**Use of the capture-recapture technique to evaluate the completeness of systematic literature searches**

BMJ 1996; 313 doi: http://dx.doi.org/10.1136/bmj.313.7053.342 (Published 10 August 1996)

**Cite this as:** BMJ 1996;313:342

1. Pat Spoor*, information manager* (p.a.spoor@leeds.ac.uk)[a](http://www.bmj.com/content/313/7053/342#aff-1),
2. Mark Airey*, senior research fellow*[a](http://www.bmj.com/content/313/7053/342#aff-1),
3. Cathy Bennett*, research fellow*[a](http://www.bmj.com/content/313/7053/342#aff-1),
4. Julie Greensill*, research assistant*[a](http://www.bmj.com/content/313/7053/342#aff-1),
5. Rhys Williams*, professor of epidemiology and public health*[a](http://www.bmj.com/content/313/7053/342#aff-1)

Author Affiliations

1. Correspondence to: Pat Spoor.
* **Accepted**6 March 1996

Capture-recapture methods were pioneered in ecology and derive their name from censuses of wildlife in which several animals are captured, marked, released, and subject to recapture. In epidemiology the technique examines the degree of overlap between two (or more) methods of ascertainment and uses a simple formula to estimate the total size of the population. When the number already identified is subtracted from this estimate the number of cases not ascertained by either (or any) of the methods can then be calculated. It has been suggested that studies which attempt to ascertain all cases of a given disease in a population should use this method to estimate the number of missing cases.[1](http://www.bmj.com/content/313/7053/342#ref-1) [2](http://www.bmj.com/content/313/7053/342#ref-2)

There are direct parallels between epidemiological studies which attempt to ascertain all available cases and systematic literature searches which attempt to identify all publications on a given topic: both should incorporate estimates of the number of cases or publications they fail to identify. Our study compared, for one journal, the results of searching an electronic literature database with those of hand searching, both carried out for the Cochrane collaborative review group on diabetes.[3](http://www.bmj.com/content/313/7053/342#ref-3)

**Methods and results**

The Medline database was searched from January 1984 to October 1994 for articles in Diabetic Medicine likely to be describing randomised controlled trials, as defined by Dickersin et al and using their search strategy.[4](http://www.bmj.com/content/313/7053/342#ref-4) Independently, a handsearch of the journal for the same period was carried out, with the same aim.

The maximum likelihood estimator, N = M(n/m), was used to estimate the total population size,[2](http://www.bmj.com/content/313/7053/342#ref-2) where M is the number of publications identified by Medline, n the number identified by hand searching, and m the number identified by both sources. The estimated number unidentified by either method was calculated by subtraction. The maximum likelihood estimator is biased for small samples, for which Chapman's method is more appropriate.[5](http://www.bmj.com/content/313/7053/342#ref-5) This estimates the total population size as N = (M + 1)(n + 1)/(m + 1) - 1. The variance of N is estimated as Var(N) = (M + 1)(n + 1) (M - m)(n - m)/((m + 1)2(m + 2)), from which 95% confidence intervals can be constructed.

Table [1](http://www.bmj.com/content/313/7053/342#T1) shows the number of publications identified by each method and the overlap. The articles missed by the hand search are attributed to human error; those not identified by the Medline search were improperly indexed, either because until recently no appropriate methodological subject heading existed or because the abstract failed to describe the study design. For our data both the maximum likelihood estimator and Chapman's method gave the same estimate of total population size (160, 95% confidence interval 158 to 164) rounded to the nearest whole number. The number of articles “missed” was 2 (0 to 6).

View this table:

* [View Popup](http://www.bmj.com/highwire/markup/423165/expansion?width=1000&height=500&iframe=true)
* [View Inline](http://www.bmj.com/content/313/7053/342)

**Table 1**

Extent of overlap in the number of publications found by Medline search of “Diabetic Medicine” (January 1984 to October 1994) and by hand searching

**Comment**

A caveat to the application of these methods is that if there is positive dependency between the two sources— that is, if an article identified by hand searching is more likely to be ascertained in Medline than one not so identified—then the estimates will underestimate the true population. If, however, Medline and the hand search are negatively dependent then the estimates will overestimate the true population.[2](http://www.bmj.com/content/313/7053/342#ref-2) Log-linear modelling offers an alternative approach to modelling dependency among data, where it is present.

The term capture-recapture is not so appropriate for the technique's use in epidemiology or literature searches since, while cases and publications may be said to be “captured,” nothing is being “recaptured.” As applied in epidemiology the method has been termed “ascertainment intersection.”[2](http://www.bmj.com/content/313/7053/342#ref-2)However, we suggest the more informative descriptor “comparison of multiple methods of ascertainment” (or COMMA) for this useful technique, which we advocate for all systematic literature searches.

We thank Mr Alan S Rigby for statistical advice.

**Footnotes**

* Funding Northern and Yorkshire Regional Health Authority and the British Diabetic Association.
* Conflict of interest None.

**References**

1. 1.[↵](http://www.bmj.com/content/313/7053/342#xref-ref-1-1)
	1. LaPorte R

.Assessing the human condition: capture-recapture techniques.BMJ 1994;308: 5–6.

[FREE Full Text](http://www.bmj.com/lookup/ijlink?linkType=FULL&journalCode=bmj&resid=308/6920/5&atom=%2Fbmj%2F313%2F7053%2F342.atom)

1. 2.[↵](http://www.bmj.com/content/313/7053/342#xref-ref-2-1)
	1. Hook EB,
	2. Regal RR

.The value of capture-recapture methods even for apparent exhaustive surveys. The need for adjustment for sources of ascertainment intersection in attempted complete prevalence studies.Am J Epidemiol 1992;135:1060–7.

[Abstract/FREE Full Text](http://www.bmj.com/lookup/ijlink?linkType=ABST&journalCode=amjepid&resid=135/9/1060&atom=%2Fbmj%2F313%2F7053%2F342.atom)

1. 3.[↵](http://www.bmj.com/content/313/7053/342#xref-ref-3-1)
	1. Airey CM,
	2. Williams DRR

.Cochrane Collaborative Review Group: Diabetes.Diabet Med 1995;12: 375–6.

[Medline](http://www.bmj.com/lookup/external-ref?access_num=7648796&link_type=MED&atom=%2Fbmj%2F313%2F7053%2F342.atom)[Web of Science](http://www.bmj.com/lookup/external-ref?access_num=A1995QX67600001&link_type=ISI)

1. 4.[↵](http://www.bmj.com/content/313/7053/342#xref-ref-4-1)
	1. Dickersin K,
	2. Scherer R,
	3. Lefebvre C

.Identifying relevant studies for systematic reviews.BMJ 1994;309: 1286–91.

[Abstract/FREE Full Text](http://www.bmj.com/lookup/ijlink?linkType=ABST&journalCode=bmj&resid=309/6964/1286&atom=%2Fbmj%2F313%2F7053%2F342.atom)

1. 5.[↵](http://www.bmj.com/content/313/7053/342#xref-ref-5-1)
	1. Chapman DG

.Some properties of the hypergeometric distribution with applications to zoological sample census.University of California Public Statistics 1951;1: 131–60.