

METHODOLOGIC ISSUES

Children are not goldfish—mark/recapture techniques and their application to injury data

Stephen N Jarvis, Philip J Lowe, Alison Avery, Sara Levene, Richard M Cormack

Abstract

Objectives—Mark/recapture (or capture-recapture) is a simple technique commonly applied to estimate the hypothetical total (including undercount) in a register composed of cases from two or more independent and separately incomplete case lists. This paper seeks to illustrate serious drawbacks in the use of the mark/recapture technique when applied to injuries.

Setting and subjects—Northumbrian children under 15 years of age who were seriously injured in motor vehicle accidents (MVAs) over a five year period ascertained from two data sources: police reports and hospital inpatient records.

Methods—Individuals (*n*) appearing in both police (*S*) and hospital (*H*) case lists are identified using various matching criteria. The separate and combined influence of age, sex, and casualty class (cyclist, passengers, pedestrians) on the probability of such matching is estimated using multivariate techniques. The hypothetical total incidence of child MVA victims (*N*) is calculated from $N = (S \times H)/n$.

Main outcomes—Estimates of the incidences of “serious” injuries in MVAs under various conditions of stratification and matching. The overall procedure is tested for conformity with accepted criteria for valid use of mark/recapture.

Results—About one third of the 1009 police and 836 hospital records could be exactly matched. There were significant variations in matching proportions by class of accident (pedestrian *v* passenger *v* cyclist). This selective recapture or “heterogeneity” was not affected by sex, but was independently influenced by the age of the child. Further uncertainty was introduced when matching criteria were slightly relaxed. Estimates of the total population of children with serious injuries vary accordingly from 1729 to 2743. A number of plausible reasons why these two data sources might not be unbiased or mutually independent samples of the total target population are proposed as explanations for this heterogeneity.

Conclusion—This typical example of two sample mark/recapture estimation in an epidemiological setting can be shown to violate virtually all the requirements for valid use of the technique. Very little can be deduced accurately about the scale or characteristics of an unobserved group by the use of mark/recapture applied to two overlapping health event registers.

(*Injury Prevention* 2000;6:46–50)

Keywords: mark/recapture; epidemiology

When researchers want to estimate the rate of occurrence of injuries they often have limited sources of information to work with. Commonly, there will be hospital records, and for traffic accidents, police records. Ideally the two should be identical, but this is never the case. When there is only partial overlap between two sources of injury information investigators may ask if there is not another group of cases unknown to either source. In order to estimate the total number of cases, under such circumstances, the mark/recapture technique is often used. This technique is popular in wildlife biology where repeated samples of a closed population are used to estimate the total. The estimate is based on recognising individuals in any sample who have been also captured previously. The technique has increasingly been applied in epidemiological settings where two or three partially overlapping lists of “cases” are used to estimate a “missing” group^{1–4} and has been advocated in editorials in medical journals.^{5,6} Because the technique has also been exploited in injury epidemiology^{7,8} it is important to explore its limitations.

The necessary conditions for this type of estimation were set out by Seber in 1982.⁹ They might be summarised in the context of two samples taken from a goldfish pond as follows.

- The population is “closed” with *N*, the true number of goldfish, being constant between samples.
- When the samples are undertaken, they are independent equal probability samples of the whole population of goldfish.
- Having been caught once should not alter the probability of a goldfish being caught a second time.

Department of Child Health, University of Newcastle upon Tyne, Donald Court House, 13 Walker Terrace, Gateshead NE8 1EB, UK

S N Jarvis
P J Lowe
A Avery

Child Accident Prevention Trust, London
S Levene

School of Mathematical and Computational Sciences, University of St Andrews
R M Cormack

Correspondence to:
Professor Jarvis

Table 1 Number of children injured in MVAs identified from hospital (HES) and police (Stats19) sources and the extent of overlap between them

	Class 1: cyclists	Class 2: passengers	Class 3: pedestrians	All classed episodes
No in HES data (using E codes)*	114	97	625	836
No in Stats19 data	126	100	783	1009
Matches	21	33	283	357
% of HES records matched	18.4	34.0	45.3	42.7
% of Stats19 records matched	16.7	33.0	36.1	35.4

*Cyclists = E (810–819) .0/.2/.6 + E (827–829) .1, passengers: E (810–819) .1, pedestrians: E (810–819) .7 + E (826–829) .0.

NB: 751 additional child admissions with E826.1 (pedal cycle accident with no motor vehicle involved) and 98 horse riders (E828.2) are excluded. Neither group had matches in Stats19.

d. Such recapture should be recognised every time it occurs.

It is not difficult to see how, even in the apparently free living circumstances of a goldfish pond, such a set of requirements might be difficult to meet. Indeed, estimates of population size arrived at from such applications of the technique have nearly always been found to be gross undercounts of the true value.¹⁰

The most likely explanation is that certain characteristics make some goldfish “easier” to catch, either on both occasions or on the second occasion (for example they are larger, or are more visible because they are marked). This non-random sampling will be revealed by differences in the characteristics of those recaptured compared with those who appear in only one sample (“heterogeneity”).

In an epidemiological example, the conditions have a somewhat different meaning⁹: “a” translates as a requirement that two (or more) lists of possible cases are samples from the same (possibly dynamic) population at risk during the period of data collection. Condition “b” represents a requirement that the probability of observing any individual should be the same for observed and not observed. This can be addressed by questioning whether the methods by which the lists are generated will have a greater probability of ascertaining some members of the target population than others. Condition “c”, meanwhile, requires that presence on one list is not in any way contingent on presence on the other list—that is, that there is not “dependence” between the lists. Failure to meet this criterion, as with b, may be revealed if there are marked differences between the characteristics of those recaptured and those captured only once (that is, heterogeneity). Condition “d” translates as a requirement for correct matching between lists when the same individual is truly present in both.

We have examined these issues, and the likely consequences for estimates of a total population, using two apparently independent lists of children who have been injured in motor vehicle accidents (MVAs).

Method

The target group which it was intended to enumerate was those children under 15 years of age from addresses in Northumbria (Tyne & Wear and Northumberland) who were “seriously” injured in local MVAs between 1 April 1990 and 31 March 1995.

Police “Stats19” data covering all road traffic accidents in Northumbria that had been reported to the police and which involved injuries to children were compared with hospital episode (HES) data covering admissions of children whose home address postcodes lay within the same area.

Valid postcodes were present on 98% of the HES records. Some of these records were diagnosis and cause coded using the *International Classification of Diseases*, 10th revision (ICD-10).¹¹ These were re-coded in ICD-9 and included with the study set. Duplicates (matching dates, address, postcode, age, and sex) were removed. Hospital episodes were then selected that had at least one ICD-9 injury code or one ICD-9 external cause code giving 24 040 episode records with 53 572 diagnoses.

Of these episodes, 23 226 had an ICD-9 injury code and 14 640 had both an ICD-9 injury code and also an ICD-9 cause code (E code). The cause coding ratio of this set is, therefore, 63%.

At this stage, the set was further reduced by selecting only those 836 HES records with ICD-9 cause codes relevant to MVAs (see table 1 footnote for E code categories used).

The Stats19 data provided by the Tyne & Wear Traffic Accident Data Unit were combined with confidential data concerning the home address postcodes of the individual children direct from Northumbria Police.¹² Earlier work with this dataset has shown it to be fully coded for the variables used in this analysis (that is, sex, age, date, postcode, casualty class, and severity). The Stats19 data set with valid Northumbria postcodes for the same accident date range as the admission date range in the HES data contained 4668 records. There were 1009 records with a severity score of 2 (serious) and 3627 with severity score 3 (slight). The records referring to deaths (severity 1) were excluded. “Serious” injury in Stats19 is defined as follows¹³:

Examples of “serious” injury are:

- Fracture
- Internal injury
- Severe cuts and lacerations
- Crushing
- Concussion
- Severe general shock requiring hospital treatment
- Detention in hospital as an inpatient, either immediately or later
- Injuries to casualties who die 30 or more days after the accident from injuries sustained in that accident¹³

For most children injured in MVAs it was considered that such injuries would lead to hospital admission (but see Discussion).

After matching to determine the overlapping numbers (n) of the Stats19 (S) and HES (H) “lists” (see the Results section for matching variables), the estimate of the total population was made using the formula $N = (S \times H)/n$. This was repeated within strata by casualty class (cyclist/passenger/pedestrian), sex and age of victims and also with more or less stringent matching criteria. The effect of heterogeneity by class/sex/age was assessed by testing the

Table 2 Fitted percentages of “recaptures” by age and class (that is, matches between hospital (HES) and police (Stats19) data)

Age (years)	Cyclists	Passengers	Pedestrians
0–4			
HES	11.8	25.5	34.0
Stats19	15.9	32.4	34.1
5–9			
HES	22.2	42.1	52.3
Stats19	20.2	39.0	40.9
10–14			
HES	17.1	34.5	44.2
Stats19	14.7	30.4	32.1

Fitted model has a residual chi square of 24.3 with 26 df.

Table 3 Number (%) of “recaptures” by severity of injury (that is, percentage of hospital cases also appearing in matched police records)

	Cyclists (%)	Passengers (%)	Pedestrians (%)
Severe* injury	7/32 (22)	10/13 (77)	112/190 (59)
Not severe injury	14/82 (17)	23/84 (27)	171/435 (39)

*Fractures of skull, of femur, or of other long bones requiring operative correction.

equality of capture probabilities in either list for these three factors, in a procedure akin to analysis of variance.¹⁴ The computations were carried out in GLIM.¹⁵

Results

MATCHING

After progressive relaxation of each criterion in turn, it was decided that if sex, age (\pm one year), date of event/admission (equal to or plus one day), and postcode coincided, a match was established. The absence of duplicate matches within the data reinforced this view. This gave 357 matching cases (see table 1). As will be seen from the “matches” row total, there were 6% (20) of these matches where the hospital cause code did not match the police class of accident. The effect of further alteration of the matching criteria is illustrated later.

MARK/RECAPTURE

At first sight, table 1 data suggest that, with more than half of relevant HESs unknown to the police and nearly two thirds of police reported child MVAs with serious injury apparently unknown to the hospitals, there might well be a substantial group with equivalent injuries “missed” by both “lists” of cases. We now examine if the conditions are met for this “unknown” group to be estimated using mark/recapture techniques.

HETEROGENEITY

Table 1 also shows the percentage of HES data that were “recaptured” in the Stats19 (police) data, and conversely the percentage of Stats19 data “recaptured” in HES. The two list capture probabilities vary among classes (reduction in

deviance of 35.12 with 4 df, $p < 0.001$). Further stratified analysis shows that these probabilities do not vary between sexes (reduction in deviance of 2.18 with 2 df, $p > 0.3$). Nor is there any interaction between classes and sex (reduction in deviance of 2.29 with 4 df, $p > 0.5$). The probabilities do, however, vary among age groups (0–4, 5–9, 10–14 years) with a reduction in deviance of 21.7 with 4 df, $p < 0.001$, but without interaction between class and age (reduction in deviance of 9.8 with 8 df, $p > 0.2$). The fitted percentages from the model (table 2) show that pedestrians have the highest chance of being recorded, cyclists the least, and that 5–9 year olds have the highest chance of being recorded, the youngest children the least.

There is thus clear evidence of heterogeneity, according to two of the three factors examined.

EVIDENCE FOR FURTHER HETEROGENEITY

In addition to class, sex and age, a number of other potential reasons for biased membership of one or other of the lists could be proposed. For instance, more severe injury might be expected to lead to more reliable ascertainment by both police and hospital. However, the allocation of severity can only be done within the hospital dataset (see table 3).

This reveals that there is significant heterogeneity of matching (recapture) by injury severity (change in deviance 27.3 on 1 df). Similar stratification within the Stats19 data by the four police force areas revealed further significant heterogeneity (not shown).

RELIABILITY OF RECAPTURE

We were concerned that some actual matches might have been missed because of our requirement for a perfect (seven characters) home address postcode match. Removing this criterion (that is, matching only on sex, age within one year, and date of event not more than one day before admission) added 131 (39%) further “recaptures”. In this process, about 10% of matched records were excluded as duplicates (for example, two possible candidates in the HES data for a single Stats19 record). Table 4 shows the distribution of these new data and may be compared with table 1.

EFFECTS ON MARK/RECAPTURE ESTIMATES

Table 5 shows the estimated numbers by class of accident using the different categorisations.

The estimate of the numbers of cases varies considerably according to the data used. Occasionally the estimated total numbers (N) are more than twice the number of cases actually observed (for example, 211 cyclists are observed in table 4, but the estimated total in table 5 is 495).

OTHER REASONS FOR UNMATCHED CASES

As noted in Methods, the Stats19 data also included 3627 children categorised by the police as having “slight” injury. Hospital admission should attract the label of “serious” injury (see definition above) but in fact an additional 125 children were identified who had “slight” injury according to the police but

Table 4 Number of overlapping cases when matching criteria exclude the requirement for identical postcode

	Cyclists	Passengers	Pedestrians	All episodes
No in HES data (using E codes)	114	97	625	836
No in Stats19 data	126	100	783	1009
Matches	29	39	370	488

Table 5 Total N calculated using mark/recapture formula

From data in	Class 1: cyclists	Class 2: passengers	Class 3: pedestrians	Total with classes separated	All classed episodes
Table 1	684	294	1729	2707	2363
Table 2 (years)					
0–4	58	79	409	546	
5–9	195	60	634	889	
10–14	443	158	707	1308	
Total age separated	696	297	1750	2743	
Table 4	495	249	1323	2067	1729

for whom there were matches with the HES data using the primary criteria.

Discussion

We discuss first the extent to which this study provides evidence for or against the use of mark/recapture techniques in the epidemiological setting. We then consider the relationship of these findings to other similar studies. To what extent then does the present study meet the “requirements” for the mark/recapture method?

As far as possible, the study has attempted to define a “closed” population—Seber’s condition “a”. Note here that “closed” does not infer that the child population of Northumbria is unchanged—rather that the reference population for individual cases is the same for both lists. In this study the reference population is children from Northumbrian addresses with “serious” injuries sustained in MVAs over a five year period. Despite the extended time over which our lists are compiled, an individual’s availability for both lists is almost concurrent. The crucial assumption, that an individual in the population from which one list is a sample, is also in the population from which the other list is a sample, is thus satisfied.

Why should any of the target cases not appear in the two lists? (requirement b). It is known that among all age groups, 50% of MVA victims with “serious” injury are not admitted to hospital.¹⁶ Admission is more likely for pedestrians (three quarters of our study group) and we expect that children will be more readily admitted to hospital than adults with equivalent injuries. Other target cases may have incorrect or absent “cause” codes in HES data. However, if the correctly cause coded group are, in effect, a random sample of true MVA hospital admissions, then this is the normal way to conduct mark/recapture population estimates. Alternatively, there are some types of MVAs that the police do not ascertain fully—for example, cyclists in minor collisions.¹⁷ Some may also be incorrectly assigned as “not serious” by the police.^{16–18} This is confirmed in the present study—one quarter (125/479) of the unmatched HES records in table 2 appearing in Stats19 “slight” injuries.

Why should any of the cases have a greater probability of appearing in a second list once they had appeared in one of the lists? (requirement c). The police, in establishing whether an injury is “serious” are recommended to contact the hospital to find out whether the child is admitted or not.¹⁶ Conversely, the patient may contact the police to inform them of the history

of a MVA injury event for insurance purposes. This latter is a plausible explanation for the high rate of matching for passengers with severe injury (table 4). Children injured in MVAs as pedestrians or cyclists rarely enter insurance claims as they are held to be “at fault”. Unfortunately, this direct dependence between the two lists can not be estimated without a third and independent data source.

How could a situation arise where a “marked” case is not recaptured—that is, matching does not occur when the same child appears in both lists? (requirement d). The analysis has illustrated the potential effect of this by excluding the matching variable most likely to have been miscoded. The subsequent presence of duplicate matches suggests that this set of extra “recaptures” will conceal some spurious matches.

In summary, therefore, although the requirement for a “closed” population is met, there are *a priori* reasons to expect that the members of the two lists are not complete, unbiased, or independent samples of this target group. This is confirmed by the finding of widespread heterogeneity, which could imply that the hypothetical unobserved cases (the undercount) are significantly different to those who do appear in the lists. Those who do appear are more likely to be injured as pedestrians, to be aged 5–9, to be more severely injured, and to live in certain police force areas. Estimates of the absolute numbers of those who do not appear (the undercount) are further undermined by the demonstrated consequences of inaccurate matching (recapture). This can substantially alter the size of the observed and unobserved fractions.

How does this relate to previous work in this area? Two early examples of the use of mark/recapture techniques in injury epidemiology are the papers by Roberts and Scragg⁷ and that by Laporte *et al.*⁸ In the first, an estimate was made of the undercount in the Auckland child pedestrian injury study by comparing active surveillance of hospital ward admission books and clinical records, to the codes attributed to public hospital discharges for Auckland resident children in the same period. One hundred and eighty four children were common to the active surveillance group (206) and the discharge coded group (238), giving an estimated undercount of six cases in addition to the 260 known to one or other source. Further analyses revealed that, although active surveillance was almost equally likely to identify non-traffic and traffic victims (0.79 and 0.77 probabilities), the discharge coding was much more likely to identify or “sample” the traffic group (0.66 *v* 0.98). In this New Zealand study, therefore, there was some evidence of a similar phenomenon to that observed in the present study, insofar as capture probability is influenced by the types of accidents. Without a third source of cases, it is impossible to gauge whether this heterogeneity reflects some form of varying dependence between the case sources, or merely unequal sampling probabilities. The authors considered dependence unlikely, but do not discuss why the low “sampling” rate of

non-traffic accidents by discharge coding should imply an overall undercount.

In the second study, Laporte *et al* attempted to estimate the number of physician treated injuries over a four month period among about 1200 Pittsburgh adolescents. The “aggregate” number of injuries (that is, those that appeared in at least one of the four available sources of injury data) was used to define the minimum true count. By this standard, some of the two source capture-recapture estimates were obviously undercounts implying dependence between the sources. Other two source combinations produced apparent overcounts. The authors then introduced data from a third source to assess possible dependency more directly. Three source capture-recapture estimates of the total number of injuries were derived from log linear models, which allowed for such dependency. These models still only explained part of the variance and resulted in wide confidence intervals for the true count. The authors then went on to say that “*when using these (capture-recapture) techniques, investigators should always be aware that they are making assumptions that the unobserved individuals will behave as the observed individuals*”.

Herein lies the essential problem, for it is not the absolute size of the case population that is the only, or even the principal, issue in epidemiology. The key problem is to ensure that risk factor analysis based on observed cases is not confounded by factors that reflect propensity for true cases to be selectively observed in the first place.

Conclusion

It would seem that deviations from the fundamental requirements for mark/recapture are more or less inevitable, given the systematic ascertainment biases that attend any health event registration and many self reporting population surveys. More appropriate settings for the technique, if used at all in humans, might be for calculations of census undercounts (that is, from geographically random sample survey) and in situations where there is an unambiguous case definition independent of service utilisation (for example, malignant neoplasms). At the least, those reporting mark/recapture estimates should seek heterogeneity of matching by stratified analyses on as many independent variables as are available in one or both datasets. Alternatively a logistic model can incorporate quantitative covariates for individuals¹⁹ but this requires the particular model to be true, and its truth confirmed as far as possible, by subsidiary analyses of the kind we recommend. Subsequent adjustments of the mark/recapture estimates are, however, only possible for significant determinants of matching which happen to be recorded in both

datasets (for example age and class, but not severity, in this study). Such adjustments may even serve to disguise ignorance of other unobserved ways in which the two contributing lists are not equal probability samples of the target set.

Faced with two such overlapping lists, and once every effort has been made to ensure that truly identical cases have been matched, then the unmatched cases must be intensively scrutinised in an attempt to identify those characteristics that might act to exclude other cases from both lists (the undercount). It is not here contended that there may not be such “undercount”, rather that their numbers cannot be properly estimated, or the distribution of their risk factors recognised by the mark/recapture technique. Injured children are not goldfish.

We would like to acknowledge the help of the Northumbria Police and the Gateshead Traffic Accident Data Unit in the provision of Stats19 data and the Department of Health for the selected HES records. The work was partially funded by the NHSE Research and Development Programme in Maternal and Child Health.

- 1 International Working Group for Disease Monitoring and Forecasting. Capture-recapture and multiple-record systems estimation. I: History and theoretical development. *Am J Epidemiol* 1995;142:1046–58.
- 2 International Working Group for Disease Monitoring and Forecasting. Capture-recapture and multiple-record systems estimation. II: Applications in human disease. *Am J Epidemiol* 1995;142:1059–77.
- 3 Razzak J, Luby S. Estimating deaths and injuries due to road traffic accidents in Karachi, Pakistan through the capture-recapture method. *Int J Epidemiol* 1998;27:866–70.
- 4 Hickman M, Sutcliffe H, Sondhi A, et al. Surveillance of problem drug use in the UK: a review of a regional drug misuse database. *J Public Health Med* 1999;21:271–7.
- 5 Laporte R. Assessing the human condition: capture-recapture techniques. *BMJ* 1994;308:5–6.
- 6 Martyn C. Capture-recapture methods in surveys of diseases of the nervous system. *J Neurol Neurosurg Psychiatry* 1998;64:2–3.
- 7 Roberts I, Scragg R. Application of capture “recapture” methodology to estimate the completeness of child injury surveillance. *J Paediatr Child Health* 1994;30:513–14.
- 8 Laporte R, Dearwater S, Chang Y, et al. Efficiency and accuracy of disease monitoring systems: application of capture-recapture methods in injury monitoring. *Am J Epidemiol* 1995;142:1069–77.
- 9 Seber G. *The estimation of animal abundance and related parameters*. 2nd Ed. London: Griffin, 1982.
- 10 Cormack R. The statistics of capture/recapture methods. *Annual Review of Oceanography and Marine Biology* 1968;6:455–506.
- 11 World Health Organisation. *ICD-10. International statistical classification of diseases and related health problems. 10th revision. Volume 1*. Geneva: WHO, 1992.
- 12 Walsh S, Barton S, Jarvis S, et al. *Children injured in road traffic accidents: a collaborative project with Northumbria Police*. Newcastle upon Tyne: North East Regional Research Laboratory, University of Newcastle upon Tyne, 1990.
- 13 Department of Transport. *Instructions for the completion of road accident reports—STATS20*. London: DoT, 1994.
- 14 Cormack R. Discussion of “A simple EM algorithm for capture-recapture data with categorical covariates” by GG Baker. *Biometrics* 1990;46:1198–200.
- 15 Baker R, Nelder J. *GLIM manual (release 3.77)*. Oxford: Royal Statistical Society, 1987.
- 16 Simpson H. *Comparison of hospital and police casualty data: a national study*. TRL report 173. Crowthorne, UK: Transport Research Laboratory, 1996.
- 17 James H. Under-reporting of RTAs. *Traffic Engineering and Control* 1991;32:574–83.
- 18 Hopkin J, Murray P, Pitcher M, et al. *Police and hospital recoding of non fatal road accident casualties: a study in Greater Manchester*. Research report 379. Crowthorne, UK: Transport Research Laboratory, 1993.
- 19 Tilling K, Sterne J. Capture-recapture models including covariate effects. *Am J Epidemiol* 1999;149:392–400.