

Estimating numbers of homeless and homeless mentally ill people in north east Westminster by using capture-recapture analysis

Nigel Fisher, Stuart W Turner, Robert Pugh, Colin Taylor

Abstract

Objectives—To use routinely collected data to provide a reliable estimate of the size and psychiatric morbidity of the homeless population of a given geographical area by using capture-recapture analysis.

Design—A multiple sample, log-linear capture-recapture method was applied to a defined area of central London during 6 months. The method calculates the total homeless population from the sum of the population actually observed and an estimate of the unobserved population. Data were collected from local agencies used by homeless people.

Subjects—Homeless people in north east Westminster residing in bed and breakfast accommodation and hostels or sleeping rough who had contacted statutory or voluntary agencies in the area.

Results—2150 contacts by 1640 homeless people were recorded. The estimated unobserved population was 3293, giving a total homeless population for the period of around 5000 (SD 1250). Mental health problems were significantly less prominent in the unobserved compared with the observed population (23% (754) v 40% (627), $P < 0.0001$). For both groups the prevalence varied greatly with age and sex.

Conclusions—Capture-recapture techniques can overcome problems of ascertainment in estimating populations of homeless and homeless mentally ill people. Prevalences of mental illness derived from surveys that do not correct for ascertainment are likely to be falsely inflated while at the same time underestimating the total size of the homeless mentally ill population. Population estimates derived from capture-recapture techniques may usefully provide a good basis for including homeless populations in capitation calculations for allocating funds within health services.

Introduction

Homelessness is associated with considerable ill health^{1,2} and increased use of acute general hospitals.^{3,4} The planning of health services led by needs for homeless people is hampered by a lack of reliable data on the size of homeless populations.⁵ For example, the 1991 census failed to identify any people sleeping rough in Birmingham on census night.⁶ The prominence of mental health problems among the homeless in inner city areas has generated increasing interest. As much as half of the homeless population may have some form of mental disorder, although estimates of the extent and severity of these disorders varies widely.⁷⁻⁹

Working for Patients gives each health authority responsibility to determine the size and morbidity of both its resident and homeless populations.^{10,11} The size of homeless populations is difficult to estimate because of problems in identifying all homeless people in a given period. This is a result of the relative rarity and elusiveness of homeless people compared with resident populations.¹² Capture-recapture techniques are increasingly being used to overcome these problems of ascertainment^{13,14} and have been recommended for use with homeless populations.¹⁵

Capture-recapture methods derive their name from censuses of wildlife populations. In medical practice the technique allows the number of cases in a defined population to be estimated by using two or more sources of cases. These could include records from hospital or general practice or any other points of contact. Taken alone each of these sources may undercount the actual number of cases, as indeed would a simple aggregate of the sources that excluded duplicate cases (cases identified at more than one source). Capture-recapture techniques use information provided by duplicate cases to allow the number of people not identified at any of the sources to be calculated.¹⁴ An estimate of the total number of cases is thus the sum of the calculated unobserved population and the observed population. In its simplest form the capture-recapture technique makes a range of assumptions about the enumerated population being homogeneous and closed to arrival or departure of cases and the sources being independent of each other. The use of multiple sources and log-linear modelling applied to this technique has been recommended when these assumptions are violated.¹³⁻¹⁶

We report a pilot application of this technique which used several sources to estimate the size of the homeless population of a fixed geographical area over a limited time. We proposed that data sources could be derived from existing lists of homeless people, such as booking in sheets in hostels. In addition, we expected that these lists would contain enough data to describe the population in terms of demography and prevalence of mental health problems.

Subjects and methods

The study area was north east Westminster, an area of central London with a large homeless population. The sources providing data for the capture-recapture analysis were derived from services used by homeless people within this area. Statutory and voluntary agencies that agreed to participate included hospitals, local authority social services, a primary health care centre for homeless people, hostels, an outreach service for the homeless mentally ill, the probation service, and the registrar of deaths for Westminster.

Homelessness was defined to include those placed in emergency bed and breakfast or rented accommodation in the private sector; residents of night shelters, traditional hostels, and short stay hostels; those sleeping rough or on the streets; those in precarious accommodation (for example, squats); and those in non-residential institutions with no other permanent accommodation (for example, hospitals). When ambiguity about homelessness existed recorded addresses were checked against lists of addresses of hostels and bed and breakfast accommodation used to place homeless people within the study area in a manner similar to that used by Victor *et al.*¹⁷

All data were obtained by examining records held at each source. Name, date of birth, and sex were recorded to identify duplicate cases. In addition, if the information was available, type of homelessness and presence of any mental health problems (including those related to use of drugs and alcohol) were

Pathfinder Community and Specialist Mental Health Services, Springfield Hospital, London SW17 7DJ
Nigel Fisher, consultant psychiatrist

Camden and Islington Community Health Services NHS Trust, Hampstead Road, London NW12 2LT
Stuart W Turner, medical director
Robert Pugh, clinical director, mental health services

National Addiction Centre, Institute of Psychiatry, London SE5 8AF
Colin Taylor, senior lecturer in statistics

Correspondence to: Dr Turner.

BMJ 1994;308:27-30

recorded. No subjects were interviewed. Data were collected at the end of two periods of three months: July to September 1991 and October to December 1991. The number of people presenting to each source, rather than the number of contacts, was recorded. Strict confidentiality was preserved. The data collected were not used for purposes other than this study.

Duplicate cases were identified by an algorithm based on name and date of birth which allowed for various name spellings. We then examined these matches and judged whether they were obvious alternative spellings of the same name or were likely to be different names.

The heterogeneity of the homeless population was acknowledged by stratifying the sample by three factors: sex, presence of mental health problems, and age (under 30 or 30 and over). This gave a total of eight population subgroups (see table II). The total population of homeless people was thus the sum of these subgroups. Violation of assumptions relating to dependencies between sources and population subgroups was managed by log-linear modelling by using the GLIM statistical package¹⁸ (see appendix). Three possible types of dependencies were incorporated into the log-linear model estimation procedure (see

appendix). The first was dependencies between sub-population groups—for example, the association of age with an increased likelihood of mental health problems. The second was dependencies between subpopulations and sources—for example, the presence of mental problems increasing the probability of presentation at a particular source. The last was dependencies between sources—that is, presentation at one source increasing the likelihood of presentation at additional sources. The model reported was the one which produced the most acceptable fit to the observed counts.

Results

THE OBSERVED SAMPLE

During the two sampling periods a total of 2150 contacts by 1640 homeless people were recorded, with 393 contacts at more than one source. No deaths of homeless people were identified from death certificates; this source was thus excluded from further analyses. There were no differences between the two sampling periods (representing summer and winter) in numbers of contacts to the remaining agencies (table I). As a result the two periods were aggregated in all further analyses.

Of the 1640 people, 1337 (82%) were men and 303 (18%) women (table II). Of the men, 896 (70%) were aged 30 and over compared with only 104 (41%) women ($\chi^2=102.3$, $df=1$, $P<0.0001$). A total of 627 (40%) people were identified as having mental health problems. For those aged 30 and over the proportions of men and women with mental health problems were similar (399 (44%) men, and 49 (47%) women; $\chi^2=0.2$, $df=1$, $P>0.5$); in the under 30s, however, there was a significant difference (138 (34%) men compared with 41 (23%) women; $\chi^2=7.1$, $df=1$, $P<0.01$). The pattern of use of services was uneven. A person was more likely to contact more than one agency if he or she was 30 and over ($\chi^2=5.0$, $df=1$, $P<0.025$) or had a mental health problem ($\chi^2=70.7$, $df=1$, $P<0.0001$) (table III).

CAPTURE-RECAPTURE ANALYSIS

We found strong interdependencies between population subgroups and sources. The older subgroup was more likely to go to hostels while the younger was more likely to present to the probation service; women were more likely to go to social services, hostels, and hospitals. Of necessity the presence of mental health problems was associated with contact with the homeless mental health team. The above heterogeneity accounted for almost all dependencies between sources apart from, firstly, social services and the homeless mental health team and, secondly, hostels and the homeless mental health team. The nature of these dependencies reflects the known use of services and gives the model validity.

UNOBSERVED POPULATION

The unobserved homeless population was calculated as 3293 (table III), giving a total estimated population of around 5000 (SD 1250). This indicates that a simple survey would have ascertained barely one third of the total homeless population. The demographic characteristics and the relative distribution of mental health problems were similar for the estimated unobserved population and the observed sample (tables II and IV). There were, however, significantly fewer mental health problems in the unobserved compared with the observed population (23% (754) v 40% (627), $\chi^2=145.8$, $df=1$, $P<0.0001$).

Discussion

The results of the capture-recapture analysis indicate, firstly, that the homeless population of

TABLE I—Numbers of homeless people observed by source and time period

Time period	No observed for each sample source (as % for time period)						
	Hospital	Social services	Primary care	Homeless mental health team	Probation service	Registrar of deaths	Hostels
July-September	110 (10%)	105 (10%)	582 (55%)	26 (2%)	104 (10%)	0	134 (13%)
October-December	97 (9%)	83 (8%)	545 (50%)	22 (2%)	129 (12%)	0	213 (20%)

TABLE II—Observed population of homeless and estimation by capture-recapture techniques of unobserved population of homeless by population subgroup

Population subgroup	No observed	Estimated No unobserved (95% confidence interval)	Total No (observed + unobserved)
Women with no mental health problem:			
Aged < 30 years	138	255 (195 to 334)	393
Aged ≥ 30 years	55	79 (56 to 112)	134
Women with mental health problem:			
Aged < 30 years	41	33 (23 to 48)	74
Aged > 30 years	49	38 (26 to 55)	87
Women of unknown age:	20	NA	20
Men with no mental health problem:			
Aged < 30 years	269	820 (656 to 1025)	1089
Aged ≥ 30 years	497	1384 (1134 to 1688)	1881
Men with mental health problem:			
Age < 30 years	138	161 (123 to 211)	299
Aged > 30 years	399	522 (421 to 647)	921
Men of unknown age	34	NA	34
Total	1640	3293	4933

TABLE III—Prevalence of contact with agencies in observed sample of homeless people

	Sex		Age (years)		Mental health problems	
	Men	Women	< 30	≥ 30	Yes	No
No in sample	1337	303	586	1000	652	988
No (%) making more than one contact	167 (12%)	31 (10%)	59 (10%)	139 (14%)*	133 (20%)	65 (7%)+

* $\chi^2=5.0$, $df=1$, $P<0.025$ (for difference between one or more than one contact by age group).

+ $\chi^2=70.7$, $df=1$, $P<0.0001$ (for difference between one or more than one contact by mental health status).

TABLE IV—Distribution of mental health problems

Population subgroups by age and sex	Proportion (%) with mental health problems		χ^2 For difference in mental health problem between observed and unobserved populations
	Observed population	Unobserved population	
Women < 30 years	41/170 (23)	33/288 (11)	10.9, $df=1$, $P<0.009$
Women ≥ 30 years	49/104 (47)	38/117 (32)	4.9, $df=1$, $P<0.025$
Men < 30 years	138/407 (47)	161/981 (16)	52.2, $df=1$, $P<0.0001$
Men ≥ 30 years	399/896 (44)	522/1906 (27)	81.2, $df=1$, $P<0.0001$
Total	627/1586 (40)	754/3293 (23)	145.8, $df=1$, $P<0.0001$

north east Westminster is substantially larger than previously estimated, and, secondly, that there are important differences between the observed and unobserved populations, particularly in the prevalence of mental health problems.

The completion of this pilot project suggests that log-linear capture-recapture analysis can be applied to existing datasets to estimate the size and nature of the homeless population for a given geographical area. Questions of validity and reliability cannot be conclusively answered as no gold standard for the size of this population exists and there are no similar studies for comparison. With this in mind the results of the study should be taken in the context of several methodological limitations.

LIMITATIONS

The population was not closed. People could have moved into or out of the area and become or stopped being homeless. Potentially this inward or outward flow could have been determined by fitting more complex log-linear models¹⁶; this, however, was beyond the remit of this pilot project. The estimate is therefore dependent on the time period of the study, with the expectation that the total would increase over time as new cases arrive. Thus the figure of 5000 should be taken to indicate the number of people who were homeless in north east Westminster for at least part of the six months studied. This population estimate remains valuable for planning health services as all the members of that population would have a valid claim to use the local health care services.

People whose age was unknown did not present at more than one source. Consequently it was not possible to complete any capture-recapture analysis on this group. Clearly it is improbable that no people from this group were unobserved. The total unobserved population was about twice the size of the observed population. If a similar relation applied to the group of unknown age then about 100 people could be added to the unobserved total.

Data on mental health problems relied on the use of information recorded at each source and cross referencing between sources. For some sources we assumed that the quality of this information was good (for example, hospitals); for other sources the information was likely to be less robust. Problems associated with drugs and alcohol were recorded separately from other mental health problems, but the numbers were too small to allow for a discrete capture-recapture analysis.

Identification of all homeless people presenting to the study sources was likely to be incomplete. This was most obvious at the registry of deaths, where no deaths of homeless people were identified from death certificates. This was an improbable finding² and may have resulted from a practice of using the last known address rather than the current situation in completion of death certificates. Some duplicate cases may have been missed if people used aliases when presenting at different sources.

COMPARISON WITH OTHER STUDIES

A recent report places this study in context. Black *et al* estimated that within what was Bloomsbury Health Authority, which encompasses north east Westminster, there were 3467 homeless people.³ The area covered by this study was about half the size of Bloomsbury Health Authority, yet 1640 homeless people were identified and the capture-recapture analysis estimated that there were a further 3293 homeless people that were unobserved during the study period.

Clearly from both the observed and estimated unobserved populations the homeless population is

Social implications

- Homelessness is associated with ill health
- Current estimates of the size of homeless populations are unreliable because of problems in identifying all homeless people for a given time period
- Capture-recapture methods overcome problems of ascertainment by calculating the size of the unobserved population
- This study estimated the population of unobserved homeless to be twice the size of the observed population
- Mental health problems are significantly less common in the unobserved or hidden homeless compared with those who are easily surveyed

heterogeneous both demographically and in distribution of mental health problems. The prevalence of mental health problems in the observed population was similar to that reported previously.⁸ Strikingly, for all age and sex categories the prevalence in the estimated unobserved population was significantly lower (table III) and in total was barely half that of the observed population. This indicates that homeless people with mental health problems are disproportionately prominent. Consequently prevalences derived from surveys that do not correct for ascertainment are likely to be falsely inflated while at the same time underestimating the total size of the homeless mentally ill population.

CONCLUSION

Capture-recapture techniques can overcome problems of ascertainment in the estimation of homeless and homeless mentally ill populations and is an advance on other current methods of estimation. Population estimates derived from capture-recapture techniques may usefully provide a good basis for including homeless populations in capitation calculations for allocating funding in health services.

Nigel Fisher was supported as a research worker in the department of psychiatry, Middlesex Hospital, by a grant from the special trustees of the Middlesex Hospital. The authors are grateful to all the agencies who cooperated during the course of this study.

- 1 Connelly J, Kelleher C, Morton S, St George D, Roderick P. *Housing or homelessness—a public health perspective*. London: Faculty of Public Health Medicine of the Royal Colleges of Physicians of the United Kingdom, 1991.
- 2 Keyes S, Kennedy M. *Sick to death of homelessness: an investigation into the link between homelessness, health and mortality*. London: Crisis at Christmas, 1992.
- 3 Black M, Scheuer M, Victor C, Benzeval M, Gill M, Judge K. Utilisation by homeless people of acute hospital services in London. *BMJ* 1991;303:958-61.
- 4 Fisher N, Turner S, Pugh R. Homeless and mentally ill. *Lancet* 1990;335:914-5.
- 5 Oldman J. *Who says there's no housing problem? Facts and figures on housing and homelessness*. London: Shelter, 1990.
- 6 Office of Population Censuses and Surveys. *1991 census—preliminary report for England and Wales. Supplementary monitor on people sleeping rough*. London: HMSO, 1991.
- 7 Morrissey J, Levine I. Researchers discuss latest findings, examine needs of homeless mentally ill persons. *Hospital Community Psychiatry* 1987;38:811-2.
- 8 Scott J. Homelessness and mental illness. *Br J Psychiatry* 1993;162:314-25.
- 9 Tessler R, Dennis D. *A synthesis of NIMH funded research concerning persons who are mentally ill*. Washington DC: National Institute of Mental Health, 1989.
- 10 Department of Health. *Working for patients*. London: HMSO, 1989.
- 11 Fisher N, Turner S, Pugh R. Working for patients: will it work in practice? *Psychiatric Bulletin* 1991;15:73-5.
- 12 Susser E, Conover S, Struening EL. Problems of epidemiologic method in assessing the type and extent of mental illness among homeless adults. *Hospital Community Psychiatry* 1989;40:261-5.
- 13 Hook E, Regal R. The value of capture-recapture methods even for apparent exhaustive surveys. *Am J Epidemiol* 1992;135:1060-7.
- 14 McCarty D, Tull E, Moy C, Kwok C, Laporte R. Ascertainment corrected rates: applications of capture-recapture methods. *Int J Epidemiol* 1993;22:559-65.

- 15 Sudman S, Sirken M, Cowan C. Sampling rare and elusive populations. *Science* 1988;240:991-5.
- 16 Wittes J, Colton T, Sidel V. Capture-recapture methods for assessing the completeness of case ascertainment when using multiple information sources. *Journal of Chronic Disease* 1974;27:25-36.
- 17 Victor CR, Connelly J, Roderick P, Cohen C. Use of services by homeless people in an inner London health district. *BMJ* 1989;299:725-7.
- 18 Cormack R. Log-linear models for capture-recapture. *Biometrics* 1989;45:395-413.
- 19 Payne C, ed. *The GLIM system, release 3.77*. Oxford: Numerical Algorithms Group, 1986.

(Accepted 1 October 1993)

Appendix

Details of the statistical analysis

CAPTURE-RECAPTURE METHODS

An estimate of the total population (N_T) in a simple two source analysis uses information on the relative frequency of duplicate cases (N) and cases in only the first sample (N_1) or the second (N_2) in relation to the number of unobserved cases where $N_T \times (N_1 + N_2) / N$. The assumption that the captures are independently made is crucial in allowing this relation to be determined. In multiple sources recapture analysis the increased information on different multiple captures allows this assumption to be tested and allows models which incorporate heterogeneity among subgroups in the population. These models are analysed with standard log-linear modelling programs,¹⁸ in this case GLIM3.77.¹⁹

MODELLING PROCEDURE

The model allowed for all interactions between subgroups. First order interactions were included for dependencies between subgroups and individual sources and for dependencies between sources. The related higher order interactions were not included. Owing to the special position of the homeless mental health team in relation to mental illness the basic model was extended to include all interdependencies between this source and all combinations of subgroups.

ESTIMATION PROCEDURE

Interactions in the fitted model (apart from those with the homeless mental health team) were removed as being unnecessary if the effect was less than four times its standard error. This was a harsh criterion deliberately used to avoid "overfitting" in such a large model, which would have been likely to inflate the overall population estimate. The retained interactions between sources and subgroups (see text) were strong ($\chi^2=243.6$, $df=8$, $P<0.0001$); the between source dependencies, which were present even after allowing for this subgroup heterogeneity of capture, were by contrast weak and although significant ($\chi^2=25.9$, $df=2$, $P<0.001$) were excluded from the final, conservative model, again to avoid any overfitting. Had they been retained in the estimation procedure the overall population estimate would have been 5% higher but of similar structure. The final model scaled deviance was acceptably close to its degrees of freedom ($\chi^2=338.7$, $df=481$), suggesting a good but not overly close fit and a total population estimate of 3293.

Lifeline Wales: experience with a computerised kidney donor registry

J R Salaman, P J A Griffin, W Ross, J Haines

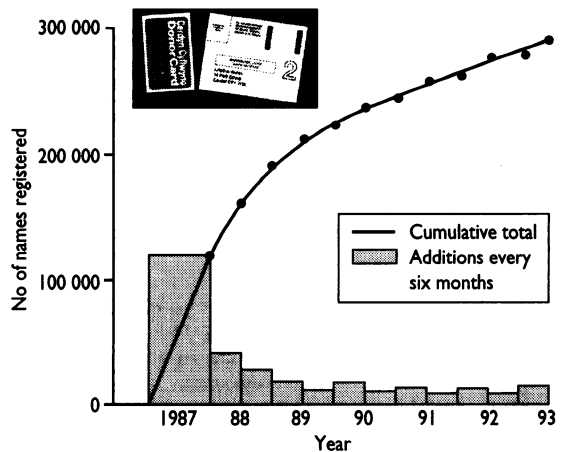
Department of Surgery and Transplantation, Royal Infirmary, Cardiff CF2 1SZ
 J R Salaman, professor of transplant surgery
 P J A Griffin, associate specialist in transplantation
 W Ross, research fellow
 J Haines, transplant coordinator

In the United Kingdom about 60 people per million of the population develop chronic renal failure each year and are accepted on to dialysis programmes, but barely half will receive kidney transplants.¹ The shortage of donors is aggravated by the reluctance of relatives to consent to donation in one third of cases.^{2,3} Relatives seem unable to give consent when the donor's wishes are not known. To address this problem various attempts have been made within the United Kingdom to record the wishes of potential donors. Computer registries can retain such information long term and can operate as an extension to the donor card scheme without a change in the law. Lifeline Wales is the largest such scheme in the United Kingdom, and this paper describes its operation over the past six years.

Methods and results

In 1983 a computerised organ donor registry was set up at Manchester University.⁴ Local intensive care units checked the database using personal computers and modems whenever they had in their care a potential organ donor whose wishes were not known. In 1986 the scheme was extended to Wales with funding from the Welsh Office and the Kidney Research Unit for Wales Foundation. Special bilingual donor cards were sent to 1.2 million Welsh homes in 1986, with regular distributions since. The card had a small tear off portion identical with the standard donor card, and this was retained by the person. The other part was returned to the registry office of Lifeline Wales, where personal details were transferred to magnetic tape and sent to Manchester for entry into the computer.

In 1990 the Manchester computer ceased operating and the Lifeline Wales database was transferred to Cardiff. Because the computer terminals in the 15 intensive care units in Wales were seldom used they



Registrations with Lifeline Wales every six months from 1987 to 1993

were withdrawn from service in 1990. Staff now check the database by consulting the scheme's transplant coordinators (WR and JH), who have their own terminal. The donor card was also modified to include a question on blood donation.

One year after launching Lifeline Wales 100 000 potential donors had been registered. The accrual of names has been more gradual since then, reaching 290 640 by January 1993 (figure). About 17 000 new names continue to be added annually. Before 1986 there were about 35 donors each year. Two years after the launch this had increased to 44 donors (representing 20 donors per million of the population), which was the highest regional procurement rate in the United Kingdom. However, only 38 donors were found during 1991 and also 1992. Fourteen people who had registered with Lifeline Wales subsequently died and became organ donors. In none of these cases was the registry consulted because the person who died had been carrying a donor card and his or her views were already known.

Comment

Whether Lifeline Wales has influenced rates of organ donation in Wales is difficult to say because Wales has traditionally scored well: in 1991, for

Correspondence to: Professor Salaman.

BMJ 1994;308:30-1