Systematic Reviews in Health Care

A Practical Guide

Paul Glasziou
School of Population Health,
University of Queensland, Australia

Les Irwig
Department of Public Health and Community Medicine,
University of Sydney, Australia

Chris Bain
Department of Social and Preventive Medicine,
University of Queensland, Australia

Graham Colditz
Channing Laboratory, Harvard School of Public Health,
Boston, MA, USA

CAMBRIDGE
UNIVERSITY PRESS

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS
The Edinburgh Building, Cambridge CB2 2RU, UK
40 West 20th Street, New York NY 10011–4211, USA
10 Stamford Road, Oakleigh, VIC 3166, Australia
Ruiz de Alarcón 13, 28014 Madrid, Spain
Dock House, The Waterfront, Cape Town 8001, South Africa

http://www.cambridge.org

First published 2001

Printed in the United Kingdom at the University Press, Cambridge

Typeface Minion 11/14.5pt *System* Poltype® [VN]

A catalogue record for this book is available from the British Library

Library of Congress Cataloguing in Publication data

Systematic reviews in healthcare; a practical guide / Paul Glasziou . . . [et al.].
p. cm.
Includes bibliographical references and index.
ISBN 0 521 79962 7
1. Systematic reviews (medical research). 2. Evidence-based medicine.
3. Meta-analysis. I. Glasziou, Paul, 1954–
R853.S94 S945 2001
610'.7'2–dc21 00-065170

ISBN 0 521 79962 7 paperback

Contents

Appendixes

1

The question

1.1 What types of questions can be asked?

Clinical problems and health policies may involve many different questions which need to be informed by the best available evidence. It is useful to have a classification of the different types of health care questions that we may ask:

- Phenomena: 'What phenomena have been observed in a particular clinical problem, e.g. what problems do patients complain of after a particular procedure?'
- Frequency or rate of a condition or disease: 'How common is a particular condition or disease in a specified group?'
- Diagnostic accuracy: 'How accurate is a sign, symptom or diagnostic test in predicting the true diagnostic category of a patient?'
- Aetiology and risk factors: 'Are there known factors that increase the risk of the disease?'
- Prediction and prognosis: 'Can the risk for a patient be predicted?'
- Interventions: 'What are the effects of an intervention?'

Answering each type of question requires different study designs, and consequently different methods of systematic review. A thorough understanding of the appropriate study types for each question is therefore vital and will greatly assist the processes of finding, appraising and synthesizing studies from the literature. A summary of the appropriate study types for each question and of the issues that are important in the appraisal of the studies is also given in Table 1.1. General information on how to find and review studies is given in the remainder of Part 1 with further details for each question type in Part 2.

Table 1.1. *Types of clinical and public health questions, ideal study types and major appraisal issues*

Question	Ideal study types	Major appraisal issues
1. Intervention	Randomized controlled trial	Randomization Follow-up complete Blinding of patients and clinicians
2. Frequency/rate (burden of illness)	Cross-sectional study or consecutive sample	Sample frame Case ascertainment Adequate response/follow-up achieved
3. Aetiology and risk	Cohort study	Groups only differ in exposure Outcomes measurement Reasonable evidence for causation
4. Prediction and prognosis	Cohort study	Inception cohort Sufficient follow-up
5. Diagnostic accuracy	Random or consecutive sample	Independent, blind comparison with 'gold standard' Appropriate selection of patients
6. Phenomena	Qualitative research	Appropriate subject selection and methods of observation

1.1.1 Interventions

An intervention will generally be a therapeutic procedure such as treatment with a pharmaceutical agent, surgery, a dietary supplement, a dietary change or psychotherapy. Some other interventions are less obvious, such as early detection (screening), patient educational materials or legislation. The key characteristic is that a person or his or her environment is manipulated in order to benefit that person.

To study the effects of interventions, it is necessary to compare a group of patients who have received the intervention (study group) with a comparable group who have not received the intervention (control group). A randomized controlled trial (RCT), which is a trial in which subjects are randomly allocated to the study or control groups, is usually the ideal design. A hierarchy of designs for the study of the effects of interventions is illustrated in Table 1.2.

1.1.2 Frequency or rate

How common is a particular feature or disease in a specified group in the population? This is measured as the frequency (proportion or prevalence) or rate (incidence) of the feature or disease. For example, the prevalence of osteoarthritis with ageing, or the rate of new cases of human immunodeficiency virus (HIV).

The appropriate study design in this case is a cross-sectional survey with a standardized measurement in a representative (e.g. random) sample of people; for a rate, the sample would need to be followed over time. If, instead of a single frequency, we become interested in the causes of variation of that frequency, then this becomes a question of risk factors or prediction (see below).

1.1.3 Diagnostic accuracy

How accurate is a particular diagnostic screening test? If there is good randomized trial evidence that an intervention for a particular condition works then it may be necessary to assess how accurately the condition can be diagnosed from a sign, symptom or diagnostic test. To do this, a comparison is needed between the test of interest and a 'gold standard' or reference standard. The most commonly used measures of accuracy are the sensitivity and specificity of the test.

If we move from an interest in accuracy to an interest in the effects on patient outcomes, then the question becomes one of intervention (that is, the effects on patients of using or not using the test, as is the case for population screening). However, we are generally content to use diagnostic accuracy as a surrogate to predict the benefits to patients.

Table 1.2. *Types of studies used for assessing clinical and public health interventions (question 1 in Table 1.1)*

Study design	Protocol
Systematic review	Systematic location, appraisal and synthesis of evidence from scientific studies (usually randomized controlled trials)
Experimental studies	
Randomized controlled trial	Subjects are randomly allocated to groups either for the intervention/treatment being studied or control/placebo (using a random mechanism, such as coin toss, random number table, or computer-generated random numbers) and the outcomes are compared
Pseudorandomized controlled trial	Subjects are allocated to groups for intervention/treatment or control/placebo using a nonrandom method (such as alternate allocation, allocation by days of the week or odd–even study numbers) and the outcomes are compared
Comparative (nonrandomized and observational) studies	
Concurrent control	Outcomes are compared for a group receiving the treatment/intervention being studied, concurrently with control subjects receiving the comparison treatment/intervention (e.g. usual or no care)
Historical control	Outcomes for a prospectively collected group of subjects exposed to the new treatment/intervention are compared with either a previously published series or previously treated subjects at the same institutions
Cohort	Outcomes are compared for groups of subjects who have been exposed, or not exposed, to the treatment/intervention or other factor being studied
Case-control	Subjects with the outcome or disease and an appropriate group of controls without the outcome or disease are selected and information is obtained about the previous exposure to the treatment/intervention or other factor being studied
Interrupted time series	Trends in the outcome or disease are compared over multiple time points before and after the introduction of the treatment/intervention or other factor being studied

Table 1.2. (*cont.*)

Study design	Protocol
Other observational studies	
Case series	A single group of subjects are exposed to the treatment/intervention
Post-test	Only outcomes after the intervention are recorded in the case series, so no comparisons can be made
Pretest/post-test	Outcomes are measured in subjects before and after exposure to the treatment/intervention for comparison (also called a 'before-and-after' study)

1.1.4 Risk factor or aetiology

Is a particular factor, such as patient characteristic, laboratory measurement, family history, etc., associated with the occurrence of disease or adverse outcomes? To answer this question a clear association between the factor and the disease must first be established. The most appropriate study type is a long-term follow-up of a representative inception cohort.

If a clear association is shown, the next stage is to determine whether that association is causal. That is, whether the factor under consideration causes the disease or outcome of interest or is merely associated with it for other reasons. This involves issues beyond the degree of association, such as the dose–response relationship and biological plausibility.

1.1.5 Prediction and prognosis

Based on one or several risk factors, what is the level of risk for a particular outcome to the person? Unlike the question of aetiology, causation is not so crucial. Strongly predictive risk markers are also useful. The most appropriate study type is a long-term follow-up of a representative inception cohort.

1.1.6 Phenomena

This question seeks to know the phenomena, subjective and objective, associated with a particular clinical situation. This represents the beginnings of studying a situation by simple observation or questioning. A common research method in health care is qualitative research, that is, the observation and questioning of patients about their experience. We will not cover the systematic reviewing of such questions in this book.

1.2 What is the relevant question?

A well-formulated question generally has four parts:

- the population (or patient group);
- the intervention (e.g. the treatment, test or exposure);
- the comparison intervention (optional, and defaults to no treatment, no test or no exposure if no comparison given); and
- the outcomes.

This question structure is known by the acronym PICO.

Since we will often be interested in all outcomes, the first two parts of the question may be sufficient (see Section 2.2).

1.3 How focused should the question be?

The question should be sufficiently broad to allow examination of variation in the study factor (e.g. intensity or duration) and across populations. For example:

What is the mortality reduction in colorectal cancer from yearly faecal occult blood screening in 40–50-year-old females?

is too narrow as an initial question.

However:

What is the effect of cancer screening on the general population?

is clearly too broad and should be broken down into cancer-specific screening questions.

A better question may be:

What is the mortality reduction in colorectal cancer from faecal occult blood screening in adults?

as this allows the effects of screening interval, age group and gender to be studied.