Epidemiology for the Uninitiated

Fifth edition

Epidemiology for the Uninitiated

Fifth edition

D Coggon

Professor of Occupational and Environmental Medicine, Medical Research Council Environmental Epidemiology Unit, University of Southampton, Southampton General Hospital, Southampton

Geoffrey Rose

late Emeritus Professor of Epidemiology, Department of Epidemiology, London School of Hygiene and Tropical Medicine, London

DJP Barker

Professor of Clinical Epidemiology and Director, Medical Research Council Environmental Epidemiology Unit, University of Southampton, Southampton General Hospital, Southampton

BMJ Books

First published 1979
Third impression 1983
Second edition 1986
Second impression 1987
Third impression 1989
Fourth impression 1990
Fifth impression 1992
Third edition 1993
Fourth edition 1997
Second impression 1999
Third impression 2000
Fourth impression 2002
Fifth edition 2003

British Library Cataloguing in Publication Data

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

ISBN 0 7279 1604 1

Typeset by SIVA Math Setters, Chennai, India
Printed and bound in Spain by GraphyCems, Navarra

Contents

1: What is epidemiology?

Epidemiology is the study of how often diseases occur in different groups of people and why. Epidemiological information is used to plan and evaluate strategies to prevent illness and as a guide to the management of patients in whom disease has already developed.

Like the clinical findings and pathology, the epidemiology of a disease is an integral part of its basic description. The subject has its special techniques of data collection and interpretation, and its necessary jargon for technical terms. This short book aims to provide an ABC of the epidemiological approach, its terminology, and its methods. Our only assumption will be that readers already believe that epidemiological questions are worth answering. This introduction will indicate some of the distinctive characteristics of the epidemiological approach.

All findings must relate to a defined population

A key feature of epidemiology is the measurement of disease outcomes in relation to a *population at risk*. The population at risk is the group of people, healthy or sick, who would be counted as cases if they had the disease being studied. For example, if a general practitioner were measuring how often patients consult him about deafness, the population at risk would comprise those people on his list (and perhaps also of his partners) who might see him about a hearing problem if they had one. Patients who, though still on the list, had moved to another area would not consult that doctor. They would therefore not belong to the population at risk.

The importance of considering the population at risk is illustrated by two examples. In a study of accidents to patients in hospital it was noted that the largest number occurred among the elderly, and from this the authors concluded that "patients aged 60 and over are more prone to accidents." Another study, based on a survey of hang gliding accidents, recommended that flying should be banned between 11 am

and 3 pm, because this was the time when 73% of the accidents occurred. Each of these studies based conclusions on the same logical error, namely, the *floating numerator*: the number of cases was not related to the appropriate "at risk" population. Had this been done, the conclusions might have been different. Differing *numbers* of accidents to patients and to hang gliders must reflect, at least in part, differing numbers at risk. Epidemiological conclusions (on risk) cannot be drawn from purely clinical data (on the number of sick people seen).

Implicit in any epidemiological investigation is the notion of a *target population* about which conclusions are to be drawn. Occasionally measurements can be made on the full target population. In a study to evaluate the effectiveness of dust control measures in British coal mines, information was available on all incident (new) cases of coal workers' pneumoconiosis throughout the country.

More often observations can only be made on a *study sample*, which is selected in some way from the target population. For example, a gastroenterologist wishing to draw general inferences about long term prognosis in patients with Crohn's disease might extrapolate from the experience of cases encountered in his or her own clinical practice. The confidence that can be placed in conclusions drawn from samples depends in part on sample size. Small samples can be unrepresentative just by chance, and the scope for chance errors can be quantified statistically. More problematic are the errors that arise from the method by which the sample is chosen. A gastroenterologist who has a special interest in Crohn's disease may be referred patients whose cases are unusual or difficult, the clinical course and complications of which are atypical of the disease more generally. Such systematic errors cannot usually be measured, and assessment therefore becomes a matter for subjective judgement.

Systematic sampling errors can be avoided by use of a random selection process in which each member of the target population has a known (non-zero) probability of being included in the study sample. However, this requires an enumeration or *census* of all members of the target population, which may not be feasible.

Often the selection of a study sample is partially random. Within the target population an accessible subset, the *study population*, is defined. The study sample is then chosen at