# EDITORIAL COMMENTARY

# Capture-recapture methods in surveys of diseases of the nervous system

This commentary relates to the paper: J H Rees, R D Thompson, N C Smeeton, R A C Hughes. Epidemiological study of Guillain-Barré syndrome in south east England. This volume pp 74–7.

Ecologists have used capture-recapture methods for carrying out censuses of wildlife populations for many years. But only recently have epidemiologists realised that they can be applied to surveys of prevalence and incidence of disease. The underlying idea is fairly simple. The size of a population can be estimated by catching and marking a sample, then releasing that sample and allowing it to mix with the other members of that population before catching a second sample. The proportion of marked to unmarked subjects in the second catch will be the same as the proportion of the number in the first catch to the number in the whole population. The figure makes the principle clearer.

Humans, of course, are effectively marked already—by, for example, some combination of name, sex, date of birth, and NHS or social security number. So an investigator who has employed more than one source of information to identify cases in a survey, can estimate how many cases have been missed from the overlap between sources. The method has proved its worth in circumstances where complete enumeration is especially difficult. Capture-recapture analyses, for example, enabled useful estimates to be made of the prevalence of HIV infection in street working prostitutes in Glasgow and of the number of homeless people in an area of central London.[1 2] They have also been used to establish rates of neurological disorders more accurately. A capture-recapture analysis of data concerning the prevalence of

seizure disorders in children indicated that, despite a system of surveillance designed to be comprehensive, rates of some types of seizure were underestimated by as much as 45%.[3] And in this issue of the *Journal*, Rees and colleagues report that it is likely that about 20% of cases were missed in their survey of Guillain-Barré syndrome in south east England.

A method that allows undetected cases to be counted seems almost too good to be true. It must be admitted that the practical application of capture-recapture analysis to human populations is not as straightforward as the figure suggests. The formula shown holds only if each member of the population has an equal chance of being caught at any particular trapping and that being caught once does not influence the chance of being caught in a second trapping. Sources of information about cases of disease are often not completely independent so that these conditions may be violated. For example, if severely affected cases tend both to have a high mortality and to be referred to specialist centres, lists compiled from registers of deaths and from surveillance of hospital specialists will not be fully independent. Statistical techniques exist that go a long way to deal with this problem.[4-6] One approach is to construct a $2^k$ contingency table (where k is the number of separate sources) containing all the collected information about which cases were identified by each source. One cell in such a table contains no observations; it corresponds to the number of unascertained cases in the population. By fitting a log linear model, the number of observed cases in the other cells can be used to predict the missing value. Interaction terms can be added to the model to investigate dependency between sources. Some degree of judgement is involved in carrying out such an analysis so it is helpful if investigators also report the raw data from which the contingency table was constructed.

Comparison of the frequency of disease between different time periods or between different places depends on precise estimation of incidence or prevalence rates. Anyone who has tried to carry out a survey will know how difficult it is to collect such data. The approach adopted by most conscientious investigators is to use as many sources of information as are available, eliminate duplicates, and assume, usually implicitly, that all cases have been identified. Capture-recapture analysis shows that this assumption is rarely justified but at the same time releases investigators from the necessity of making it. It allows them actually to measure the degree of undercounting. Laporte, who has done more than most to bring capture-recapture methods to the attention of the medical community, has argued that rates should be reported only after the data has been evaluated and adjusted for underascertainment.[7] It is hard to disagree since the application of capture-recapture methods demands little in the way of extra resources or effort from investigators. They need only record the sources from which individual cases were identified and recruit a statistician who can guide them through the process of log linear modelling. The likely benefit is that wider



$$S_1/N = S_{1,2}/S_2$$
$$N = S_1 \times S_2/S_{1,2}$$
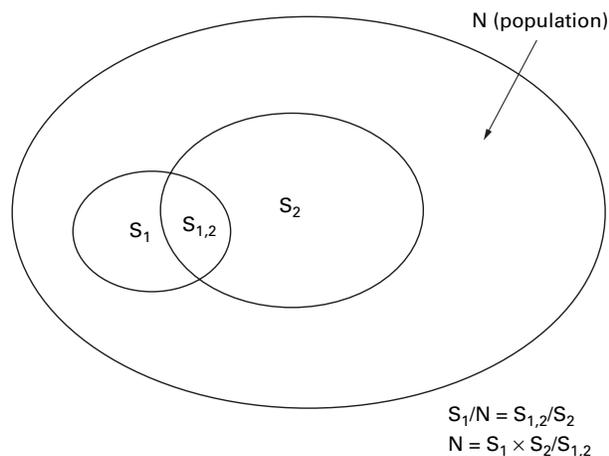
*The principle of the capture-recapture method. $S_1$ represents the sample caught and marked on the first trapping. $S_2$ is the sample caught on the second trapping. The overlap between these samples— $S_{1,2}$—represents those caught in both trappings. The ratio of $S_1$ to the total number in the population, N, is the same as the ratio of $S_{1,2}$ to $S_2$. To avoid estimating the size of the total population as infinite when no individuals are caught in both trappings (where $S_{1,2}$=0), the formula $N=S_1 x S_2/S_{1,2}$ can be approximated as $N=((S_1+1)\times(S_2+1)/(S_{1,2}+1))-1$.*

use of capture-recapture methods will begin to resolve some of the interminable and usually fruitless arguments about whether observed differences in rates of disease are real or an artefact caused by variation in the completeness of case ascertainment.

CHRISTOPHER N MARTYN
*MRC Environmental Epidemiology Unit, Southampton General Hospital, Southampton SO16 6YD, UK*

1 McKeganey N, Barnard M, Leyland A, *et al.* Female streetworking prostitution and HIV infection in Glasgow. *BMJ* 1992;**305**:801–4.

2 Fisher N, Turner SW, Pugh R, *et al.* Estimating numbers of homeless and homeless mentally ill people in north east Westminster by using capture-recapture analysis. *BMJ* 1994;**308**:27–30.
3 Bobo JK, Thapa PB, Anderson JR, *et al.* Acute encephalopathy and seizure rates in children under age two years in Oregon and Washington state. *Am J Epidemiol* 1994;**140**:27–38.
4 Cormack RM. Log-linear models for capture-recapture. *Biometrics* 1989;**45**: 395–413.
5 McCarty DJ, Tull ES, Moy CS, *et al.* Ascertainment corrected rates: applications of capture-recapture methods. *Int J Epidemiol* 1993;**22**:559–65.
6 Hook EB, Regal RR. Capture-recapture methods in epidemiology: methods and limitations. *Epidemiol Rev* 1995;**17**:243–64.
7 LaPorte RE. Assessing the human condition: capture-recapture techniques. *BMJ* 1994;**308**:5–6.

# Unreported trial registration form

Register any controlled trial which has not been published in full, including trials that have only been published as an abstract. Registration can be undertaken by anyone able to provide the registration information, even if they are unable to provide the actual trial data. Please complete one form for each trial being registered.

*Contact details*

Surname:Forename(s):

Postal address:

Phone (with regional codes)

Fax (with regional codes)

email:

Trials details

Approximate number of participants in the trials: _____

Type of participants (for example, people with head injury, women at risk of breast cancer):

_____ _____

Type of intervention (for example, steroids versus placebo, annual mammography versus standard practice):

_____ _____

versus

_____ _____

*Please post or fax completed registration forms to:*

The Editor
*Journal of Neurology, Neurosurgery, and Psychiatry*
Division of Neuroscience and Psychological Medicine
Room 10E15
Imperial College School of Medicine
Charing Cross Hospital
Fulham Palace Road
London W6 8RF

Alternatively, the above information can be sent by email to: meta@ucl.ac.uk

Fax: 0181 846 7730