

- 7 Forman D, Rider L, eds. *Yorkshire Cancer Registry report 1996*. Leeds: Yorkshire Cancer Organisation, 1996.
- 8 Campbell AJ, Buchner DM. Unstable disability and the fluctuations of frailty. *Age Ageing* 1997;26:315-8.
- 9 Balducci L, Mowrey K, Parker M. Pharmacology of antineoplastic agents in older patients. In: Balducci L, Lyman GH, Ershler WB, eds. *Geriatric oncology*. Philadelphia: JB Lippincott Co, 1992:169-80.
- 10 Shank WA, Jr., Balducci L. Recombinant hemopoietic growth factors: comparative hemopoietic response in younger and older subjects. *J Am Geriatr Soc* 1992;40:151-4.
- 11 Begg CB, Carbone PP. Clinical trials and drug toxicity in the elderly. The experience of the Eastern Cooperative Oncology Group. *Cancer* 1983;52:1986-92.
- 12 Evans WK, Radwi A, Tomiak E, Logan DM, Martins H, Stewart DJ, et al. Oral etoposide and carboplatin. Effective therapy for elderly patients with small cell lung cancer. *Am J Clin Oncol* 1995;18:149-55.
- 13 Morandi U, Stefani A, Golinelli M, Ruggiero C, Brandi L, Chiapponi A, et al. Results of surgical resection in patients over the age of 70 years with non small-cell lung cancer. *Eur J Cardiothorac Surg* 1997;11:432-9.
- 14 Bufalari A, Ferri M, Cao P, Cirocchi R, Bisacci R, Moggi L. Surgical care in octogenarians. *Br J Surg* 1996;83:1783-7.
- 15 Farrow DC, Hunt WC, Samet JM. Temporal and regional variability in the surgical treatment of cancer among older people. *J Am Geriatr Soc* 1996;44:559-64.
- 16 Pignon T, Gregor A, Schaake KC, Roussel A, Van Glabbeke M, Scalliet P. Age has no impact on acute and late toxicity of curative thoracic radiotherapy. *Radiother Oncol* 1998;46:239-48.
- 17 Pignon T, Horiot JC, Bolla M, van Poppel H, Bartelink H, Roelofsens F, et al. Age is not a limiting factor for radical radiotherapy in pelvic malignancies. *Radiother Oncol* 1997;42:107-20.
- 18 Rostom AY, Pradhan DG, White WF. Once weekly irradiation in breast cancer. *Int J Radiat Oncol Biol Phys* 1987;13:551-5.
- 19 Olmi P, Ausili-Cefaro G. Radiotherapy in the elderly: a multicentric prospective study on 2060 patients referred to 37 Italian radiation therapy centers. *Rays* 1997;22:53-6.
- 20 Robertson JF, Todd JH, Ellis IO, Elston CW, Blamey RW. Comparison of mastectomy with tamoxifen for treating elderly patients with operable breast cancer. *BMJ* 1988;297:511-4.
- 21 Balducci L, Extermann M, Fentiman I, Monfardini S, Perrone F. Should adjuvant chemotherapy be used to treat breast cancer in elderly patients (≥ 70 years of age)? *Eur J Cancer* 1997;33:1720-4.
- 22 Cleary JF, Carbone PP. Palliative medicine in the elderly. *Cancer* 1997;80:1335-47.
- 23 Fletcher A. Screening for cancer of the cervix in elderly women. *Lancet* 1990;335:97-9.
- 24 Age Concern. *Not at my age: why the present breast screening system is failing women aged 65 or over*. London: Age Concern England, 1996.
- 25 Van Dijk JA, Verbeek AL, Beex LV, Hendriks JH, Holland R, Mravunac M, et al. Mammographic screening after the age of 65 years: evidence for a reduction in breast cancer mortality. *Int J Cancer* 1996;66:727-31.
- 26 Haigney E, Morgan R, King D, Spencer B. Breast examinations in older women: questionnaire survey of attitudes of patients and doctors. *BMJ* 1997;315:1058-9.
- 27 Yellen SB, Cella DF, Leslie WT. Age and clinical decision making in oncology patients. *J Natl Cancer Inst* 1994;86:1766-70.
- 28 Mead GE, Pendleton N, Pendleton DE, Horan MA, Bent N, Rabbit P. High technology medical interventions: what do older people want? *J Am Geriatr Soc* 1997;45:1409-11.
- 29 Given CW, Given BA, Stommel M. The impact of age, treatment, and symptoms on the physical and mental health of cancer patients. A longitudinal perspective. *Cancer* 1994;74:2128-38.
- 30 Weinrich SP, Weinrich MC. Cancer knowledge among elderly individuals. *Cancer Nurs* 1986;9:301-7.
- 31 Wetle T. Age as a risk factor for inadequate treatment. *JAMA* 1987;258:516.
- 32 Newcomb PA, Carbone PP. Cancer treatment and age: patient perspectives. *J Natl Cancer Inst* 1993;85:1580-4.
- 33 Pearlman RA, Uhlmann RF. Quality of life in chronic diseases: perceptions of elderly patients. *J Gerontol* 1988;43:M25-30.
- 34 Hazzard WR, Woolard N, Regenstein DI. Integrating geriatrics into the subspecialties of internal medicine: the Hartford Foundation/American Geriatrics Society/Wake Forest University Bowman Gray School of Medicine initiative. *J Am Geriatr Soc* 1997;45:638-40.
- 35 Sainsbury R, Rider L, Smith A, MacAdam A. Does it matter where you live? Treatment variation for breast cancer in Yorkshire. The Yorkshire Breast Cancer Group. *Br J Cancer* 1995;71:1275-8.
- 36 Calman K, Hine D. *A policy framework for commissioning cancer services. A report by the Expert Advisory Group on Cancer to the Chief Medical Officers of England and Wales*. 1995. London: Department of Health, 1995.
- 37 Selby P, Gillis C, Haward R. Benefits from specialised cancer care. *Lancet* 1996;348:313-8.
- 38 Monfardini S. What do we know on variables influencing clinical decision-making in elderly cancer patients? *Eur J Cancer* 1996;32A:12-4.
- 39 Fentiman IS. Are the elderly receiving appropriate treatment for cancer? *Ann Oncol* 1996;7:657-8.
- 40 Coebergh JW. Significant trends in cancer in the elderly. *Eur J Cancer* 1996;32A:569-71.

(Accepted 6 April 1999)

Methods in health services research

Interpreting the evidence: choosing between randomised and non-randomised studies

Martin McKee, Annie Britton, Nick Black, Klim McPherson, Colin Sanderson, Chris Bain

This is the first of four articles

Correspondence to:
Martin McKee
m.mckee@lshtm.
ac.uk

Series editor:
Nick Black

continued over

BMJ 1999;319:312-5

website
extra

Further references
are listed on the
BMJ's website

www.bmj.com

Evaluations of healthcare interventions can either randomise subjects to comparison groups, or not. In both designs there are potential threats to validity, which can be external (the extent to which they are generalisable to all potential recipients) or internal (whether differences in observed effects can be attributed to differences in the intervention). Randomisation should ensure that comparison groups of sufficient size differ only in their exposure to the intervention concerned. However, some investigators have argued that randomised controlled trials (RCTs) tend to exclude, consciously or otherwise, some types of patient to whom results will subsequently be applied. Furthermore, in unblinded trials the outcome of treatment may be influenced by practitioners' and patients' preferences for one or other intervention. Though non-randomised studies are less selective in terms of recruitment, they are subject to selection bias in allocation if treatment is related to initial prognosis.

These issues have led to extensive debate, although empirical evidence is limited. This paper is a brief summary of a more detailed review¹ of the impact of these potential threats.

Summary points

Treatment effects obtained from randomised and non-randomised studies may differ, but one method does not give a consistently greater effect than the other

Treatment effects measured in each type of study best approximate when the exclusion criteria are the same and where potential prognostic factors are well understood and controlled for in the non-randomised studies

Subjects excluded from randomised controlled trials tend to have a worse prognosis than those included, and this limits generalisability

Subjects participating in randomised controlled trials evaluating treatment of existing conditions tend to be less affluent, educated, and healthy than those who do not; the opposite is true for trials of preventive interventions

Threats to validity of evaluative research and possible solutions

Type of validity	Threatening factor	Proposed solution
Internal	Allocation bias (risk of confounding)	Randomisation
		Risk adjustment and subgroup analysis (analysis)
	Patient preference	Preference arms or adjustment for preference (design)
External	Exclusions (eligibility criteria)	Expand inclusion criteria
	Non-participation (centres/practitioners)	Multicentre, pragmatic design
	Not invited (practitioner preference or administrative oversight)	Encourage practitioners to invite all eligible patients
	Non-participation (patients)	Less rigorous consent procedures

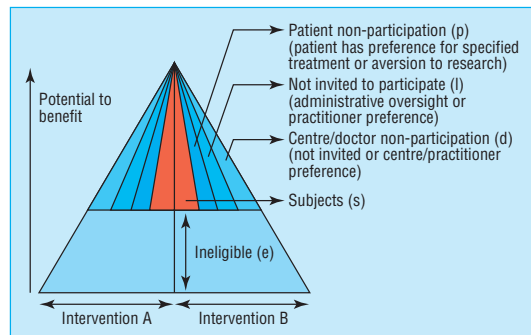


Fig 1 Differences in inclusion and participation. Shaded areas represent the study population

Nature of the evidence

The review focused on threats to internal and external validity of evaluations of effectiveness and on the strategies proposed to overcome them (table). Various factors act through their effect on the distribution of the potential to benefit among different groups. This can be illustrated schematically (fig 1). The reference population is defined by an envelope, represented here as a triangle but potentially taking many shapes. At some point, a threshold is reached, below which the overall risks outweigh the benefits. As patients are excluded or do not participate, the study population becomes a progressively smaller subset of the reference population, in principle increasing the scope for selection bias and raising the question of whether it is valid to apply the results obtained to the reference population.

We used systematic reviews to explore the potential and actual importance of factors that lead to selective recruitment, examining four questions:

- Do non-randomised studies give systematically different measurements of treatment effect from RCTs?
- Are there systematic differences between the subjects included in or excluded from studies, and do these influence the measured treatment effect?
- To what extent is it possible to overcome known or unknown baseline differences between groups that are not allocated randomly?
- How important are patients' preferences for an intervention and, if patients are randomised to a treatment they would not choose, how does this affect their outcome?

Findings

Comparing results of RCTs and non-randomised studies

Eighteen papers were identified where a single intervention was evaluated by both methods (a full list

is available on the *BMJ's* website). A review was published just after our original report; on the basis of eight comparisons it found that, on average, non-randomised studies overestimate effect size.² In contrast, of the seven studies in our review where the two methods detected effects in the same direction, in three the effect size was greater in the randomised trial and in four it was greater in the non-randomised study. The key finding in our study is that neither method consistently gave larger estimates of treatment effect.

In addition to chance, there are several potential explanations for different measurements of treatment effects. The overall impact will reflect the relative importance of each issue in a particular case. A randomised controlled trial may produce a greater effect if the patients enrolled in it receive higher quality care or are selected so that they have greater capacity to benefit than patients in non-randomised studies. But it may produce a lower estimate of treatment effect for several reasons:

- In non-randomised studies, patients tend to be allocated to treatments that are correctly considered most appropriate for their individual circumstances;
- Exclusions from a RCT create a sample with less capacity to benefit than in a non-randomised study;
- An unblinded RCT fails to capture patients with strong preferences for a particular treatment who show an enhanced response to treatment;
- Non-randomised studies of preventive interventions include disproportionate numbers of individuals who, by virtue of their health related behaviour, have greater capacity to benefit;
- Publication bias leads to negative results being less likely to be published from non-randomised studies than from RCTs.

The limited evidence indicates that the results of non-randomised studies best approximate to results of RCTs when both use the same exclusion criteria and when potential prognostic factors are well understood, measured, and appropriately controlled in non-randomised studies.³

In summary, the results of RCTs and non-randomised studies of similar patients may not, after adjustment, be substantially different in relative effect size. Any variations are often no greater than those between different RCTs or among non-randomised studies. Differences in effect sizes could be due to chance or differences in populations studied, timing, or nature of the intervention.

Exclusions

Randomised controlled trials vary widely in their inclusiveness. Medical reasons cited for exclusion from trials include a high risk of adverse effects and belief

London School of Hygiene and Tropical Medicine, London WC1E 7HT

Martin McKee
professor of European public health

Annie Britton
research fellow

Nick Black
professor of health services research

Klim McPherson
professor of public health epidemiology

Colin Sanderson
senior lecturer in health services research

University of Queensland Medical School, Brisbane 4006, Australia

Chris Bain
reader in social and preventive medicine

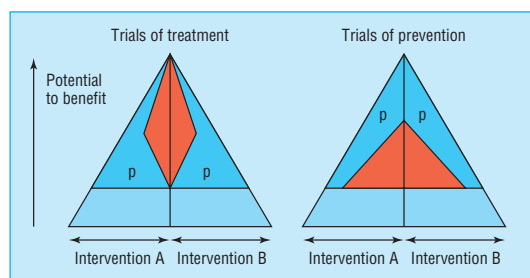


Fig 2 Effect of differences in participation in trials of prevention and of treatment. p=eligible non-participants; shaded areas represent the study population

that benefit, or lack of it, has already been established for some groups.

Scientific reasons include greater precision in estimating treatment effects by having a homogeneous sample,⁴ and reduced risk of bias by excluding individuals most likely to be lost to follow up.⁵ In addition, many RCTs have blanket exclusions,⁶ the reasons for which are often unstated, of categories of patients such as the elderly, women, and ethnic minorities.

Few studies have examined differences in prognostic factors between included and excluded patients, but some have used clinical databases to examine this.^{7,8} The patients included in such databases tended to have a poorer prognosis than those in trials: in one study, a subset selected to meet eligibility criteria of RCTs produced treatment effects of similar size to those obtained from RCTs.³

Participation

Evaluative research is undertaken predominantly in university or teaching centres, but non-randomised studies are more likely than RCTs to include non-teaching centres, and criteria for participation in RCTs may include the achievement of a specified level of clinical outcome. The available evidence suggests that this may exaggerate the measured treatment effect.⁹

Most evaluative studies fail to document adequately the characteristics of eligible patients who do not participate. The effect of non-participation differs between RCTs that evaluate interventions designed to treat an existing condition and those directed at preventing disease (fig 2).¹⁰ Participants in the former tend to be less affluent, less educated, and more severely ill than eligible patients who do not participate.¹¹ In contrast, participants in RCTs evaluating preventive interventions tend to be more affluent, better educated, and more likely to have adopted a healthy lifestyle than patients who decline.¹² On the basis of the evidence from the comparisons discussed earlier, it is plausible that low participation in RCTs of treatment may exaggerate treatment effects by including more skilful practitioners and subjects with a greater capacity to benefit, while RCTs of prevention may underestimate effects as participants have selectively less capacity to benefit.

Impact of patients' preferences

There is little empirical research on the impact on outcome of patients' preferences. The four studies that attempted to measure preference effects either were small or have yet to report full results.¹³⁻¹⁶ In theory, preference could have an important impact on results

of RCTs, especially where the true effect is small. Such effects could account for some observed differences between results of RCTs and non-randomised studies. There are methods that may detect preference effects reliably; though these may contribute to understanding this phenomenon, none provides a complete answer.¹⁷ This is mainly because randomisation between preferring a treatment and not is impossible, and confounding may bias any observed comparison.

Adjustment for baseline differences in non-randomised studies

Despite the evidence that the results of RCTs and non-randomised studies are often similar, differences in baseline prognostic factors clearly can be important. Absence of randomisation can produce groups that differ in important ways, and it is necessary to consider whether it is possible to adjust for such differences. Adjustment for imbalance in baseline prognostic factors between arms of non-randomised studies commonly changes the size of the measured treatment effect, but such changes are often small and inconsistent.¹

Overall, the limited evidence suggests that differences in the populations studied by RCTs and non-randomised studies are likely to be of at least as much importance in explaining any differences and that the two methods should be compared only after patients not meeting eligibility criteria for the RCT are excluded.

Recommendations

A large, inclusive, fully blinded RCT incorporating appropriate subgroup analysis is likely to provide the best possible evidence of effectiveness, but there will always be circumstances in which randomisation, especially on an inclusive basis, is unethical or impractical.¹⁸ In circumstances where there are genuine reasons for not randomising,¹⁹ non-randomised studies can provide useful evidence. In such studies, adjustment for baseline imbalances should always be attempted, as rigorously and extensively as possible, and the procedures should be reported explicitly to help readers' evaluations. However, adjustment cannot be relied on to approximate the prognostic balance of randomisation, given unknown or unmeasurable confounding.

Investigators conducting evaluative research (using any design) must improve the quality of reporting. Authors should define the population to whom they expect their results to be applied; what has been done to ensure that the study population is representative of this wider population, and any evidence of how it differs; whether centres that participated differ from those that declined; and the numbers and characteristics of patients eligible to be included who either were not invited to do so or were invited and declined.

The findings of such studies have implications for the way in which evidence is interpreted. When faced with data from any source, whether randomisation has been used or not, it is important first to pursue alternative (non-causal) explanations thoroughly and examine the possible influence of chance, bias, and confounding, perhaps using sensitivity analyses where feasible.

Where only non-randomised data are available, the potential for allocation bias should be considered and any attempts at risk adjustment should be assessed.

Where only randomised trials are found, preference effects should also be considered. To obtain an uncontaminated estimate of the physiological effect of a treatment, RCTs should be blind to everyone involved, but for many interventions this will be impossible. Also, the advantages of narrowing inclusion criteria to ensure high participation in RCTs should be balanced by the potential need for subgroup analysis. It should not be assumed that a summary result applies to all potential patients.

When both randomised and non-randomised studies have been conducted it is important to ascertain whether estimates of treatment effect are consistent for patients at similar risk across studies. If so, it may be reasonable to accept the results of the less exclusive non-randomised studies. Differences in results cannot be assumed to be solely due to the presence or lack of randomisation—differences in study populations, characteristics of the intervention, and the effects of patients' preferences may also affect the results.

Whichever design is used, generalisability needs attention. One approach involves examining the relation between reduction in relative risk (as a measure of effect size) against the percentage of events in the control arm (as an indirect measure of inclusiveness)²⁰; this is sometimes referred to as metaregression.²¹ Where sufficient data are available from RCTs, it may be possible to identify separate measures of benefit and harm. If, as has been shown for giving anticoagulants to prevent stroke, the percentage reduction in relative risk remains constant at all levels of severity and the increase in absolute risk of an adverse effect remains constant, the reduction in absolute risk for a given patient can then be estimated.²²

In conclusion, RCTs and non-randomised studies can provide complementary evidence—but it is important that clinicians using this evidence are aware of the strengths and weaknesses of each method.

This article is adapted from *Health Services Research Methods: A Guide to Best Practice*, edited by Nick Black, John Brazier, Ray Fitzpatrick, and Barnaby Reeves, published by BMJ Books in 1998.

Funding: This work was supported by a grant from the NHS Health Technology Assessment Programme.

Competing interests: None declared.

- 1 Britton A, McKee M, Black N, McPherson K, Sanderson C, Bain C. Choosing between randomised and non-randomised studies: a systematic review. *Health Technol Assess* 1998;2(13).
- 2 Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. *BMJ* 1998;317:1185-90.
- 3 Horvitz RI, Viscoli CM, Clemens JD, Sadock RT. Developing improved observational methods for evaluating therapeutic effectiveness. *Am J Med* 1990;89:630-8.
- 4 Yusuf S, Held P, Teo KK. Selection of patients for randomized controlled trials: implications of wide or narrow eligibility criteria. *Stat Med* 1990;9:73-86.
- 5 Haynes RB, Dantes R. Patient compliance and the conduct and interpretation of therapeutic trials. *Controlled Clin Trials* 1987;8:12-9.
- 6 Gurwitz JH, Nananda F, Avorn J. The exclusion of the elderly and women from clinical trials on acute myocardial infarction. *JAMA* 1992;268:1417-22.
- 7 Jones RH, Kesler K, Phillips HR, Mark DB, Smith PK, Nelson CL, et al. Long-term survival benefits of coronary artery bypass grafting and percutaneous transluminal angioplasty in patients with coronary artery disease. *J Thorac Cardiovasc Surg* 1996;111:1013-25.
- 8 Hartz AJ, Kuhn EM, Pryor DB, Krakauer H, Young M, Heudeberg, et al. Mortality after coronary angioplasty and coronary artery bypass surgery (the National Medicare Experience). *Am J Cardiol* 1992;70:179-85.
- 9 Stukenborg GJ. Comparison of carotid endarterectomy outcomes from randomized controlled trials and medicare administrative databases. *Arch Neurol* 1997;54:826-32.
- 10 Hunninghake DB, Darby CA, Probstfield JL. Recruitment experience in clinical trials: literature summary and annotated bibliography. *Controlled Clin Trials* 1987;8:6S-30S.
- 11 Barofsky I, Sugarbaker PH. Determinants of patients nonparticipation in randomised clinical trials for the treatment of sarcomas. *Cancer Clin Trials* 1979;2:237-49.
- 12 Davies G, Pyke S, Kinmonth AL. Effect of non-attenders on the potential of a primary care programme to reduce cardiovascular risk in the population. *BMJ* 1994;309:1553-6.
- 13 McKay JR, Alterman AI, McLellan T, Snider EC, O'Brien CP. Effect of random versus nonrandom assignment in a comparison of inpatient and day hospital rehabilitation for male alcoholics. *J Consulting Clin Psychol* 1995;63:70-8.
- 14 Nicolaides K, de Lourdes Brizot M, Patel F, Snijders R. Comparison of chorionic villus sampling and amniocentesis for fetal karyotyping at 10-13 weeks' gestation. *Lancet* 1994;344:435-9.
- 15 Torgerson DJ, Klaber-Moffett J, Russell IT. Patient preferences in randomised trials: threat or opportunity. *J Health Serv Res Policy* 1996;1:194-7.
- 16 Fallowfield LJ, Hall A, Maguire GP, Baum M. Psychological outcomes of different treatment policies in women with early breast cancer outside a clinical trial. *BMJ* 1990;301:575-80.
- 17 McPherson K, Britton AR, Wennberg JE. Are randomized controlled trials controlled? Patient preferences and unblind trials. *J Roy Soc Med* 1997;90:652-6.
- 18 Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ* 1996;312:1215-8.
- 19 McPherson K. The Cochrane Lecture. The best and the enemy of the good: randomised controlled trials, uncertainty, and assessing the role of patient choice in medical decision making. *J Epidemiol Commun Health* 1994;48:6-15.
- 20 Davey Smith G, Song F, Sheldon T. Cholesterol lowering and mortality: the importance of considering initial risk. *BMJ* 1993;206:1367-73.
- 21 Holme I. Cholesterol reduction and its impact on coronary artery disease and total mortality. *Am J Cardiol* 1995;76:10-7C.
- 22 Glasziou PP, Irwig LM. An evidence based approach to individualising treatment. *BMJ* 1995;311:1356-9.



Available from the BMJ bookshop (www.bmjbookshop.com)

Email submissions from outside the United Kingdom

We are now offering an email submission service for authors from outside the UK. The address is papers@bmj.com

Ideally our email server would link seamlessly with our manuscript tracking system, but for now it does not, which is why we are offering the service only to authors outside the UK. Most post in the UK arrives the next day, so UK authors have the least to gain in speed of delivery from email delivery. As soon as our systems improve we will invite email submissions from everyone.

If you choose to send your submission by email please would you send the text and any tables and figures as attached files, together with a covering letter giving all your contact details (postal

address, phone, fax, and email address). We can read files created with most word processing, graphics, and spreadsheet programs.

When your submission is received in our email box you will receive an automatic acknowledgment to show that it has arrived. If the submission is incomplete we will contact you and ask you to resend the missing information.

Once the submission is complete we will register it on our manuscript tracking system and you will receive a standard acknowledgment in the post.

Letters to the editor should continue to be sent direct to www.bmj.com as rapid responses or to letters@bmj.com
papers@bmj.com