in databases, potential for heterogeneity, errors in non-systematic pre-defined searches and other unforeseen circumstances.

Declaration

Funding: none.
Ethical approval: none.
Conflict of interest: none.

References


