Epidemiologic Reviews
Vol. 17, No. 2
Copyright © 1995 by The Johns Hopkins University School of Hygiene and Public Health
All rights reserved
Printed in U.S.A.

Capture-Recapture Methods in Epidemiology: Methods and Limitations

Ernest B. Hook[1] and Ronald R. Regal[2]

BACKGROUND

Capture-recapture methods in epidemiology are attempts to estimate or adjust for the extent of incomplete ascertainment using information from overlapping lists of cases from distinct sources. As the name implies, these methods have their antecedents in animal ecology (1), from which they draw their name. (A related earlier but not identical approach to population estimation was undertaken by LaPlace (2), as cited by Seber (3).) Demographers often use the term "dual" (i.e., two-source) or "multiple record systems" to denote capture-recapture approaches to estimates of population size (4, 5). Some other terms that have been used by epidemiologists or biostatisticians to describe estimates derived using these methods are "ascertainment corrected rates" (6), "Bernoulli census" estimates (7), or "closed" or "open" population estimates, although technically the latter refer to the nature of the population being estimated, not to the methods used to estimate them. Thus, such methods appear in the literature under a variety of different terms. Consistent usage of the term "capture-recapture" instead of other terms could ensure that workers in all fields readily retrieve pertinent work in the literature.

SOME EPIDEMIOLOGIC APPLICATIONS

There are several different potential usages of capture-recapture methods in epidemiology, all related to estimation of the population affected. These include 1) estimation when the investigator has clearly incomplete data available from two or more sources; 2) refinement of prevalence or incidence estimates derived from attempted exhaustive population surveys; 3) attempted evaluation only of source completeness, or of completeness of a registry that receives reports from many different sources; and 4) attempts to derive only plausible upper or lower limits on the total affected. We describe some examples of these approaches.

After prenatal diagnosis became widely available, Hook and Chambers (8) wished to derive maternal age-specific rates of Down syndrome by 1-year maternal age interval, to enable more accurate genetic counseling. The only pertinent population-based data source readily available to them was the reports of this diagnosis on birth certificates in Upstate New York. Reporting of diagnoses on birth certificates was known to be incomplete in other jurisdictions, but they had no information at hand for their own data set. After matching a list of cytogenetically diagnosed cases obtained from laboratories in the jurisdiction to names of those affected on birth certificates, they undertook a capture-recapture analysis and estimated that the birth certificates were about 35 percent complete. They adjusted the observed age-specific rates on birth certificates accordingly, to derive risk figures for genetic counseling (8). This was the first single year maternal age schedule that was published, and while there were problems detected later in some aspects of the analysis (9), the derived rates have been remarkably close to those found in later much more extensive analyses and sophisticated studies (10–12).

To estimate the prevalence of encephalopathies and three different types of seizure disorders in children, Bobo et al. (13) established a comprehensive surveillance system in Oregon and Washington, which attempted to identify all those affected. Rather than assume that they had found all cases, they applied a capture-recapture analysis and adjusted the observed frequencies accordingly. Bobo et al. used several different sources, and thus they were able to adjust for possible source dependence (a concept discussed further below). Capture-recapture adjustment derived a rate for afebrile seizures that was 45 percent greater than the observed rate. Other disorders studied in this surveillance system had much higher completeness of reporting; the upward correction of complex febrile

Received for publication September 28, 1994, and in final form October 20, 1995.

Abbreviations: MLE, maximum likelihood estimator; AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion; d.f., degrees of freedom; HIV, human immunodeficiency virus.

[1] School of Public Health, University of California, Berkeley, CA, and Department of Pediatrics, University of California, San Francisco, CA.

[2] Department of Mathematics and Statistics, University of Minnesota—Duluth, Duluth, MN.

Reprint requests to Prof. Ernest B. Hook, University of California, School of Public Health, Warren Hall, Berkeley, CA 94720–7360.

Downloaded from academic.oup.com/epirev/article/17/2/243/499931 by Michigan State University user on 12 October 2026

seizures was only 11 percent. The latter disorder is much more severe, and thus greater estimated ascertainment by the same sources appears plausible.

Robles et al. (14), working with data from a cancer registry in Ontario, wished to estimate the completeness of reporting in 1981 of a variety of malignancies. They were able to do so readily by capture-recapture methods, because, unlike many other registries, they had distinguished the multiple sources of reporting of each case, and could therefore undertake matching of reports to different sources. They determined that for the vast majority of malignancies, reporting was likely to be very close to complete (90 percent or over) and for all reports it was about 95 percent complete, an inference which appeared plausible before the analysis but which they were at least able to substantiate quantitatively. They did not attempt to derive corrected incidence and prevalence figures but of course could have done so readily from these completeness estimates. We consider below a later report from this same registry for the subsequent year (15).

Egeland et al. (16) used data from relatively available distinct data sources to estimate the prevalence of fetal alcohol syndrome in Alaskan native American children. The frequency of this disorder is notoriously difficult to establish. In at least one age group, they were unable to estimate a point estimate, but derived a plausible upper boundary, a figure still of value for policy and planning purposes.

These are just four examples of many that could be cited of epidemiologic use of these methods. For those interested in other applications in the literature, we cite in our list of references those that were published or of which we were aware before January 1, 1995. These include the reports of Abeni et al. (17); Alho (18); Barg and Huether (19); Benn et al. (20); Bishop et al. (21); Bloor et al. (22); Bobo et al. (13); Brecht and Wickens (23); Bruno et al. (24); Chapman (25); Chiu et al. (26); Cochi et al. (27); Davis et al. (28); Doscher and Woodward (29); Drucker and Vermund (30); Espeland et al. (31); Fisher et al. (32); Frischer (33); Frischer et al. (34–36); Goldberg and Wittes (37); Gutteridge and Collin (38); Hartnoll et al. (39); Hewitt and Milner (40); Hilsenbeck et al. (41); Hook and Chambers (8); Hook et al. (42); Hook (42a); Hook and Regal (43, 44); Hser (45, 46); Huether et al. (47); Kehoe et al. (48); Kozinetz et al. (49); LaPorte et al. (50, 51); Lewis and Hassanein (52); Lie et al. (53); Mastro et al. (54); McCarty et al. (6, 55); McKeganey et al. (56); Modesitt et al. (57); Neugebauer (58); Newmeyer (59); Palfrey et al. (60); Prevots et al. (61); Robles et al. (14); Rubin et al. (62); Sachs et al. (63); Schouten et al. (63a); Serrano Rios et al. (64); Sutter et al. (65); Sutter and Cochi (66); Taylor (67); van Haastrecht et al. (68); Walter et al. (15); Watts et al. (69); Wickens (70); Wittes (7); Wittes and Sidel (71); Wittes et al. (72); Woodward et al. (73); and York and Madigan (74, 75).

METHODOLOGIES: BACKGROUND

From a historical epidemiologic perspective, we believe the most significant contributions have been those of Wittes who, to our knowledge, first introduced and adapted these approaches to epidemiology (7, 71, 72), and of Fienberg (76) (see also Bishop et al. (21)) who demonstrated how log-linear models may be applied to adjust for dependencies among multiple sources. This overcame the need to assume that all sources were independent, a previous limitation which, while of lesser concern to ecologists and demographers, had probably considerably restricted application of capture-recapture methods by epidemiologists. Log-linear methods are, we believe, the methods of choice for data from multiple sources. But they are so relatively easily applied in such instances that investigators run the risk of losing sight not only of their intrinsic limitations but also of the great ancillary information that may be provided by simpler analyses of the data even when multiple sources are available.

Extensive reviews of the methodology and lists of references relevant both to applications in ecology and demography as well as some background to considerations in epidemiology appear in Seber (3, 77, 78); El-Khorazaty et al. (5); and Fienberg (79).

In this review, we present first the nature and rationale of the usual capture-recapture approaches in epidemiology. We discuss critically their application and consider their limitations and possible methods to circumvent these, and briefly consider alternatives.

THE NATURE OF A "SOURCE"

The term "source" is widely used in applications of capture-recapture methods in epidemiology. It denotes a list of cases usually but not necessarily with some unifying quality about their ascertainment. We may distinguish an "original" source, e.g., a particular list of cases from some institution, from an "analytic" source, which is a list which may be constructed by the investigator from one or more original sources.

Any application of capture-recapture methods to epidemiology implicitly or explicitly makes a critical decision as to how to define the analytic sources. Certainly, in some instances, how to do so may appear obvious to the investigator. For example, one analytic source might be all cases whose diagnoses appear on death certificates in the jurisdiction and time interval

Downloaded from academic.oup.com/epirev/article/17/2/243/499931 by Michigan State University user on 12 October 2026

covered by the investigation. But different lists may be pooled to derive a larger group that may be considered as an additional source for analysis. For example, if there are several hospitals within a jurisdiction, one could treat the lists of in-patients from each as a separate analytic source. But because they are likely to cover quite different areas, one might well prefer to pool the lists from all such institutions into one list that is used as a single source in further analyses. However, even hospitals may differ among themselves, and one might prefer to keep as one analytic source the records of a referral hospital which draws cases from throughout a region, and only pool those of local hospitals.

As this discussion indicates, the investigator must use flexibility and judgment in constructing the analytic sources, taking into account not only the nature of the original sources, but whether the numbers are in any sufficient to warrant treating it as distinct for purposes of analysis. Wittes (7) appears to have been the first to discuss these alternative approaches explicitly at least with regard to epidemiologic analyses. If there are both "major" and "minor" original sources with large and small numbers of cases, respectively, the investigator may choose to pool all the minor sources together and treat them as one larger analytical source, to fuse some or all of them with various of the major sources, or to ignore them altogether and only use data from the major sources in analyses. Indeed, the investigator may wish to attempt several different methods of fusing these lists and defining analytic sources.

In animal population surveys, the analog of original sources is usually the result of successive trappings of animals. Thus, in animal ecology, the "sources" have a natural temporal ordering which is usually lacking in epidemiologic application. Moreover, the traps are usually "planned" to optimize sampling of the population. Because of this difference in the nature of sources, a great deal of literature on capture-methodology in the animal ecology literature is not directly relevant to epidemiologic application.

DATA COLLECTION

An investigator who undertakes a prevalence or incidence study using two or more data sources can ultimately only apply capture-recapture methods if he or she has collected the data in a format which allows one to derive the number in all possible "intersections" of sources.

Many registries collect data from multiple sources but do not note in their records a subsequent report after the first notification of an affected individual. This precludes capture-recapture analysis, at least on the collected information. But each new report of a case requires matching against existing records to determine if the case has been previously reported. Instead of just discarding such potentially useful information, the registry need expend little extra effort to note the fact and nature of any new ascertainment.

The diagnostic quality of the data and the accuracy of identifying information that enables matching are of course also critical issues. We discuss these latter items further in the section below on "Limitations and Complexities."

REPORTING OF DATA

An investigator who collects or uses data from two or more data sources for prevalence or incidence estimates should attempt not only to keep track of the intersection of ascertainment sources, but report them in any publication, irrespective of utilization of capture-recapture estimates to refine estimates. This will enable others at least to explore the plausibility of the overall estimate derived from simply adding all distinct cases that are found. For example, the useful reports of data on Huntington disease by Folstein (80) and on amyotrophic lateral sclerosis by Annegers et al. (81) enabled others to explore the application of capture-recapture estimates to the findings in these surveys (6, 44).

In addition, we strongly urge that editors of journals insist that authors present the observed data from which they derived capture-recapture estimates. If there are so many covariates that it is not feasible to present the information at each level of covariate then 1) the information pooled over the entire population and some substrata should at least be presented, and 2) the more extensive data should be deposited in a data repository.

There are several interrelated reasons for these recommendations. First, providing the raw data enables others to confirm that at least an acceptable estimate has been presented. Second, new investigators could derive any other estimate that may seem preferable based on other decision criteria (see below). In other words, subsequent workers could check to see if at least the estimates have been calculated correctly, and they could explore further implications of the data and other models than those selected by the original investigators.

SOURCE DEPENDENCY

The concept of source dependency is critical to an understanding of the application of capture-recapture methods. Sources are "dependent" if they are not "independent." Any two sources A and B are independent if the overall, i.e., average, probability of members of

Downloaded from academic.oup.com/epirev/article/17/2/243/499931 by Michigan State University user on 12 October 2026

a population who appear in their intersection or "overlap" is equal to the product of the average probabilities of appearing in A, and in B. The extremes of dependence for two sources may be readily conceptualized. If two sources (say, for example, with probabilities of ascertainment of 0.2 and 0.3) have no cases in common whatsoever, then clearly the probability of being in both of them is zero. This of course is less than the product of the probabilities of being in any one of them (0.06), so the sources are dependent. For such a situation, by convention, we speak of a "negative" dependence, as the difference (0 − 0.06) is negative. If, however, the sources are nested so that all of the cases in the smaller source (with probability 0.2 in this case) are within the larger source (probability 0.3), then the probability of being in both sources is of course that of being in the smaller source alone, 0.2 in this example, and greater than the product of the probabilities of being in either source separately, 0.06. Again the sources are dependent, but the dependence is "positive" by convention, because the difference (0.2 − 0.06) is greater than zero. As we will elaborate below, positive dependence of sources will tend to produce an underestimate of the true population size, and negative dependence will tend to produce an overestimate.

Multiple sources may be said, analogously, to be independent if the overall (or "average") probability of members of a population being in the overlap of any particular subset of all the sources is equal to the product of the (overall) probabilities of being in any source within this subset. Multiple sources are dependent if not independent. While there is no formal definition of "positive" or "negative" dependence of multiple sources of which we are aware, we suggest an operational definition analogous to the outcomes for two-source dependencies. If the assumption of independent multiple sources results in a capture-recapture estimate that is an underestimate of the true total, then we suggest defining these sources as having "net" or "overall" positive dependence, and if an overestimate of the true total, we suggest defining the sources as having "net" or "overall" negative dependence. The qualifying terms are necessary because the various subsets of multiple sources may differ considerably in the nature of their dependencies and the (net) dependency of any particular subset of sources may be quite different from that of all sources considered together.

The extreme deviations from independence for multiple sources occur in analogous situations to those considered above for two sources. If sources are all disjoint, with no cases in common with any other, then this is the extreme of negative dependence of multiple sources, and indeed all subsets of the multiple sources are also negatively dependent. If the sources are nested from smallest to largest within each other, this is the realization of the extreme of positive dependence. And of course any subset of sources also has net positive dependence in this case.

TWO-SOURCE ANALYSES

As implied above, we may subdivide capture-recapture approaches into two-source and multiple-source analyses. The former is in some ways the most useful. The latter is more powerful, but often is associated with greater difficulty in interpretation of derived estimates.

Table 1 explains the basic rationale of the application of these methods for the simplest case in which there are only two sources. In the text that follows, we will use the terms "a," "b," "c," and "x" to denote both the names of the cells themselves and the total in each cell. The context will make clear the particular meaning. One may estimate either the missing cases or the complete population affected using the formulae given in table 1. The maximum likelihood estimator (MLE) indicated is so intuitively evident that it has been developed independently by Petersen (1), Lincoln (82), and a number of other investigators. The term "Petersen estimate" or "Lincoln-Petersen estimate" has been most widely employed. Use of the adjustment originally suggested by Chapman (83), and shown by Wittes (84) to have optimal properties under a wide range of conditions, will diminish the small sample bias of the maximum likelihood estimate and will result in what may be termed a "nearly unbiased estimator."

As illustrated in table 1, investigators who conduct an ad hoc study of the completeness of ascertainment of a specific source are in essence providing data pertinent to a capture-recapture estimate of the size of the entire group affected, even though they may not use such information to derive such estimates. The two-source capture-recapture population estimate can also be expressed as the number of observed cases in any source, say Y_0 in table 1, divided by the estimate of source completeness derived by matching information in any other source, Z (or estimated in any other manner (see below)). More generally, of course, any estimate of completeness of pooled multiple sources, e.g., of a registry of cases, may analogously generate an estimate of the total affected.

As an aside, we note that if an investigator estimates source completeness not by matching but in another manner, then there is no precise analogy with capture-recapture methods as generally used in epidemiology. Application of completeness estimates derived in this way to calculate a population total is generally termed

Downloaded from academic.oup.com/epirev/article/17/2/243/499931 by Michigan State University user on 12 October 2026

TABLE 1. Two-source model

		Source Y: Yes	Source Y: No	Total
Source Z	Yes	a	b	$a + b = Z_0$
	No	c	x	
	Total	$a + c = Y_0$		$N = a + b + c + x$

Estimated values		Maximum likelihood estimator (MLE)	Nearly unbiased estimator (NUE)
Unobserved cell:	$\hat{x}$	bc/a	$bc/(a + 1)$
Completeness of source Y:	$\hat{Y}_c$	$a/(a + b) = a/Z_0$	$(a + 1)/(a + b) = (a + 1)/Z_0$
Completeness of source Z:	$\hat{Z}_c$	$a/(a + c) = a/Y_0$	$(a + 1)/(a + c) = (a + 1)/Y_0$
Total population:	$\hat{N}$	$a + b + c + (bc/a)$ or, $(a + b)(a + c)/a$ or, $Y_0/\hat{Y}_c$ or, $Z_0/\hat{Z}_c$	$a + b + c + (bc/(a + 1))$ or, $[(b + 1)(c + 1)/(a + 1)] - 1$

the ratio estimator method. As an example, Newmeyer (59) (cited by Brecht and Wickens (23)) estimated the number of heroin users in San Francisco, California, by independently obtaining an estimate of the probability of death of a heroin user in any one year, roughly 1/200. He divided the total number of recorded heroin deaths in San Francisco, 76, by this figure to estimate the total number of heroin users as 76/0.005, or about 15,000. But he did not match the 76 observed deaths with the recorded number in another source to derive the probability of death. (LaPlace (2) used a similar approach in his estimate of the population of France.) While the logic of the method is similar to that used in capture-recapture analysis, we suggest the latter term be reserved, in epidemiology at least, for those instances in which matching of lists is used to derive the completeness estimate.

UNDERLYING ASSUMPTIONS

Application of capture-recapture methods in essence makes several assumptions. Some of these are implicit, i.e., that each case has been diagnosed accurately, that "matching" of cases that appear in different ascertainment sources has been done appropriately, and that the cases included are in fact within the time-space unit under study. Some further assumptions usually more explicitly recognized are: 1) that for any single source, each case in the population has the same "catchability" or probability of ascertainment (although any two sources may differ in this probability), 2) that for two sources at least, ascertainment of any case by the sources is independent, and 3) that the population under study is "closed," i.e., that there are no entries or losses during the study period.

In fact, assumption 3 is subsumed in assumption 1. Cases that stay in a jurisdiction clearly are likely to have higher catchability than those who migrate in or out, and cases that die may be selectively reported by pathologists' records and other sources, but be less likely to be ascertained by some other sources.

As we discuss below, from some analytical perspectives, assumptions 1 and 2 each imply the other (see below) and, moreover, all variation from independence may originate biologically from variable catchability. Nevertheless, for heuristic purposes, they are best regarded, we believe, as distinct.

Even when the explicit assumptions are violated, one may use two-source capture-recapture methods to derive useful estimates that may be of considerable practical value, as we will elaborate on further. Here we note that most violations of these explicit assumptions may be modeled most readily as variations of the assumption of independence alone. Therefore, we consider this condition first.

Downloaded from academic.oup.com/epirev/article/17/2/243/499931 by Michigan State University user on 12 October 2026

If the investigator can predict the likely direction of any dependency of two sources in the entire population, then he or she can determine if the two-source capture-recapture estimate is likely to be a plausible under- or overestimate of the true total, a result which may be very useful. By application of numerous two-source analyses, Hook and Regal (44) were able to derive plausible lower and upper boundaries on the prevalence of Huntington disease in one jurisdiction. By knowing that two sources used were, if anything, negatively dependent, Egeland et al. (16) put a plausible upper limit on chart-verifiable cases of fetal alcohol syndrome in Alaska in one age group. In many instances, knowledge of a plausible upper or lower boundary derived from two sources may be more useful for public health policy purposes than an estimate, of unknown validity, of the true total derived from multiple sources.

MULTIPLE SOURCES

When data from more than two sources are available, the investigator's approach is more complex because multiple estimates are possible. These complexities are most easily illustrated by considering three sources, the simplest multiple-source case. Table 2 illustrates the data layout and the application of log-linear methods developed for derivation of an esti-

TABLE 2. Three-source model

		Source 1: Yes		Source 1: No	
		Source 2: Yes	Source 2: No	Source 2: Yes	Source 2: No
Source 3: Yes		a	b	e	f
Source 3: No		c	d	g	x

$$N_{obs} = a + b + c + d + e + f + g$$
$$N_1 = a + b + c + d$$
$$N_2 = a + c + e + g$$
$$N_3 = a + b + e + f$$

Maximum likelihood estimates of x using alternative models

d.f.*	Model	Estimator†
3	Independent	$\hat{x} = \hat{N} - N_{obs}$ ($\hat{N}$ is the solution of $(\hat{N} - N_1)(\hat{N} - N_2)(\hat{N} - N_3) = \hat{N}^2(\hat{N} - N)$)
2	1–2 interactions	$\hat{x} = (c + d + g)(f)/(a + b + e)$
2	1–3 interactions	$\hat{x} = (c + d + f)(g)/(a + c + e)$
2	2–3 interactions	$\hat{x} = (e + f + g)(d)/(a + b + c)$
1	1–2, 1–3 interactions	$\hat{x} = gf/e$
1	1–2, 2–3 interactions	$\hat{x} = gd/c$
1	1–3, 2–3 interactions	$\hat{x} = df/b$
0	1–2, 1–3, 2–3 interactions	$\hat{x} = (adfg)/(bce)$

Two-source restricted estimators of same population*

	$\hat{N}_{MLE}$	$\hat{N}_{NUE}$
1 versus 2	$(N_1)(N_2)/(a + c)$	$(N_1)(N_2)/(a + c + 1)$
1 versus 3	$(N_1)(N_3)/(a + b)$	$(N_1)(N_3)/(a + b + 1)$
2 versus 3	$(N_2)(N_3)/(a + e)$	$(N_2)(N_3)/(a + e + 1)$

* d.f., degrees of freedom; MLE, maximum likelihood estimator; NUE, nearly unbiased estimator.
† Where not given explicitly, $\hat{N} = \hat{x} + N_{obs}$.

Downloaded from academic.oup.com/epirev/article/17/2/243/499931 by Michigan State University user on 12 October 2026

mate. As above, we use the terms *a*, *b*, etc., to denote both the names of the cells and the number observed in each, and *x* for the unobserved cell and the number in it.

To derive an estimate using the usual log-linear approach, the investigator assumes there is no "three-way" interaction of sources, and, more generally, if there are *k* sources available that there are at most $k - 1$ source interactions. With three sources, there are eight possible models (using all the data) excluding the three-way interaction. The associated estimates are illustrated in table 2. As the number of sources increases, the number of possible models increases quite rapidly. With four sources, there are 114 different models (using all the data), so, excluding the four-way interaction, there are 113 associated estimates. With five sources, there are 6,893 analogous estimates excluding the five-way interaction (W. J. Redfearn, University of California, Berkeley, personal communication, 1994).

In the discussion below, we will concentrate on four general types of models: 1) the independent model, which assumes all sources are independent; 2) models that reduce to or are equivalent to two independent sources (e.g., models 1–2, 1–3, and 2–3 in table 3) or two independent subsets of sources (e.g., models 1–2, 1–3; 1–2, 2–3; and 1–3, 2–3 in table 3), which we designate as "full two-source models"; 3) the model that assumes all possible interactions except that of *k* sources (type 8 in table 3); because there are no degrees of freedom in this model, we may denote it as "saturated"; and 4) other models. (There are none such in the three-source case.)

We denote these as "full-source" estimates because they use or are based on considerations derived from the complete data set. There are also other types of estimates. For instance, as we illustrate in the bottom of table 2, if there are three sources, then there are three different possible two-source estimates that one could derive by ignoring information in one of the

TABLE 3. Spina bifida in black livebirths, Upstate New York, 1969–1974

		Source 1*				
		Yes		No		
		Source 2*		Source 2		
		Yes	No	Yes	No	*x*
Source 3*	Yes	0	3	1	8	
	No	8	13	3	0	$N_{obs} = 36$

Three-source estimates							
Model	d.f.†	G^2	AIC†	BIC†	$\hat{x}$	$\hat{N}$	90% CI†
Independent	3	8.3	2.3	3.0	18	54	44–77
1–2	2	3.0	−1.0	−0.5	48	84	52–186
1–3	2	4.5	0.5	1.0	8	44	38–60
2–3	2	6.1	2.1	2.6	14	50	41–69
1–2, 1–3	1	2.6	0.6	0.9	24	60	40–311
1–2, 2–3	1	2.4	0.4	0.6	34	70	46–165
1–3, 2–3	1	0.6	−1.4	−1.1	4	40	37–52
1–2, 1–3, 2–3	0	0	0	0	0	36	36–89
Weighted BIC						54	38–75

Restricted two-source estimates*		
1 versus 2	$\hat{N} = 96$	$\hat{N}_{NUE} = 72$
1 versus 3	$\hat{N} = 36$	$\hat{N}_{NUE} = 32$
2 versus 3	$\hat{N} = 144$	$\hat{N}_{NUE} = 72$

* Source 1 = birth certificate reports; Source 2 = death certificate records; Source 3 = medical rehabilitation files. Data from Hook et al., *Am J Epidemiol* 1980;112:750–8 (42), and Hook, *Am J Epidemiol* 1993;137:1285 (42a).

† d.f., degrees of freedom; AIC, Akaike Information Criterion; BIC, Bayesian Information Criterion; CI, confidence interval; NUE, nearly unbiased estimator.

Downloaded from academic.oup.com/epirev/article/17/2/243/499931 by Michigan State University user on 12 October 2026

sources. We term these "restricted" two-source estimates to distinguish them from the "full" two-source estimates, type 2 above.

To derive an estimate from such an analysis, the investigator has several options. We consider first non-Bayesian methods which involve choosing the estimate associated with the single model judged optimal, usually as evaluated by some rather informal procedure related to the value of the likelihood ratio statistic, G^2.

For any model i and observed data set,

$$G^2 = -2 \sum \text{Obs}_j \log (\text{Obs}_j/\text{Exp}_{ji}), \quad (1)$$

where Obs_j is the observed number in each cell, j, and Exp_{ji} denotes the expected number in cell j under model i. The lower the value of G^2, the better the fit of the model. Investigators generally choose the least complex model whose fit appears "adequate," i.e., is not rejected at some nominal level of statistical significance.

Formal procedures for making such a choice have been described but have not been widely used in capture-recapture applications. Here we describe two such procedures and also a method that uses one of these to derive a weighted average of all the estimates associated with each model.

An information criterion may be associated with a model that is a function of its likelihood ratio statistic. At least two such information criteria have been proposed in the literature: 1) the criterion proposed by Akaike (85) and designated by him "an information criterion" or AIC, and often termed by others the "Akaike Information Criterion" (86); and 2) the Bayesian Information Criterion (BIC) (87). For log-linear models, the values for these criteria for any model i are as given by the following formulae:

$$\text{AIC} = G^2 - 2 (\text{d.f.}) \quad (2)$$

$$\text{BIC} = G^2 - (\log N_{\text{obs}}/2\pi) (\text{d.f.}), \quad (3)$$

where d.f. is the number of degrees of freedom of the model and N_{obs} is the number of observed cases. (The usual formula for BIC for log-linear models as given, e.g., Agresti (88) omits the 2π divisor. Draper (89) suggests that its inclusion is more appropriate, although not all statisticians agree (90, 91).) The optimal model is the model with the lowest value of the associated information criterion. For the saturated model, because both G^2 and degrees of freedom are zero, AIC and BIC are also zero. Thus, by these criteria, only models with negative values of the criteria are preferable to the saturated model. If there are more than about 46 observed cases, then use of the BIC picks simpler models, i.e., those with more degrees of freedom, than does use of the AIC. Some theoretical considerations suggest BIC is preferable (87), but AIC is more widely used. Large-scale simulations that we have undertaken (92), assuming a total population of 100 and a large variety of different parameters, suggest that, at least under the conditions considered, there is little difference in performance of AIC and BIC. But of course their use will often result in selection of different models for any particular data set.

In general, as the complexity of the model increases, the "stability" of the estimate tends to diminish and the associated confidence interval widens. Note in particular that the estimate derived from the saturated model is very sensitive to small values in the b, c, and e cells that appear in the denominator of the expression for three sources.

If the saturated model is selected by any criterion, the investigator should be particularly cautious about using the associated estimate. This outcome implies that the saturated model, which assumes $k - 1$ interactions, is preferable to any simpler model. The saturated model could be correct. But a complex unknown k-way interaction in the population analyzed would produce the same observed outcome. Without independent information, the investigator could not decide which of these possibilities was more likely. Unfortunately, the converse does not necessarily follow. The fact that a simple model appears to fit the observed data well does not imply there is no k-way interaction present nor that the true total is estimated by any of the models considered. We amplify these considerations below in discussion of the limitations of capture-recapture methods.

These methods all use a single model to derive an estimate. The difficulty with this approach is illustrated in table 3 with the use of data on spina bifida in black livebirths in Upstate New York in 1969–1974 (42). These data are relatively sparse, and we emphasize that one must be cautious about drawing any strong inference about the total number affected, if for no other reason than that the sampling error about any estimate is relatively large. Nevertheless, as they illustrate a great number of issues, we consider them in detail. Model 1–3, 2–3 has the lowest AIC and BIC and predicts four missing cases and 40 total affected, close to the estimate of the saturated model that predicts a population of 36 and, implausibly, no missing cases at all. But the second best model by either criterion, 1–2, predicts 40 missing and 84 total cases, more than double the other estimates. There is still a considerable "probability" that the model that predicts the much larger number is in fact the correct one. The result using either criterion provides a rationale for an

Downloaded from academic.oup.com/epirev/article/17/2/243/499931 by Michigan State University user on 12 October 2026

attempt to adjust for "model uncertainty" by seeking some intermediate estimate (93).

Before addressing this issue, however, we consider how we might use information from the restricted two-source estimates. As indicated in table 3, the 1 versus 3 model provides a total estimate of 32 to 36 cases total depending on which estimator is used. The total number actually found was 36. We believe it highly unlikely that no cases were missed by any source. So the almost certain underestimate from the 1 versus 3 model suggests that sources 1 and 3 are at least somewhat positively dependent in this population and a log-linear approach which included this interaction would be preferable to one that did not. Of course, any such inferences with such sparse data must be made cautiously. Bayesian methods do allow adjustment for model uncertainty. In essence, they weight the estimate from each model by the likelihood of the observed values assuming the model is correct, multiplied by the "prior probability" of the model. (York and Madigan (74, 75) provide detailed discussion and references of numerous approaches to this literature. Some later commentary appears in Zelterman et al. ((94), and Zelterman et al., Bayesian estimates of a population size, unpublished manuscript).) These methods are relatively complex and are, we believe, not as yet routinely applicable for epidemiologic research.

Recently, we have explored a simple approximation to a Bayesian approach that might be applied relatively easily to data from up to four sources and with some modification to any number of sources. This method, which we designate as "weighted BIC," is derived from Draper (89). In essence, the method weights the estimate from each model i by $[1/\exp(\mathrm{BIC}_i/2)]$. Note that the weight associated with the estimate generated by the saturated model is thus 1.0; models with negative values of BIC (and by this criterion preferable to the saturated model) have weights greater than 1.0; models with positive values of BIC have weights less than 1.0. We have found that not only is this method easy to apply, but that in large-scale simulations (92) it performs better overall than approaches that selected a single model using either AIC or BIC.

Of course, *if* the investigator *knows* with certainty which model describes the population, e.g., that all sources are independent, then he or she should use the estimate associated with this model, and assume that any deviation of the observed values from a perfect fit are due solely to statistical fluctuation, whatever the probability of getting the observed result by chance alone. But almost always the investigator cannot be certain of this in advance. The weighted BIC allows easy adjustment for such model uncertainty. In essence, the weighted BIC is an approximation to a Bayesian approach with "flat priors" because each model is accorded an equal prior probability. The weighted BIC could be further adjusted to introduce the investigators' subjective beliefs about the applicability of various models, i.e., to allow for variable prior probabilities of the models. The advantage of such an approach compared with use of flat priors as in the weighted BIC approach would depend on how well the investigator's assigned prior probabilities agreed with the actual relation of the sources in the population.

The weighted BIC may be extended to four sources with 113 possible models. For five sources (6,893 models) or more, one may readily program a strategy which derives an estimate from the subset of models that provide the best fit (i.e., lowest BIC values) but restricts attention to a relatively small number. For example, one might wish to exclude all models whose weight is no more than, say, 1 percent or 10 percent of the optimal model. This is an area of ongoing investigation.

Another issue that affects estimates, especially of small populations, is also illustrated by the data in table 3. The observation of a null entry in one or more cells may result in the saturated model at least producing either a zero estimate (if $a = 0$) as did occur with this data source, an infinite estimate (b, c, and/or $e = 0$), or no possible estimate (if both conditions hold), irrespective of the entries in any other cell. An adjustment for the multiple sources similar to that of Chapman (83) for two sources would avoid the latter two anomalous or implausible outcomes at least, and would also tend to diminish small-sample bias of the estimator. Of course, if the optimal multiple-source model was one which reduced to two sources (type 2 models above), then one could apply Chapman's adjustment readily (95).

Extensive large-scale simulations that assume a population of about 100 indicate that with any method of model selection or with the weighted BIC, among a variety of approaches which add 0.5 or 1.0 to different combinations of cells, the optimal adjustment overall among them is to add 1.0 to the cells b, c, and e of table 2. (Note that these are the cells which are in the denominator of the expression of the estimate associated with the saturated model.) Calculation of the values of G^2, BIC and AIC associated with each model after, rather than before, making any cell adjustment, improves the performance of any method of model selection or of the weighted BIC (92). These observations have been made on three sources. A natural extension to four or more sources is to add 1.0 to all cells which appear in the denominator of the equation for the estimate of the missing cell given by the

Downloaded from academic.oup.com/epirev/article/17/2/243/499931 by Michigan State University user on 12 October 2026

saturated model. Evans and Bonett (96) have recently suggested an alternative adjustment—addition of 0.25 to each cell if there are three sources, and $(0.5)^{k-1}$ if there are k sources. Large-scale simulations for three sources that we have undertaken indicate that this correction performs much less well for log-linear capture-recapture methods than the correction we propose above.

While these approaches improve the *overall* performance of the estimator in simulations, we emphasize that such simulations are on data sets that meet all the underlying assumptions of the methods. In no sense is there any proof or reassurance that, *for any particular observed data on real populations*, their application results in a valid estimate, nor even necessarily produces an estimate closer to the true value than some alternative approach. We illustrate such considerations below.

CONFIDENCE INTERVALS

The utility of any estimate is restricted by any uncertainty about its accuracy. A confidence interval adjusts for the uncertainty associated with sampling fluctuation. Most epidemiologic applications of capture-recapture methods in the literature have used variance-based approaches to calculate a confidence interval. These use multiples of the standard error about the estimate (see Bishop et al. (21)). In general, because of their symmetry, such methods underestimate both the upper and lower limits, although the error may not be large. But uncritical use of a variance-based interval has led some investigators to give a lower limit for the total population that is less than the observed number of cases, which is equivalent to giving a negative value to the number of missing cases, clearly an impossibility. (References in the literature to such instances will be provided on request.)

Methods based on likelihood ratio statistics (97, 98) lack this disadvantage. Because their complexities, as described in the statistical literature, may not be readily accessible to some readers, we describe their essence here. One method (using a "goodness of fit" approach) fixes the "observed" cell entry and enters in the unobserved cell the value (of x) estimated by any particular model. The investigator calculates the likelihood ratio statistic G^2 (as defined above in the section titled Multiple Sources) for this *complete* table under the model chosen. One then varies the value of the unobserved cell in either direction from that estimated and calculates 1) the G^2 for the new associated complete table, and 2) the difference $\Delta\ G^2$ between this last G^2 and that of the original complete table. The 90 percent (95 percent) interval is that set of totals over the complete table associated with a $\Delta\ G^2$ of 2.645 (3.84) or less (97). The approach of Cormack (98) is similar, except that one evaluates not G^2 but what we define here as $C^2 = (G^2 - f(x,N))$, where $f(x,N) = (x^x N!)/(x^x N! N^N)$ and then proceeds analogously to evaluate $\Delta\ C^2$. We expect these methods to perform similarly (W. J. Redfearn, University of California, Berkeley, personal communication, 1994), but their relative behavior has not yet, to our knowledge, been compared with the use of simulations. An exact method that may be applied to the independence model has been described by Pickands and Raghavachari (99).

The primary advantage to a variance-based interval is its ease of calculation, certainly a benefit if a computer is not readily available to the investigator—a circumstance, as Lauckner (100) notes, often not appreciated by commentators in well-developed countries. For the G^2-based method of Regal and Hook, we have found that, with some trial and error, an investigator can derive an interval for two or three sources with a pocket calculator. This result would not be achievable with the use of bootstrap or exact methods.

Lastly, a confidence interval only adjusts for sampling fluctuation, and it does not adjust for any uncertainty as to whether any of the underlying assumptions are violated. As implied above, a related limitation is that a confidence interval about the estimate from any model does not adjust for uncertainty about whether the model itself holds (93). Model uncertainty has been termed the "Achilles heel of statistics" (101) and a "quiet scandal" (Breiman (102), cited by Chatfield (103)). We consider some of the implications of this further below. The posterior interval associated with Bayesian methods (e.g., York and Madigan (74, 75)) does adjust for model uncertainty at the price of considerable additional complexity in calculation. Methods to calculate intervals associated with the weighted BIC method for log-linear capture-recapture analysis are the subject of research (Regal and Hook, unpublished manuscript).

LIMITATIONS AND COMPLEXITIES

The relative ease of calculation of capture-recapture estimates, at least in epidemiologic applications, has tended to obscure some of the difficulties, except perhaps for the problems that arise from issues of model selection discussed above. Most of these limitations pertain to the possibility, indeed almost the certainty, of the violations of the underlying assumptions. Here we consider such issues.

Downloaded from academic.oup.com/epirev/article/17/2/243/499931 by Michigan State University user on 12 October 2026

Accuracy of data

As for the results of all epidemiologic investigation, the credibility of any estimate will be enhanced to the extent that the investigator may be able to investigate and confirm the accuracy of all information used, such as diagnosis, location of the case within the space-time interval analyzed, and appropriate case matching. With capture-recapture methods, errors are highly likely to have a more than an additive effect on estimates. There is little work, of which we know, that models quantitatively the consequences of errors on log-linear capture-recapture estimates in epidemiology.

For the two-sample case, one may easily derive an estimate of the effects of inaccurate diagnosis in a single source. Recall that in the two-sample case, for any source, say Z, if Z_o is the observed number of cases in this source and Z_c is the estimated completeness, then the estimated total is Z_o/Z_c. And if Z_a is the "accuracy" of diagnoses in the source (i.e., the proportion of those in the source diagnosed as affected who in fact are affected), then one may show readily that the appropriate estimate of N is, instead,

$$N = (Z_o/Z_c)/Z_a = (Z_o)(Z_a)/(Z_c). \quad (4)$$

That is, the best population estimate is the observed number in a source divided by the (estimate of) source completeness and multiplied by source accuracy. (Source "accuracy" may be regarded as the "positive predictive value" of being diagnosed and in the source.) This approach assumes the investigator has some prior knowledge of source accuracy or has estimated that parameter from some survey.

The completeness estimate of Z described in earlier sections, $Z_c = a/Y_{obs}$ is only appropriate if in fact the other source, Y, is 100 percent accurate. Data pertinent to such an instance appear in an analysis by Johnson et al. (104) in which Z was birth certificate reports of Down syndrome, which turned out to be about 5–10 percent inaccurate, i.e., $Z_a = 0.90–0.95$, and Y was the reports of cytogenetic laboratories (presumably 100 percent accurate), results that suggest that earlier capture-recapture estimates of maternal age-specific rates of Down syndrome that have used birth certificate reports were biased on this account, giving 5–10 percent overestimates.

If the second source, Y, is not 100 percent accurate as well, then the situation becomes more complex. The estimate of the completeness of source Z, $Z_c = a/Y_{obs}$ is falsely low, and a separate correction is needed here as well that may be subject to considerable uncertainty. The investigator must decide if accuracy of reporting by any two sources is itself independent. This is an issue separate from source dependency. Two sources may be negatively dependent, for instance, because of referral patterns, but may both use the same method of diagnosis, one of which is not completely accurate. In this case, there would be a positive dependence of misdiagnosis, so to speak, despite the negative dependence of reporting.

Analogous problems arise if cases are included falsely that are outside the time or space interval under analysis or if records are mismatched or unmatched inappropriately, a phenomenon that is sensitive to name changes. A related issue is the nature of the information used to make the diagnosis. Ideally, the investigator should independently evaluate all cases ascertained and confirm the diagnoses, done, for example, by Folstein (80) for the data set on Huntington disease. But usually this is impractical. The investigator may choose, rather, to accept at face value diagnoses reported in the sources (and run the risk of diagnostic errors) or, preferably, use certain key clinical features in the patient's medical record to reach a diagnosis without necessarily accepting the clinician's own diagnosis.

The last approach was done, for instance, in studies of amyotrophic lateral sclerosis (6, 81) and fetal alcohol syndrome (16). An interesting, related problem arose in the latter study. The criteria for verification of the diagnosis of fetal alcohol syndrome required the notation of five separate findings (but not necessarily the diagnosis itself) in the patient's records. For two out of 50 total observed cases, the diagnosis could only be regarded as confirmed by pooling information in both sources that were used. That is, inclusion as a confirmed case was conditional on the case being noted in both sources. Had either of these cases been reported in only one of the sources, it would have been excluded as a chart-verified case (although classified as a potential case). Probably these cases should not have been included in the final analysis. A related issue applies to prevalence estimates of affected individuals in what may be termed an "open" population, considered below.

In the absence of identifying information on those with sensitive diagnoses such as drug addiction or human immunodeficiency virus (HIV) positivity, investigators can use only surrogate variables to make putative matches (e.g., Frischer et al. (36)). This introduces additional uncertainty as one can have both false positive and false negative matches. While the investigator may, through careful checking, attempt to minimize the error, it is not clear to us that one may safely conclude that any such scheme results only in small errors on capture-recapture estimates. An alternative approach is to make matches using both rela-

Downloaded from academic.oup.com/epirev/article/17/2/243/499931 by Michigan State University user on 12 October 2026

tively strict criteria and relatively loose criteria, and to explore the consequent variation in matching on the overall estimate. The problem of matching error has also arisen with corrections to human census estimates. This issue is discussed extensively by Hogan and Wolter (105) and is beyond the scope of this review.

Lastly, we note that if exploratory capture-recapture analysis (especially that of two sources) reveals estimates and suggests dependencies which the investigator believes are implausible, this may well be an indicator of errors in the data set. We are aware of two instances in which incomplete matching was initially suspected because of such errors and subsequently confirmed.

Variable catchability

Members of a population may differ considerably in the probability of ascertainment by any particular source because of geographic or socioeconomic variables if no other variables. Analogous with approaches in animal ecology (3), this heterogeneity in ascertainment by a source may be termed "variable catchability" (106) or perhaps "variable sightability." The problem this heterogeneity introduces into the analysis is most easily illustrated in the two-source case.

Even if two sources ascertain individual members or subgroups of the population independently, these sources may well not be independent in the entire population. Table 4, from Hook and Regal (106), illustrates this point. It assumes a hypothetical population of Down syndrome cases in which 20 percent occur in stratum 1 and 80 percent in stratum 2. We assume that within each stratum, the two sources, cytogenetic records and birth certificate reports, are independent. With regard to the first source, birth certificates, the probability that a case that occurs in the minority, subgroup 1, will be reported is 15 percent. For a case in the majority, subgroup 2, this probability is 40 percent. We consider three alternatives with regard to the second source, cytogenetic reports: that the subgroup 1 probability of appearing in this source is less than, greater than, or equal to the subgroup 2 probability of being in this source (30 percent vs. 60 percent; 60 percent vs. 30 percent; and 20 percent in both, respectively). Then the capture-recapture maximum likelihood estimates of the entire population are, respectively, an underestimate, overestimate, or correct estimate. Yet, in both subgroups, recall that the sources are independent, and note that the maximum likelihood capture estimates of the entire population are, respectively, an underestimate,

TABLE 4. Variation in capture-recapture estimates of total affected illustrated by substrata variation in probability of source capture*

		Subgroup 1, source B†			Subgroup 2, source B			Pooled strata, source B		
		Yes	No	Total	Yes	No	Total	Yes	No	Total
		Probability of cytogenetic report in subgroup 2 > that in subgroup 1‡								
	Yes	9	51	60	192	288	480	201	339	540
Source C†	No	21	119	140	128	192	320	149	311	460
	Total	30	170	200	320	480	800	350	650	1,000
		n_{MLE}† = 200 = n_1†			n_{MLE} = 800 = n_2†			n_{MLE} = 940 < n† = 1,000		
		Probability of cytogenetic report in subgroup 2 < that in subgroup 1§								
	Yes	18	102	120	96	144	240	114	246	360
Source C	No	12	68	80	224	336	560	236	404	640
	Total	30	170	200	320	480	800	350	650	1,000
		n_{MLE} = 200 = n_1			n_{MLE} = 800 = n_2			n_{MLE} = 1,105 > n = 1,000		
		Probability of cytogenetic report in subgroup 2 = that in subgroup 1‖								
	Yes	6	34	40	64	96	160	70	130	200
Source C	No	24	136	160	256	384	640	280	520	800
	Total	30	170	200	320	580	800	350	650	1,000
		n_{MLE} = 200 = n_1			n_{MLE} = 800 = n_2			n_{MLE} = 1,000 = n		

* Reproduced with permission from Hook and Regal, *Am J Epidemiol* 1993;137:1148–66 (108).

† B, birth certificates; C, cytogenetic reports; n_{MLE}, maximum likelihood estimate; n_1, total in subgroup 1; n_2, total in subgroup 2; n, total in the population.

‡ The probability of capture by B is 0.15 in subgroup 1 and 0.4 in subgroup 2; by C it is 0.3 in subgroup 1 and 0.6 in subgroup 2.

§ The probability of capture by B is 0.15 in subgroup 1 and 0.4 in subgroup 2; by C it is 0.6 in subgroup 1 and 0.3 in subgroup 2.

‖ The probability of capture by B is 0.15 in subgroup 1 and 0.4 in subgroup 2; by C it is 0.20 in both subgroups.

Downloaded from academic.oup.com/epirev/article/17/2/243/499931 by Michigan State University user on 12 October 2026

overestimate, or correct estimate. Recall that in both subgroups the sources are independent, and note that the maximum likelihood capture-recapture estimates of either subgroup is correct in all three analyses, despite the fact that the estimates of the entire group is erroneous in two instances. This illustrates that even when the investigator knows sources are acting independently at least in subgroups, he or she may be misled because substrata differences in catchability may result in a deviation from independence in the entire group.

As an aside, we note one further trend of interest illustrated by these data. For the first data set in table 4, ascertainment by source B is lower than that by source C in both subgroups. For the second data set, ascertainment by source B is greater in one subgroup and less in the other. For the third data set, ascertainment by the sources is the same in one subgroup although differing in another. In the first case, the maximum likelihood estimate is less than the true total; in the second it is greater; and in the third it equals the true value. This illustrates a more general result, namely that if catchability varies in the same direction in both sources, one will derive an underestimate of the true total, if in opposite directions, an overestimate, and if ascertainment is identical in at least one subgroup, an unbiased estimate (see table 3 of Hook and Regal (106), but note a corrigendum to this (107)).

We may express the deviation from the true value of any two-source capture-recapture estimate of a population in terms of two components. One component, denoted as V, represents the contribution of variable catchability by sources of the different substrata. The other component, denoted by D, represents the contribution of source deviation from independence in the subgroups. We will define these more precisely below, but note here that the latter contribution is of course zero in the above example, and all of the deviation considered in table 4 arises from V, variable catchability.

We now consider the formal definitions of these two components in terms of the variables defined below. All of these values are illustrated for the specific data in table 5. Suppose there are not two substrata of the entire population as above but, rather, k subgroups and p_i $(i = 1, ..., k)$ represents the proportion of the total in each, i.e., the "catchability" of cases by each subgroup. Assume two sources B and C, and that b_i, c_i, and $(bc)_i$ denote, respectively, the probabilities of ascertainment by source B, C, or both B and C jointly, of cases in source i. Let Δb_{ij} and Δc_{ij} denote, respectively, $(b_i - b_j)$ and $(c_i - c_j)$ for any two subgroups i and j, that is, the differences in "catchability" by the sources of the two subgroups. Let $r_i = (bc)_i/(b_i)(c_i)$ and $\Delta r_i = (r_i - 1)$.

Define

$$K = \sum p_i b_i c_i \tag{5}$$

$$\Delta = \sum p_i p_j \Delta b_{ij} \Delta c_{ij}, \tag{6}$$

for all combinations of subgroups i and j

$$V = \Delta/K \tag{7}$$

TABLE 5. An example of bias in capture-recapture estimation resulting both from source dependencies in subgroups and variable catchability of cases by each source*

		Subgroup 1, source B			Subgroup 2, source B			Pooled strata, source B		
		Yes	No	Total	Yes	No	Total	Yes	No	Total
Source C	Yes	15	45	60	150	330	480	165	375	540
	No	15	125	140	170	150	320	350	275	460
	Total	30	170	200	320	480	800	350	650	1,000
		$n_{MLE} = 120 < n_1 = 200$			$n_{MLE} = 1,024 > n_2 = 800$			$n_{MLE} = 1,145 > n = 1,000$		
		$p_1 = 0.2$ $b_1 = 0.15$ $c_1 = 0.30$			$p_2 = 0.8$ $b_2 = 0.40$ $c_2 = 0.60$					
								$K = 0.2010$		
		$r_1 = 15/9 = 1.6667$ $\Delta r_1 = 0.6667$ $\nabla_1 = 0.0060$			$r_2 = 150/192 = 0.7813$ $\Delta r_2 = -0.2187$ $\nabla_2 = -0.0420$					
								$\Delta = 0.0120$ $\nabla = -0.0360$ $B_r = 1.1455$ $\Delta' = -0.0292$ $\nabla' = -0.0255$	$V = 0.0597$ $D = -0.1194$ $V' = -0.1455$ $D' = -0.1270$	

* See text and table 4 for definitions of symbols and terms.

Downloaded from academic.oup.com/epirev/article/17/2/243/499931 by Michigan State University user on 12 October 2026

$$\nabla i = p_i b_i c_i (\Delta r_i) \quad (8)$$

$$\nabla = \sum \nabla_i \quad (9)$$

$$D = \nabla/K. \quad (10)$$

Then, as slightly modified from results in Hook and Regal (106), the relative bias (B_r) of a two-source capture-recapture population estimate may be modeled by a single expression in terms of variable catchability among subgroups of the population, V, and deviation from independence of the sources within these subgroups D, as

$$B_r = (K - \Delta)/(K + \nabla) = (1 - (\Delta/K))/(1 + (\nabla/K)) = (1 - V)/(1 + D). \quad (11)$$

This expression is denoted as "relative accuracy" in Hook and Regal (106), but it is more appropriately regarded as a measure of relative bias as here. Note that we have defined V, as D, so that positive values result in a bias toward an underestimate, negative values to a bias to overestimate. The direction of the overall bias B_r, of course depends on the relative magnitudes of the two components.

Moreover, if B_r is greater than 1.0, then the capture-recapture estimate of the entire population is biased high and we may say there is "net negative dependency" among the sources in the entire population whatever the dependencies in the subgroups and the value of D and ∇. Similarly, if $B_r = 1.0$, we have "net independence," and if $B_r < 1.0$, we have "net positive dependence." We may express this "net" or "overall" dependency in terms of a variable D', analogous to D, or another variable ∇', analogous to ∇, defined by the equations

$$D' = \nabla' / K = (1 / B_r) - 1 \quad (12)$$

so that

$$B_r = K / (K + \nabla') = 1 / (1 + D'). \quad (13)$$

That is, the (relative) bias of the capture-recapture estimate of the entire population may be expressed as if it were derived *only* from (possibly non-independent) sources within which there was no variable catchability, and for which the overall or net dependence is D'.

Similarly, we could express the relative bias *as if it derived only* from variable catchability within sources which were all independent. We define this "net" or overall variable catchability as V', and it and a related variable Δ', are given by the equations:

$$V' = \Delta' / K = 1 - B_r \quad (14)$$

so that

$$B_r = (K - \Delta) / K = (1 - V'). \quad (15)$$

So the (relative) bias may be expressed also as if it were derived *only* from subgroups for which the cases had variable catchability but in each of which the sources were independent. While the concepts of "net" or "overall" dependence D' and "net" or "overall" catchability V' have been derived within the context of two-source analyses, their definitions in terms of the relative bias B_r may of course be extended readily to applications involving any number of sources.

The analysis of the hypothetical data set in table 5 illustrates these matters quantitatively. The contribution of variable catchability to the bias in terms of the formulation above is $V = +0.0597$, and of source dependency is $D = -0.1194$. But this could also be expressed algebraically, as a contribution entirely of deviations from source independence in subsets summarized as $D' = -0.1270$, or entirely due to variable catchability of sources, $V' = -0.1455$. In this sense, variable catchability and nonindependence of sources are intertwined concepts. In fact, it appears likely that in most circumstances variable catchability is the "prior" concept so to speak, i.e., that most (and perhaps all) observed variations from independence have their ultimate origin biologically simply in variable catchability among numerous very small subgroups of the population. Of course, identical twins aside, all affected individuals are likely to have at least slightly different probabilities of ascertainment by any source. Thus, expression of the bias solely in terms of variable catchability by subgroups would imply (twins aside), at the extreme, as many subgroups as there are individuals in the population (i.e., $k = N$) in each of which there is a single member. Clearly, it would be practically impossible to dissect out the sources of bias by such an analysis. For this reason, the investigator will often find it practical to conceptualize and characterize the bias, whatever its origin, as attributable in whole, or at least in part, to source dependence in larger subgroups or in the entire population.

In multiple-source capture-recapture analysis, investigators using log-linear methods make the same assumption of constant catchability. If the assumptions are violated, then not only may different models apply to different subsets of the population, but they may differ from the model, however simple, that appears to describe the whole population on the assumption of constant catchability, let alone the model that does in fact hold. Methods have been proposed to model the consequences of variable catchability for log-linear capture-recapture estimates derived from multiple sources in human census (108) or animal population

Downloaded from academic.oup.com/epirev/article/17/2/243/499931 by Michigan State University user on 12 October 2026

studies (109). Such estimates still make assumptions about multisource independence of ascertainment of individuals (or subgroups). While they may well be appropriate for most census and animal ecology applications, we believe they are much less likely to apply generally in epidemiology. Sources used typically in epidemiology are lists of cases already collected for other administrative purposes, and may well have some intrinsic relation. They are not typically the results of different random and/or attempted complete surveys of the population, where multiple "source" independence may be much more likely to hold, at least for individuals.

In any event, because of the intrinsic nature of human variation, many populations, no matter how small, have variable catchability by the sources usually available to the epidemiologist. This relates not just to geographic and socioeconomic divarication but to variation in severity of disease. The assumption of constant catchability is, we believe, rarely if ever satisfied.

The main assumption that underlies the usual application of log-linear methods for capture-recapture analysis to k sources is that any variable catchability and source dependencies that are present in the population analyzed result in no more than a net $k - 1$ interaction. A secondary assumption of course is that he or she can model that interaction with the observed data.

An investigator can hardly ever prove these assumptions to be correct. About the best he or she can do is attempt multiple different evaluations of the data to determine if the estimates implied by various approaches are consistent or inconsistent with each other and with prior knowledge or expectation of population size and source interaction.

One strategy that at least enhances the plausibility of this assumption is the use of as many sources and as many qualitatively different types of sources as possible. As the number of sources increases, the number of possible models that can be fit to the data increases markedly; the single k-way interaction model becomes a smaller and smaller fraction of the total possible. For $k = 3$ sources, the k-way interaction model is one of nine possible, for $k = 4$, the k-way interaction model is one of 114 possible, and for $k = 5$ sources, the five-way interaction model is one of 6,894 possible. This illustrates one of the major advantages to analyzing data from four or five sources say, rather than, for example, collapsing them to three before any analysis is begun. Another advantage is considered below in the discussion of analysis of validity of estimates.

The use of qualitatively different types (or regions of geographic coverage) of sources provides in principle a different advantage by diminishing the likelihood that any significant subset of the affected population is virtually uncatchable by any source used.

The penalty paid for analyzing a large number of sources rather than collapsing to no more than a few, is the additional complexity of analysis and the variable, inconsistent, or unstable estimates that may result. Moreover, no matter how many sources and how many qualitatively different types available, the investigator nevertheless must always consider what subsets of the population might be particularly likely to be missed by any of them, and whether it is worthwhile to limit inferences accordingly, say to the most severely affected.

Egeland et al. (16), for example, excluded infants under age 3 years from capture-recapture prevalence estimates of chart verifiable cases of fetal alcohol syndrome because of concern that full physical manifestations of the condition were not fully developed until after that age. As this implies, if manifestations of the disorder vary considerably, the investigator may well find it worthwhile to make estimates separately for distinctive types of manifestations. As Bonaiti and Demenais (110) note for retinoblastoma, which may affect one or both eyes, it would be useful to attempt separate estimates for unilateral or bilateral affliction because of their great differences in ascertainment. Unilaterally affected cases, for instance, would be unlikely to be found on diagnostic lists of institutions for the blind.

In any event, if data on covariates are available to the investigator, such as race, age, geographic area, etc., he or she may well find it worthwhile to attempt independent estimates on some of the subgroups defined by these variables, not only because of their intrinsic interest, but also to determine directly if the sum of the subgroup estimates is equal to the direct population estimate. If these are grossly discrepant, the estimate derived from the sum of the subgroups is preferable, and the possibility of a net k-way interaction in the entire group should be considered.

Another related approach may be used if the investigator has data on the values of relevant covariates for each case. This is to derive an estimate using a log-linear regression analysis simultaneously adjusting for covariates. This approach has been used infrequently in epidemiologic application; perhaps the best examples are those of Ahlo (18) and Walter et al. (15) (see also Robles et al. (14) and Hilsenbeck et al. (41)—the latter derived standardized rates of malignancy in an entire population from capture-recapture estimates of rates in strata). The relative ease of application of this method with statistical packages for log-linear methods such as Generalised Linear Interactive Modelling

Downloaded from academic.oup.com/epirev/article/17/2/243/499931 by Michigan State University user on 12 October 2026

(GLIM) (111), cited by Walter et al. (15), makes it very attractive. The investigator, however, may be led to overlook some of the intrinsic limitations or subtleties in application of the method and accept the estimate and stated confidence interval at face value without any exploratory analysis. Moreover, some anomalous outcomes in the report of Walter et al. (15) suggest that caution should be used in application of such packaged programs that adjust for covariates. These workers report estimates of completeness of reporting to their registry of over 100 percent for several malignancies and imply that their capture-recapture analysis derived negative estimates of missing cases, an impossible outcome. There is clearly a problem with the particular method used by the authors, or with its application to the particular data set.

Open populations

If individuals can migrate into an area and/or be lost because of death or out-migration in the space-time interval under study, we may term the population "open." Otherwise, it is "closed." Those who migrate in or out clearly have lower "catchability" by any source than the remainder. Those who die may have higher or lower catchability than other individuals depending on the nature of the source. (Catchability might be higher for autopsy records but lower for outpatient clinic records, for instance.) As an open population virtually by definition is one with variable catchability, it does not meet the theoretical assumption of constant catchability underlying the application of log-linear models to population estimates. For this reason, many discussions in the literature of log-linear applications to estimate population size describe them as methods for estimating closed populations (e.g., York and Madigan (74)), because only these can *in principle* meet the underlying assumption. As implied by the discussion above, however, virtually all closed populations available to epidemiologists, at least, violate the assumption of constant catchability *in any case*. Thus, the difference between closed and open populations from this perspective is simply one of degree, i.e., the extent to which the assumption is likely to be violated. Certainly, other things being equal, one has fewer grounds for concern about an estimate of a closed population because there are fewer possible components of variable catchability.

Some open populations may be closed to "in-migration" but not to "out-migration" (or death) or the reverse. We may call these "semiclosed." Other things being equal, the extent of variable catchability is intermediate between "open" and closed populations. Probably the vast bulk of populations studied in epidemiology with capture-recapture methods are at best semiclosed, because of at least some out-migration during the interval of ascertainment as well as death of cases. (One might limit inference to cases known to be in a jurisdiction at the start of some interval and still alive and within it at the end, but rarely do epidemiologists have such complete information. An exception first analyzed by Wittes (112) is presented by Fienberg (76) and Bishop et al. (21).) One may introduce directly methods for modeling the in- and out-migration of a population, although such methods are usually restricted by the need for additional ad hoc data on migration or some sequential temporal ordering of the sources. Cormack (113), for instance, has proposed methods in animal ecology for extending log-linear approaches to open populations with the use of data from sequential trappings of animals. This method could only be used in epidemiologic application if the lists used had some natural temporal ordering, as do trappings. Such an application, for instance, might use data from a single source at several different time intervals.

Brecht and Wickens (114) generated simulations of mobile populations of drug addicts and compared estimates derived from three different methods: 1) the log-linear approach for a closed population, as discussed above (although they did not explicitly consider methods of model selection); 2) an approach developed for study of open animal populations (115, 116); and 3) an approach of their own that employs latent-state Markov models. Brecht and Wickens (114) interpreted their results to indicate that all but the Jolly-Seber method produced "reasonable" estimates close to the hypothetical true values used in the simulations, and that the best result was with their Markov model. (The Jolly-Seber method makes assumptions likely to be more applicable to successive trappings of wildlife, i.e., multiple reuses of one source, than to multiple different ascertainment sources of human disorder, and so its failure is understandable here.) The Markov model that Brecht and Wickens found to be optimal for open populations necessitates use of external information concerning the ratios of "exit probabilities" from various states, information which the epidemiologist is unlikely to have but which was assumed available in the simulations.

Consequences of assumption violations for confidence intervals

The assumption of constant catchability that underlies application of log-linear capture-recapture methods to human populations almost certainly does not hold. For this reason alone, if for no other reason, we must also remain skeptical about any confidence interval, however it may be derived. As long as the

Downloaded from academic.oup.com/epirev/article/17/2/243/499931 by Michigan State University user on 12 October 2026

investigator appreciates this limitation, an appropriately calculated confidence interval about any estimate may be of some use. But we suspect strongly that a confidence interval misleads many individuals about the extent of uncertainty of the result. As illustrated below, it may be more useful if the investigator derives a number of different plausible estimates using different methods and presents simply their "plausible range" in the derived estimates, instead or in addition to presenting some particular interval at some putative level of confidence.

METHODS OF VALIDATION AND THEIR APPLICATION TO REFINEMENT OF ESTIMATES

Increasing application of capture-recapture methods in epidemiology has inevitably generated skepticism because of their dependence, as discussed above, on unproven or unprovable assumptions (117–120). One study of particular interest is that by Carothers (121), who, knowing the total number of registered taxicabs in Edinburgh, Scotland, "captured" them by noting their identification numbers on successive occasions in different locations. His study was primarily designed to determine the reliability of methods to test for the absence of variable catchability in animal populations. He concluded that not only were they not reliable, but that even for taxicabs there is "variable catchability." (Persons in midtown Manhattan who have tried to hail a cab on successive rainy and sunny days are already well aware of this, although we are referring here to a different type of "catchability.") Carothers concluded that virtually all natural populations have variable catchability (see also Cone et al. (122)). We find it of additional interest that despite knowledge of the total number of registered taxicabs, Carothers still had to make a rough subjective *estimate* as to the (almost certainly small) number of taxicabs that were completely out of service in the interval analyzed (i.e., had essentially zero catchability), although his final conclusions were reached separate from this additional complication.

As noted above, Brecht and Wickens (114) found that with the use of simulations using data from five sources they could reasonably estimate the size of a hypothetical open population by applying log-linear capture-recapture methods that were developed in principle for closed populations. Despite the reassurance of this announced result, simulations are somewhat unsatisfying for verification of the validity of any approach because one never knows how well the simulated conditions come to those of "real" populations.

We are aware of two validity studies of capture-recapture estimates in epidemiology that have used data from "real populations." Barg and Huether (19) compared the number of Down syndrome cases generated by a two-source estimate from reports on birth certificates and in diagnostic cytogenetic laboratory records with the total number of cases found in an exhaustive study of all known sources. The estimate was reasonably close to the (presumed) true value. The only multiple-source capture-recapture validity study in epidemiology known to us that has evaluated data from a real population, as opposed to a simulated population, compared estimates of various overlapping *sub*populations of a single population with their known totals (43). While the size of these subpopulations were all known, they were part of one particular larger population of unknown size. For this reason, we call them here studies of "relative" or "internal validity," because, even though they could validate capture-recapture methods applied to all the subsamples investigated, they could not prove that any particular capture-recapture estimate of the *entire* population was valid. The subpopulations analyzed were in fact the total number observed in each of four ascertainment sources. (Information on cases in the intersection of any source with those in all other sources was used to estimate the number unique to the source, an estimate then compared with the observed number. The results of such an approach are most useful when a large number of data from four or more sources are available.) This analysis demonstrated that, at least for the subpopulations analyzed, the capture-recapture analysis did provide in each case reasonable estimates of the total subpopulation.

This approach to analysis of "internal validity" may also be used to adjust estimates of the entire population when there are data from at least four sources (125). In essence, the method examines how well-known subsets of the total are predicted by models that are subsets of the log-linear model chosen to predict the entire population, and it adjusts derived estimates of the total accordingly.

As an illustration, we consider four sources, denoted as A, B, C, and D, and we assume that the log-linear model selected by the investigator for estimation postulates that there is an interaction of sources C and D, i.e., that A, B, and the union of C and D are independent. The method first undertakes selected multiple restricted estimates of the entire population. By assuming that A, B, and C are independent, one derives an initial estimate of the total, which we will denote as $\hat{N}_{ABC}$. One then undertakes an internal validity analysis by comparing the analogous estimate of D, derived by assuming that A, B, and C are independent, $\hat{N}_{D/ABC}$, with the known total in D, N_D. Their ratio is an estimate of the relative bias of this three-source estimate. One then adjusts $\hat{N}_{ABC}$, the estimate of the total

Downloaded from academic.oup.com/epirev/article/17/2/243/499931 by Michigan State University user on 12 October 2026

for this bias by dividing this value by the estimated relative bias, i.e., one derives a final estimate by this particular approach as

$$\hat{N} = (\hat{N}_{ABC})/(\hat{N}_{D/ABC}/N_D)$$

$$= (\hat{N}_{ABC})(N_D)/\hat{N}_{D/ABC} \ . \qquad (16)$$

The step is then repeated analogously for other "submodels" of the optimal model chosen. In the case of this example, in which this is A vs. B vs. (C union D), these are: A vs. B vs. D, A vs. (C union D), and B vs. (C union D). The associated estimates of the totals from these latter approaches would be respectively,

$$\hat{N} = (\hat{N}_{ABD})\ (N_C)/\hat{N}_{C/ABD}\ , \qquad (17)$$

$$\hat{N} = (\hat{N}_{A(CD)})(N_B)/\hat{N}_{B/A(CD)}\ , \qquad (18)$$

$$\hat{N} = (\hat{N}_{B(CD)})(N_A)/\hat{N}_{A/B(CD)}. \qquad (19)$$

No single such estimate in itself may be regarded as optimal, but the median of all of them provides an alternative capture-recapture estimate for the entire population.

We recently applied this approach to data from a four-source analysis of intravenous drug users in Glasgow, Scotland, reported by Frischer et al. (36). There were four sources: police records (POL), a needle and syringe exchange registry (NES), an HIV registry (HIV), and a records of drug treatment center (TRE). The conventional log-linear capture-recapture analysis estimated 8,494 cases from the model with minimum AIC, and 8,319 from the weighted BIC. The optimal model by AIC was POL vs. (HIV union NES union TRE), i.e., the model that postulates police records are independent of the other three sources pooled. All three of the appropriate internal validity analyses derived estimates greater than the value predicted by this model: about 8,700, 9,900, and 10,200 (123). Clearly, one cannot state with certainty which estimates are closest to the true values. However, we believe that the internal validity analyses provide strong evidence that there is a positive dependency of the police records with the other sources pooled together (a relation we believe to be plausible because of the nature of the sources), so that the true number of cases is likely to be greater than those of the original analysis.

The results of any particular internal validity analysis may not always be so consistent and may result in a much greater range of estimates, and one which overlaps that produced by the more traditional approach to estimation. Moreover, there may not be a natural "submodel" for generation of analogous internal validity analyses. But we suggest that whenever data from four or more sources are available, the investigator attempt such an internal validity analysis (if such is possible) to evaluate the plausibility of the estimate reached on other grounds.

The internal validity analysis illustrates some of the attraction of two-source methods. However limited they may seem, they have the benefit not only of ease in derivation of estimates but also of conceptual simplicity of interpretation. We often find it very difficult to interpret what multiple-source interactions in a population imply about the structure of a population. Yet, we usually find it easy to comprehend and often predict the likely nature of dependencies between any two sources, or two combinations of sources. In our view, the critical evaluation of the multiple results from many different two-source analyses may well outweigh the apparent advantage of a single more complex analysis.

ALTERNATIVES TO CAPTURE-RECAPTURE METHODS

We have already discussed the ratio estimator. With multiple sources, one alternative approach to capture-recapture methods is to use a truncated distribution. That is, one summarizes the observations according to the number of cases captured by just one source, by two sources, by three sources, etc., and then derives the number of cases in no source predicted by the observed distribution, on the assumption that it is a truncated realization of some theoretical distribution, e.g., a binomial or Poisson whose parameters are derivable from the observations. This may well be the oldest method used for undertaking ascertainment adjustment of human disorders (124–128). These approaches have been widely used by genetic epidemiologists, who must adjust for ascertainment in undertaking segregation analysis, i.e., in analyzing the probability that a family "at risk" will have an affected child (127–133). They have been used in epidemiologic application independent of human genetic analysis (e.g., by Morton et al. (134) and Zelterman (135)). The use of truncated distributions has been applied in a number of other areas, for example, to estimate the number of as yet undiscovered species of some type or the well-known attempt to estimate the total number of words that Shakespeare knew but did not use (136).

A truncated binomial or truncated Poisson is relatively easy to apply but assumes not only constant catchability but equal ascertainment probabilities and that all sources are independent, assumptions that are even more implausible in most epidemiologic analyses than those of log-linear capture-recapture analysis. The Skellam distribution (137), also known as a beta-binomial distribution, has the advantage of allowing at least for variable catchability, although not for deviation from independence (over and beyond that im-

Downloaded from academic.oup.com/epirev/article/17/2/243/499931 by Michigan State University user on 12 October 2026

posed by variable catchability). Morton et al. (134) applied such an analysis to estimation of the prevalence of congenital malformations in Hawaii. The internal validity analysis of Hook and Regal (43) compared the accuracy of log-linear, truncated binomial, and truncated beta-binomial estimates derived from four sources in one data set. While the log-linear estimates were optimal, the estimates derived by means of the other two methods were surprisingly good, with average mean error of about 10 percent and with no more than 6 percent greater mean (absolute) error than that of the log-linear estimates for any of the four sources analyzed. The confidence intervals associated with the truncated distribution methods, however, were notably larger than those associated with the log-linear methods.

Yet another alternative is simply to make no adjustment. For purposes of surveillance of prevalence rates in any jurisdiction, this may be adequate. The investigator in such circumstances is less interested in the actual numbers of those affected than whether the proportion of those at risk is increasing or decreasing. However, eventually he or she is likely to confront the question of whether some apparent cluster or other trend to increase resulted from better ascertainment or whether diminished ascertainment has led to some missed episode. Certainly, if pertinent data are available, there appears to be no reason not to apply capture-recapture methods for surveillance, as has been strongly recommended by LaPorte (138), but with due caution in interpretation of trends.

The option of making no adjustment may appear to be attractive when an apparently exhaustive survey has been carried out. But this requires a different assumption, namely that all possible cases have been found. Capture-recapture methods enable at least some plausible test of this assumption, whether or not final population adjustments are made (44). Similar issues have arisen in the evaluation of results from the US census (139, 140).

CONCLUSIONS AND RECOMMENDATIONS

Any investigator who collects or uses data from multiple data sources for prevalence or incidence estimates should attempt to keep track of the intersection of ascertainment sources, and they should note such intersections in any publication, irrespective of utilization of capture-recapture estimates to refine estimates. Similarly, registries should keep track of all ascertainments of cases by source of report.

Once the data set is in hand and the investigator has attempted to ensure the accuracy of the reported data to the extent possible, we suggest that he or she undertake an "exploratory" analysis by reviewing critically the range in predicted results with multiple different two-source analyses (assuming more than two sources are available) in both the entire population as well as in available strata. If the results are not congruent with the investigator's expectations and prior knowledge about the nature of the sources, we suggest he or she consider the possibility that there are still residual inaccuracies in the data. The advantage of conceptual ease of interpretation of the results of two-source analyses may well reward the rather simple expenditure of effort on them.

As implied by our discussion above, we suggest consideration of many different approaches to application of multiple-source log-linear analyses both in the total population and in subgroups, and that the investigator search for some internal consistency of results which cohere with his or her knowledge of or reasonable expectation of the population. Lastly, we also urge that the investigator consider what types of cases are intrinsically unlikely to be caught by any of the sources used, and that he or she limit inferences about the population estimated accordingly.

ACKNOWLEDGMENTS

Preparation of this paper was supported in part by cooperative agreement no. U83/CCU908 694–02 with the National Center for Health Statistics.

We thank Johanna Weber for extensive assistance, William J. Redfearn for cited communications, an anonymous reviewer for a number of suggestions concerning organization and clarification of the text, and Richard Cormack for many comments. We emphasize that these remarks do not imply that anyone we have acknowledged necessarily agrees with all of the viewpoints expressed.

REFERENCES

1. Petersen CGJ. The yearly immigration of young plaice into the Limfjord from the German sea. Rep Dan Biol Stat 1896; 6:1–48.
2. LaPlace PS. Sur les naissances, les mariages, et les morts. (In French). In: Historie de l'Academie Royale des Sciences, Année, 1783:693. Cited by: Seber GAF. The estimation of animal abundance and related parameters. 2nd ed. London, England: Charles Griffin, 1982.
3. Seber GAF. The estimation of animal abundance and related parameters. 2nd ed. London, England: Charles Griffin, 1982.
4. Chandrasekar C, Deming WE. On a method of estimating birth and death rates and the extent of registration. J Am Stat Assoc 1949;44:101–15.
5. El-Khorazaty M, Imreg PB, Koch GG, et al. Estimating the total number of events with data from multiple-record systems: a review of methodological strategies. Int Stat Rev 1977;45:129–57.
6. McCarty DJ, Tull ES, Moy CS, et al. Ascertainment corrected rates: applications of capture-recapture methods. Int J Epidemiol 1993;22:559–65.
7. Wittes JT. Applications of a multinomial capture-recapture

Downloaded from academic.oup.com/epirev/article/17/2/243/499931 by Michigan State University user on 12 October 2026

model to epidemiological data. J Am Stat 1974;69:93–7.
8. Hook EB, Chambers GM. Estimated rates of Down's syndrome in livebirths by one year maternal age intervals for mothers aged 20 to 49 in a New York State study—implications of the "risk" figures for genetic counseling and cost benefit analysis of prenatal diagnosis programs. In: Bergsma D, Lowry RB, Trimble BK, et al., eds. Numerical taxonomy of birth defects and polygenic disorders. (Birth defects original article series, vol 13, no. 3A.) New York, NY: Alan R Liss, 1977:123–41.
9. Hook EB, Cross PK. Interpretation of recent data pertinent to genetic counseling for Down's syndrome: maternal-age specific rates, temporal trends, adjustments for paternal age, recurrence risks, risks after other cytogenetic abnormalities, recurrence risk after remarriage. In: Willey AM, Carter TP, Kelly SM, et al., eds. Clinical genetics: problems in diagnoses and counseling. New York, NY: Academic Press, 1982:119–45.
10. Lindsten J, Marsk L, Berglund K, et al. Incidence of Down's syndrome in Sweden during the years 1969–1977. In: Burgio GR, Fraccaro M, Tiepolo L, et al., eds. Trisomy 21: an international symposium, Convento delle Clarisse, Rapallo, Italy, November 8–10, 1979. Berlin, Germany: Springer-Verlag, 1981:195–210.
11. Cuckle HS, Wald NJ, Thompson SG. Estimating a woman's risk of having a pregnancy associated with Down's syndrome using her age and serum alpha-fetoprotein level. Br J Obstet Gynaecol 1987;94:387–402.
12. Hook EB. Chromosome abnormalities: prevalence, risks and recurrence. In: Brock DJH, Rodeck CH, Ferguson-Smith MA, eds. Prenatal diagnosis and screening. Edinburgh, Scotland: Churchill-Livingstone, 1992:351–92.
13. 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.
14. Robles SC, Marrett LD, Clarke EA, et al. An application of capture-recapture methods to the estimation of completeness of cancer registration. J Clin Epidemiol 1988;41:495–501.
15. Walter SD, Birnie SE, Marrett LD, et al. The geographic variation of cancer incidence in Ontario. Am J Public Health 1994;84:367–76.
16. Egeland GM, Perham-Hester KA, Hook EB. Use of capture-recapture analyses in fetal alcohol syndrome surveillance in Alaska. Am J Epidemiol 1995;141:335–41.
17. Abeni DD, Brancato G, Perucci CA. Capture-recapture to estimate the size of the population with human immunodeficiency virus type 1 infection. Epidemiology 1994;5: 410–14.
18. Alho JM. Logistic regression in capture-recapture models. Biometrics 1990;46:623–35.
19. Barg MS, Huether CA. A study of underreporting of Down's syndrome on birth certificates in an Ohio County, 1970–78. Public Health Rep 1983;98:78–84.
20. Benn RT, Leck I, Nwene UP. Estimation of completeness of cancer registration. Int J Epidemiol 1982;11:362–7.
21. Bishop YMM, Fienberg SE, Holland, PW. Discrete multivariate analysis: theory and practice. Cambridge, MA: MIT Press, 1975:229–56.
22. Bloor M, Leyland A, Barnard M, et al. Estimating hidden populations: a new method of calculating the prevalence of drug-injecting and non-injecting female street prostitution. Br J Addict 1991;86:1477–83.
23. Brecht ML, Wickens TD. Application of multiple-capture methods for estimating drug use prevalence. J Drug Issues 1993;23:229–50.
24. Bruno G, Bargero G, Vuolo A, et al. A population-based prevalence survey of known diabetes mellitus in northern Italy based upon multiple independent sources of ascertainment. Diabetologia 1992;35:851–6.
25. Chapman CJ. Ethnic differences in the incidence of cleft lip and/or cleft palate in Auckland, 1960–1976. N Z Med J 1983;96:327–9.
26. Chiu WT, Dearwater SR, McCarty DJ, et al. Establishment of accurate incidence rates for head and spinal cord injuries in developing countries: a capture-recapture approach. J Trauma 1993;35:206–11.
27. Cochi SL, Edmonds LE, Dyer K, et al. Congenital rubella syndrome in the United States, 1970–1985: on the verge of elimination. Am J Epidemiol 1989;129:349–61.
28. Davis SF, Strebel PM, Atkinson WL, et al. Reporting efficiency during a measles outbreak in New York City, 1991. Am J Public Health 1993;83:1011–15.
29. Doscher ML, Woodward JA. Estimating the size of subpopulations of heroin users: applications of log-linear methods to capture/recapture sampling. Int J Addict 1983;18:167–82.
30. Drucker E, Vermund SH. Estimating population prevalence of human immunodeficiency virus infection in urban areas with high rates of intravenous drug use: a model of the Bronx in 1988. Am J Epidemiol 1989;130:133–42.
31. Espeland MA, Murphy WC, Leverett DH. Assessing diagnostic reliability and estimating incidence rates associated with a strictly progressive disease: dental caries. Stat Med 1988;7:403–16.
32. 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.
33. Frischer M. Estimated prevalence of injecting drug use in Glasgow. Br J Addict 1992;87:235–43.
34. Frischer M, Bloor M, Finlay A, et al. A new method of estimating prevalence of injecting drug use in an urban population: results from a Scottish city. Int J Epidemiol 1991;20:997–1000.
35. Frischer M, Green ST, Goldberg DJ, et al. Estimates of HIV infection among injecting drug users in Glasgow, 1985–1990. AIDS 1992;6:1371–5.
36. Frischer M, Leyland A, Cormack R, et al. Estimating the population prevalence of injection drug use and infection with human immunodeficiency virus among injection drug users in Glasgow, Scotland. Am J Epidemiol 1993;138: 170–81.
37. Goldberg JD, Wittes JT. The estimation of false negatives in medical screening. Biometrics 1978;34:77–86.
38. Gutteridge W, Collin C. Capture-recapture techniques: quick and cheap. (Letter). BMJ 1994;308:531.
39. Hartnoll R, Lewis R, Mitcheson M, et al. Estimating the prevalence of opioid dependence. Lancet 1985;1:203–5.
40. Hewitt D, Milner J. Prevalence of hemophilia in Ontario, 1966. Can Med Assoc J 1970;102:174–7.
41. Hilsenbeck SG, Kurucz C, Duncan RC. Estimation of completeness and adjustment of age-specific and age-standardized incidence rates. Biometrics 1992;48:1249–62.
42. Hook EB, Albright SG, Cross PK. Use of Bernoulli census and log-linear methods for estimating the prevalence of spina bifida in livebirths and the completeness of vital record reports in New York State. Am J Epidemiol 1980;112: 750–8.
42a. Hook EB. Re: "Use of Bernoulli census and log-linear methods for estimating the prevalence of spina bifida in livebirths and the completeness of vital record reports in New York State." (Letter). Am J Epidemiol 1993;137:1285.
43. Hook EB, Regal RR. Validity of Bernoulli census, log-linear, and truncated binomial models for correcting for underestimates in prevalence studies. Am J Epidemiol 1982;116: 168–76.
44. Hook EB, Regal RR. The value of capture-recapture methods even for apparent exhaustive surveys: the need for adjustment for source and ascertainment intersection in attempted complete prevalence studies. Am J Epidemiol 1992;135: 1060–7.
45. Hser YI. Population estimates of intravenous drug users and HIV infection in Los Angeles County. Int J Addict 1993;28: 695–709.

Downloaded from academic.oup.com/epirev/article/17/2/243/499931 by Michigan State University user on 12 October 2026

46. Hser YI. Population estimation of illicit drug users in Los Angeles County. J Drug Issues 1993;23:323–34.
47. Huether CA, Gummere GR, Hook EB, et al. Down's syndrome: percentage reporting on birth certificates and single year maternal age risk rates for Ohio 1970–79: comparison with upstate New York data. Am J Public Health 1981; 71:1367–72.
48. Kehoe L, Hall W, Mant A. Estimates of the number of injecting drug users in a defined area. Aust J Public Health 1992;16:232–7.
49. Kozinetz CA, Skender ML, MacNaughton N, et al. Epidemiology of Rett syndrome: a population based registry. Pediatrics 1993;91:445–50.
50. LaPorte RE, Tull ES, McCarty D. Monitoring the incidence of myocardial infarctions: applications of capture-mark-recapture technology. Int J Epidemiol 1992;21:258–62.
51. LaPorte RE, McCarty D, Bruno G, et al. Counting diabetes in the next millennium: application of capture-recapture technology. Diabetes Care 1993;16:528–34.
52. Lewis CE, Hassanein KM. The relative effectiveness of different approaches to the surveillance of infections among hospitalized patients. Med Care 1969;7:379–84.
53. Lie RT, Heuch I, Irgens LM. Maximum likelihood estimation of the proportion of congenital malformations using double registration systems. Biometrics 1994;50:433–44.
54. Mastro TD, Kitayaporn D, Weniger BG, et al. Estimating the number of HIV-infected injection drug users in Bangkok: a capture-recapture method. Am J Public Health 1994;84: 1094–9.
55. McCarty DJ, Kwoh CK, LaPorte RE. The importance of incidence registries for connective tissue diseases. (Editorial). J Rheumatol 1992;19:1–7.
56. McKeganey N, Barnard M, Leyland A, et al. Female streetworking prostitution and HIV infection in Glasgow. BMJ 1992;305:801–4.
57. Modesitt SK, Hulman S, Fleming D. Evaluation of active versus passive AIDS surveillance in Oregon. Am J Public Health 1990;80:463–4.
58. Neugebauer R. Application of a capture-recapture method (the Bernoulli census) to historical epidemiology. Am J Epidemiol 1984;120:626–34.
59. Newmeyer JA. The prevalence of drug use in San Francisco in 1987. J Psychoactive Drugs 1988;20:185–9.
60. Palfrey JS, Haynie M, Porter S, et al. Prevalence of medical technology assistance among children in Massachusetts in 1987 and 1990. Public Health Rep 1994;109:226–33.
61. Prevots DR, Sutter RW, Strebel PM, et al. Completeness of reporting for paralytic poliomyelitis, United States, 1980 through 1991: implications for estimating the risk of vaccine-associated disease. Arch Pediatr Adolesc Med 1994;148: 479–85.
62. Rubin G, Umbach D, Shyu SF, et al. Using mark-recapture methodology to estimate the size of a population at risk for sexually transmitted diseases. Stat Med 1992;11:1533–49.
63. Sacks JJ, Sattin RW, Bonzo SE. Dog bite-related fatalities from 1979 through 1988. JAMA 1989;262:1489–92.
63a. Schouten LJ, Straatman H, Kiemeney LA, et al. The capture-recapture method for estimation of cancer registry completness: a useful tool? Int J Epidemiol 1994;23: 1111–16.
64. Serrano Rios M, Moy CS, Martin Serrano R, et al. Incidence of type 1 (insulin-dependent) diabetes mellitus in subjects 0–14 years of age in the Comunidad of Madrid, Spain. Diabetologia 1990;33:422–4.
65. Sutter RW, Cochi SL, Brink EW, et al. Assessment of vital statistics and surveillance data for monitoring tetanus mortality, United States, 1979–1984. Am J Epidemiol 1990;131: 132–42.
66. Sutter RW, Cochi SL. Pertussis hospitalizations and mortality in the United States, 1985–1988: evaluation of the completeness of national reporting. JAMA 1992;267:386–91.
67. Taylor R. Global Programme on AIDS. A review of methods for estimating the size of subgroups particularly at risk of infection with HIV, and development of proposals which could be used to enumerate these populations in the field: with particular reference to the use of capture-release methods for estimating populations of IV drug users and prostitutes. Geneva, Switzerland: World Health Organization, 1989.
68. van Haastrecht HJ, van den Hoek JA, Bardoux C, et al. The course of the HIV epidemic among intravenous drug users in Amsterdam, The Netherlands. Am J Public Health 1991;81: 59–62.
69. Watts C, Zwi A, Wilson D, et al. Capture-recapture as a tool for programme evaluation. (Letter). BMJ 1994;308:858.
70. Wickens TD. Quantitative methods for estimating the size of a drug-using population. J Drug Issues 1993;23:185–216.
71. Wittes J, Sidel VW. A generalization of the simple capture-recapture model with applications to epidemiological research. J Chronic Dis 1968;21:287–301.
72. Wittes JT, Colton T, Sidel VW. Capture-recapture methods for assessing the completeness of case ascertainment when using multiple information sources. J Chronic Dis 1974;27: 25–36.
73. Woodward JA, Bonett DG, Brecht ML. Estimating the size of a heroin-abusing population using multiple-recapture census. NIDA Res Monogr 1985;57:158–71.
74. York JC, Madigan D. Bayesian methods for estimating the size of a closed population. Technical report no. 234. Seattle, WA: Department of Statistics, University of Washington, 1992.
75. York JC, Madigan D. Bayesian methods for estimation of the size of a closed population. Biometrika (in press).
76. Fienberg SE. The multiple-recapture census for closed populations and incomplete 2^k contingency tables. Biometrika 1972;59:591–603.
77. Seber GAF. A review of estimating animal abundance. Biometrics 1986;42:267–92.
78. Seber GAF. A review of estimating animal abundance. II. Int Stat Rev 1992;60:129–66.
79. Fienberg SE. Bibliography on capture-recapture modeling with application to census undercount adjustment. Surv Methodol 1992;18:143–54.
80. Folstein SE. Huntington's disease: a disorder of families. Baltimore, MD: Johns Hopkins University Press, 1989:103.
81. Annegers JF, Appel S, Lee JR, et al. Incidence and prevalence of amyotropic lateral sclerosis in Harris County, Texas, 1985–1988. Arch Neurol 1991;48:589–93.
82. Lincoln FC. Calculating waterfowl abundance on the basis of banding returns. Circular no. 118. Washington DC: US Department of Agriculture, 1930:1–4.
83. Chapman CJ. Some properties of the hypergeometric distribution with applications to zoological censuses. U California Public Stat 1951;1:131–60.
84. Wittes JT. On the bias and estimated variance of Chapman's two-sample capture-recapture population estimate. Biometrics 1972;28:592–7.
85. Akaike HA. A new look at the statistical identification model IEEE. Trans Auto Control 1974;19:716–23.
86. Sakamoto Y, Ishiguro M, Kitigawa G. Akaike information criterion statistics. Tokyo, Japan: KTK Scientific, 1986.
87. Schwarz G. Estimating the dimension of a model. Ann Stat 1978;6:461–4.
88. Agresti A. Categorical data analysis. New York: John Wiley and Sons, 1990:251.
89. Draper D. Assessment and propagation of model uncertainty. J R Stat Soc [B] 1995;57:45–70.
90. Raftery AE. Discussion of: Draper D. Assessment and propagation of model uncertainty. J R Stat Soc [B] 1995;57:78–9.
91. Kass RE, Wasserman L. Discussion of: Draper D. Assessment and propagation of model uncertainty. J R Stat Soc [B] 1995;57:84–85.
92. Hook EB, Regal RR. Small sample corrections, optimal model selection and adjusting for model uncertainty in three source capture-recapture estimates of closed populations.

Downloaded from academic.oup.com/epirev/article/17/2/243/499931 by Michigan State University user on 12 October 2026

(Abstract). In: Proceedings of the XVIIth International Biometric Conferences, Hamilton, Ontario, Canada, August 8–12, 1994. Vol II: Contributed papers. Alexandria, VA: Biometric Society, 1994:194.
93. Regal RR, Hook EB. The effects of model selection on confidence intervals for the size of a closed population. Stat Med 1991;10:717–21.
94. Zelterman D, Hook EB, Regal, RR. Bayesian estimates of the size of a closed population. (Abstract no. 0209). Abstracts of the 1994 Joint Statistical Meetings in Toronto, Ontario, Canada. Washington, DC: American Statistical Association, 1994.
95. Regal RR, Hook EB. Small sample adjustments for some capture-recapture models with multiple sources. Technical report TR93–8. Duluth, MN: University of Minnesota, Duluth, Department of Mathematics and Statistics, 1993.
96. Evans MA, Bonett DG. Bias reduction for multiple-recapture estimators of closed population size. Biometrics 1994;50: 388–95.
97. Regal RR, Hook EB. Goodness-of-fit based confidence intervals for estimates of the size of a closed population. Stat Med 1984;3:287–91.
98. Cormack RM. Interval estimation for mark-recapture studies of closed populations. Biometrics 1992;48:567–76.
99. Pickands J, Raghavachari M. Exact and asymptotic inference for the size of a population. Biometrika 1987;74:355–63.
100. Lauckner FB. Letter. Biometrics Bull 1991;8:3.
101. Madigan D. Discussion of Draper D. Assessment and propagation of model uncertainty (89). J R Stat Soc [B] 1995;57: 85.
102. Breiman L. The little bootstrap and other methods for dimensionality selection in regression: X-fixed prediction error. J Am Stat Assoc 1992;87:738–54.
103. Chatfield C. Discussion of Draper D. Assessment and propagation of model uncertainty (89). J R Stat Soc [B] 1995;57: 77.
104. Johnson KM, Huether CA, Hook EB, et al. False-positive reporting of Down syndrome on Ohio and New York birth certificates. Genet Epidemiol 1985;2:123–31.
105. Hogan H, Wolter K. Measuring accuracy in a post-enumeration survey. Surv Methodol 1988;14:99–116.
106. Hook EB, Regal RR. Effect of variation in probability of ascertainment by sources ("variable catchability") upon "capture-recapture" estimates of prevalence. Am J Epidemiol 1993;137:1148–66.
107. Hook EB, Regal RR. Re: "Effect of variation in probability of ascertainment by sources (variable catchability) upon capture-recapture estimates of prevalence." (Erratum). Am J Epidemiol 1994;140:856.
108. Darroch JN, Fienberg SE, Glonek GFV, et al. A three-sample multiple-recapture approach to census population estimation with heterogeneous catchability. J Am Stat Assoc 1993;88: 1137–48.
109. Agresti A. Simple capture-recapture models permitting unequal catchability and variable sampling effort. Biometrics 1994;50:494–500.
110. Bonaiti C, Demenais F. Diseases having two classes of severity: estimation of the proportions of each manifestation by maximum likelihood. Biometrics 1980;36:447–60.
111. GLIM: Generalised linear interactive modelling. Version 3.77. London, England: Royal Statistical Society, 1978.
112. Wittes JT. Estimation of population size: the Bernoulli census. PhD thesis. Cambridge, MA: Harvard University, 1970.
113. Cormack RM. Log-linear methods for capture-recapture. Biometrics 1989;45:395–413.
114. Brecht ML, Wickens TD. A comparison of multiple-recapture models for drug prevalence estimation. In: Proceeedings of the Social Statistics Section, American Statistical Assocication, 1991:556–61.
115. Jolly GM. Explicit estimates from capture-recapture data with both death and immigration-stochastic model. Biometrika 1965;52:225–47.
116. Seber GAF. A note on the multiple-recapture census model. Biometrika 1965;52:249–59.
117. Kiemeney LA, Schouten LJ, Straatman H. Ascertainment corrected rates. Int J Epidemiol 1994;23:203–4.
118. Black JF, McLarty D, Mtasiwa DE. Capture-recapture techniques: difficult to use in developing countries. (Letter). BMJ 1994;308:531.
119. Waters WE. Capture-recapture techniques: more unreliable in humans than in birds. (Letter). BMJ 1994;308:531.
120. Neugebauer R, Wittes J. Voluntary and involuntary capture-recapture samples—problems in the estimation of hidden and elusive populations. Am J Public Health 1994;84:1068–9.
121. Carothers AC. Capture-recapture methods applied to a population with known parameters. J Anim Ecol 1973;42: 125–46.
122. Cone RS, Robson DS, Krueger CC. Failure of statistical tests to detect assumption violations in the mark-recapture population: estimation of brook trout in Adirondack ponds. North Am J Fisheries Management 1988;8:489–96.
123. Hook EB, Regal RR. Internal validity analysis: a method for adjusting capture-recapture estimates of prevalence. Am J Epidemiol 1995;142:S48-S52.
124. Weinberg W. Über methoden und Fehlerquellen der untersuchung auf mendelische zahlen beim menschen. (In German). Arch Rassen Gesellsch Biol 1912;9:165–74.
125. Weinberg W. Zur Vererbung der Anlage der Bluterkrankheit mit methodogischen Erganzungen meiner Geschistermethode. (In German). Arch Rassen Gesellsch Biol 1912;9: 694–709.
126. Weinberg W. Mathematische grundlage der probandenmethode. (In German). Z Abstammungs Verebungslehre 1928;48: 179–228.
127. Stene J. The incomplete, multiple ascertainment model: assumptions, applications and alternative models. Genet Epidemiol 1989;6:247–51.
128. Stene J. Ascertainment problems in family data: methodological aspects. In: Lynch HT, Tautu P, eds. Recent progress in the genetic epidemiology of cancer. Berlin, Germany: Springer-Verlag, 1991:95–101.
129. Morton NE. Genetic tests under incomplete ascertainment. Am J Hum Genet 1959;11:1–16.
130. Morton NE. Segregation and linkage. In: Burdette WJ, ed. Methodology in mammalian genetics. San Francisco, CA: Holden-Day, 1963:17–52.
131. Neel JV. Discussion. In: Burdette WJ, ed. Methodology in mammalian genetics. San Francisco, CA: Holden Day, 1963: 47–9.
132. Morton NE, Shields DC, Collins A. Genetic epidemiology of complex phenotypes. Ann Hum Genet 1991;55:301–14.
133. Ewens WJ. Ascertainment biases and their resolution in biological surveys. In: Rao CR, Chakraborty R, eds. Statistical methods in biological and medical sciences. Handbook of statistics, vol 8. Amsterdam, The Netherlands: Elsevier Science Publishers, 1991:29–61.
134. Morton NE, Chung CS, Mi MP. Genetics of interracial crosses in Hawaii. Monographs in human genetics, vol 3. Basel, Switzerland: Karger, 1967:94–103.
135. Zelterman D. Robust estimation in truncated discrete distributions with application to capture-recapture experiments. J Stat Planning Inference 1988;18:225–37.
136. Efron B, Thisted R. Estimating the number of unseen species: how many words did Shakespeare know? Biometrika 1976;63:435–47.
137. Skellam JG. A probability distribution derived from the binomial distribution by regarding the probability of success as variable between the sets of trials. J R Stat Soc [B] 1948;10:257–61.
138. LaPorte RE. Assessing the human condition: capture-recapture techniques. (Editorial). BMJ 1994;308:5–6.
139. Wolter KM. Accounting for America's uncounted and miscounted. Science 1991;253:12–15.
140. Schenker N. Undercount in the 1990 Census. J Am Stat Assoc 1993;88:1044–7.

Downloaded from academic.oup.com/epirev/article/17/2/243/499931 by Michigan State University user on 12 October 2026