

Accuracy of Capture-Recapture Estimates of  
Prevalence

**Ernest B. Hook**, *University of California, Berkeley*

**Melanie S. Hsia**, *University of California, Berkeley*

**Ronald R. Regal**, *University of Minnesota, Duluth*

**Recommended Citation:**

Hook, Ernest B.; Hsia, Melanie S.; and Regal, Ronald R. (2012) "Accuracy of Capture-Recapture Estimates of Prevalence," *Epidemiologic Methods*: Vol. 1: Iss. 1, Article 1.  
DOI: 10.1515/2161-962X.1000

©2012 De Gruyter. All rights reserved.

# Accuracy of Capture-Recapture Estimates of Prevalence

Ernest B. Hook, Melanie S. Hsia, and Ronald R. Regal

## Abstract

Capture recapture methods using data on cases appearing on overlapping incomplete lists have been widely used for prevalence estimates. Nevertheless, workers have, in general, employed several different methods and there appears no consensus as to which is optimal.

The authors, using several different methods, undertook an “empirical” investigation of data in 21 reports available in the literature, for which they could compare the known total in a group with the capture recapture estimate derived from data given on three overlapping lists. Comparisons were undertaken of the accuracy of estimates derived from one of six methods, Bayesian or non-Bayesian.

The authors found use of the most complex log-linear model for three sources, the *all-two-way interaction model*, a method not in general use, generated estimates notably more accurate and with greater coverage of the true value by both calculated and distributional intervals than those generated by more frequently used methods. Moreover, the lower limits on these estimates were “better” (i.e. lower than the true values but closer to them). The upper limits of the estimates generated by *all-two-way interaction model*, were however, “worse”, very often uselessly large.

The authors suggest investigators consider estimates and lower limits derived from the *all-two-way interaction model* in addition to and in comparison with those derived from any other methods traditionally employed.

**KEYWORDS:** capture-recapture, confidence limits, log-linear model, prevalence

**Author Notes:** We thank two reviewers for their comments and suggestions.

Authors: Hook, Ernest B.: School of Public Health, University of California, Berkeley. Hsia, Melanie S.: Formerly, School of Public Health, University of California, Berkeley. Regal, Ronald R.: Department of Mathematics and Statistics, University of Minnesota, Duluth.

Please send all correspondence to: Ernest B. Hook, School of Public Health, University of California, Berkeley 50 University Hall; M.C. 7360 Berkeley, CA 94720-7360, USA

Phone: (510) 642-4490, Fax: (510) 643-9588, E-mail: ebhook@berkeley.edu

## Materials and methods

### *Introduction*

For over the past two decades, numerous epidemiologists have used a method for estimating the prevalence of various conditions called “capture recapture”. The terminology derives from use of such methods in ecology, e.g. ornithology. Birds are captured, “marked”, released, and recaptured. The proportion of marked birds among those recaptured enables, assuming independence of the captures, a ready estimate of the population total.

Capture recapture estimates in humans are derived from data on two or more incomplete overlapping lists or “sources” of individuals with some aspect of interest, e.g. a specific disease. Almost always, lists or sources of cases are not random samples from a population but samples of convenience. They may well be positively or negatively dependent.

Two sources generate a unique capture recapture estimate. But if they are negatively dependent, then the two-source capture recapture estimate is biased to an overestimate and, if positively dependent, biased to an underestimate. With three or more sources, investigators have used several competing statistical procedures to adjust for possible dependencies. A goal of this paper is to evaluate some procedures suggested in the literature and a new procedure.

If information is available from a study with data from four overlapping incomplete sources of cases in a population, designated say, A, B, C and D, then as exhibited in table one, the investigator can compare the number known in one source, say A, which we designate, per the suggestion of a reviewer, the “set-aside source”, with any capture recapture estimate derived from information available on cases in the other three sources B, C, and D and also in A. One may then repeat this comparison of a derived estimate and the known number, treating sources B, C and D as set-aside sources (Hook and Regal, 1982, 1997).

### *Data*

We analyzed data from 21 studies reported in English which included information from four or more sources, had presented data observed in each of the  $2^4 - 1$  possible intersections of sources, were published since earlier analyses (Hook and Regal, 1982, 1997, 2000), and were available to us by July 2006 (Hay, 1997, 2000; Orton et al., 2001; Aaron et al., 2003; Armstrong et al., 2003; Pezzotti et al., 2003; Hickman et al., 2004; Rosenman et al., 2004; Coffman et al., 2005; King et al., 2005; Uusküla et al., 2005; Hall et al., 2006; Holland et al., 2006; Rosenman et al., 2006; Du, 1997; Laing et al., 1997; Bishop et al., 1975). Two reports (Hay, 2000; Hickman et al., 2004) presented data from more than one

area: one from urban and rural regions (Hay, 2000), the other on three different cities (Hickman et al., 2004). The report on the three cities stratified data by age and gender. The largest sub-stratum of data, by a wide margin, in each city included males, aged 30-44. To enable homogeneity, we confined our analyses of data in each of the cities to this group specifically. We found no apparent intra-study correlation of results on the populations analyzed in either report. We analyzed the data from each of these five as from a "distinct" study. Three studies (Hook and Regal, 2000; Du, 1997; Laing et al., 1997) provided data from five sources. We removed the source with the smallest number observed and analyzed the data from the remaining four sources.

### *Methods of Analysis*

While as discussed further below, there are a number of methods for generating estimates, almost all epidemiologists have used log-linear methods (Bishop et al., 1975). These were first applied to prevalence estimates by Wittes (1970) in a seminal work, and have been developed and applied extensively since their application in a textbook (Bishop et al., 1975). Using log-linear methods with three lists, the investigator chooses estimates generated by one of eight models, from the simplest, independence of all sources, to the most complex, the presence of all two-source interactions (ATSI), or some combination of all eight, as discussed further below.

In the literature we have found, further, that when undertaking model selection, virtually all epidemiologists have used the estimate generated by the simplest model not rejected at a particular probability level by a chi-square test, e.g. 0.05 (which we designate here as "traditional") or the model meeting the information criterion of Akaike (AIC). A Bayesian information criterion (BIC) has been proposed for model selection, and Draper (1995) proposed an empirical Bayes procedure in which the estimate is derived by weighting the estimates associated with each model by their BIC, which we denote as "weighted BIC". We have suggested an analogous procedure, which we denote as "weighted AIC" (Hook and Regal, 1997). This, rather than AIC or the "traditional" method, has been employed in at least one report deriving prevalence estimates (Hall et al., 2006).

We derived confidence limits of the weighted BIC and weighted AIC estimates in a manner analogous to that used to generate the estimates. We also consider here the model that always chooses the most complex model available with three sources, the presence of all two-source interactions (ATSI). As it has no degrees of freedom, one may designate it as "saturated".

We compared these six methods on data from three sources, derived as described above, i.e. 1) the traditional, using  $p = 0.05$ , 2) BIC, 3) weighted BIC, 4) AIC, 5) weighted AIC, and 6) the all-two-source-interaction model.

*Small sample adjustment*

Each three-source or, equivalently, three-list data set generates a "table" with  $(2^3 - 1) = 7$  entries. Each entry in the table is the number observed in one of seven "cells", i.e. a distinct intersection of lists. (See table 1, which illustrates this for a set-aside set.) From the known values in each cell the investigator generates an estimate of the number in the cell whose value is not known, i.e. the number appearing on no list, which in table 1 is h, and from this of course may estimate the total in the entire set-aside set, A.

**Table 1. Individuals in Source A\***

	<b>B yes</b>	<b>B no</b>		<b>B yes</b>	<b>B no</b>
	<b>C yes</b>	<b>C no</b>		<b>C yes</b>	<b>C no</b>
<b>D yes</b>	a	b		e	f
<b>D no</b>	c	d		g	h

\*There are  $a + b + c + d + e + f + g$  individuals in the "set-aside" source, Source A, and at least one other source, and h individuals only in Source A. The approach treats h and the total in Source A as unknowns. It estimates, with various capture recapture methods, the number in Source A from knowledge only of the numbers of individuals in Source A and Source B and/or C and or/ D. As discussed in the text, any estimate is evaluated by comparing it with the known total in Source A.

We applied an adjustment proposed by Evans and Bonnett (1994), where, if there are k sources adds  $1/2^{k-1}$  to the number in each cell. With three sources as here, we added 0.25 to each cell. Even if the entire number observed is very large, a null value in some cells may result in an indefinitely large estimate.

In any study, the investigator may observe in one (or more) source(s) a relatively small number of cases, e.g. fewer than 50. Estimates, especially those generated by the all-two-source-interaction model, are sensitive to small numbers in cells even if an adjustment is applied. Therefore, we subdivided the data sets into those in which the investigator observed at least 50 cases in all sources in any one study and those in which the investigator observed fewer than 50 cases in one

or more sources in that study. We designate data sets in the former type as from large studies and from the latter as from small studies,

### *Measurement of accuracy*

As previously, we measured accuracy by evaluating what we term the "log relative bias" of any estimate. If  $N$  is the number known in a source and  $N_e$  is the capture recapture estimate, then

$$\text{log relative bias } (N_e) = \ln (N_e/N)$$

If the value is zero, then the estimate is perfect; if negative, then an underestimate; if positive, then an overestimate. Of any two estimates the one whose value is closest to zero is more accurate. We compared methods by comparison of the means, one-tailed 5% confidence limits, coverage by the 90% confidence intervals, medians and distribution boundaries.

To account for potentially correlated results from data sets created from the same study and from measures generated on the same table, we assessed the significance of the comparisons of the estimation methods by analyzing the results with a mixed model with random effects for studies and tables within studies.

## **Results**

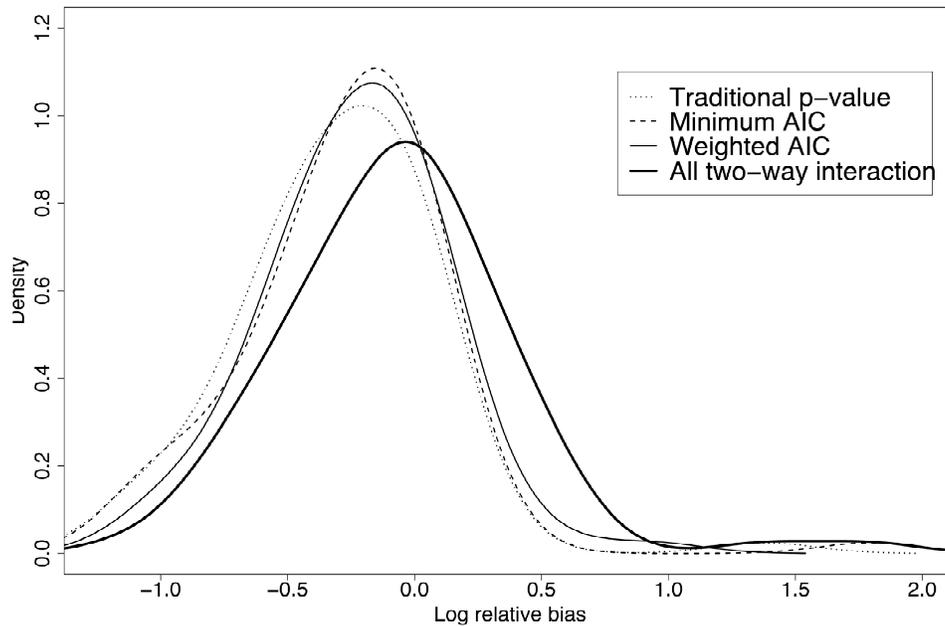
These appear in the figures and table 2. (Note that the values of log relative bias in table 2 must be exponentiated to derive the values of percent underestimates or overestimates. For instance, a log relative bias of  $-0.30$  results in an underestimate of  $(1-\exp(-0.3))$  or 26%.) Estimates from models selected based on AIC and weighted AIC performed consistently, albeit in general modestly, better than estimates derived from using BIC and weighted BIC. For clarity we exclude the results derived from BIC and weighted BIC from table 2.

**Table 2. Log relative bias calculated with various methods of model selection or adjustment.**

	Studies in which at least one source has < 50 cases				Studies in which all sources have ≥ 50 cases				All			
	Trad	AIC	Wt'd	ATSI	Trad	AIC	Wt'd	ATSI	Trad	AIC	Wt'd	ATSI
Number of Studies	9				12				21			
Number of sources	36				48				84			
Method of estimation*	Trad	AIC	Wt'd	ATSI	Trad	AIC	Wt'd	ATSI	Trad	AIC	Wt'd	ATSI
Mean	-0.40	-0.33	-0.30	-0.14	-0.22	-0.21	-0.18	-0.01	-0.30	-0.26	-0.23	-0.06
(SD)	0.34	0.50	0.39	0.56	0.39	0.29	0.29	0.34	0.38	0.40	0.34	0.45
Median	-0.32	-0.32	-0.27	-0.20	-0.19	-0.14	-0.17	0.00	-0.27	-0.15	-0.21	-0.08
90% Coverage	13/36	14/36	16/36	32/36	15/48			39/48	28/84	32/84	59/84	71/84
					16/48		31/48					
5% limit	-0.95	-1.16	-0.94	-0.79	-0.86	-0.69	-0.91	-0.57	-0.92	-0.91	-0.79	-0.80
95% limit	0.16	0.50	0.35	1.36	0.42	0.26	0.49	0.56	0.32	0.39	0.33	0.67
5th percentile	-1.05	-1.04	-0.93	-0.77	-0.75	-0.73	-0.56	-0.43	-0.95	-1.02	-0.81	-0.76
10th percentile	-0.94	-0.99	-0.81	-0.75	-0.64	-0.59	-0.52	-0.36	-0.75	-0.80	-0.63	-0.61
25th percentile	-0.58	-0.54	-0.53	-0.52	-0.50	-0.39	-0.39	-0.22	-0.52	-0.47	-0.46	-0.35
75th percentile	-0.16	-0.10	-0.02	0.21	0.01	-0.00	0.01	0.15	-0.02	-0.02	0.00	0.16
90th percentile	0.00	0.03	0.09	0.43	0.11	0.10	0.14	0.32	0.05	0.05	0.11	0.40
95th percentile	0.06	0.06	0.15	0.54	0.17	0.16	0.23	0.48	0.14	0.14	0.25	0.51

\* Trad, Min, Wt'd, ATSI denote the traditional P-value, minimum AIC, weighted AIC, and all-two source interaction methods respectively.

**Figure 1. Density distributions of log relative biases by model**



There was an overall bias to underestimate although it was of small magnitude for estimates derived from the all-two-source-interaction model. For example for all studies, for the traditional, the AIC, the weighted AIC, and the all-two-way-interaction approaches the mean estimates were respectively, 26%, 23%, 21% and 6% underestimates of the known values. For the median estimates the analogous figures were 24%, 21%, 19%, and 4% respectively. With all approaches, the estimates were notably better in the analysis of the "large" than of the "small" studies.

However, the upper confidence limits, or upper percentile values, associated with the all-two-source-interaction approach were notably higher than those associated with the others evaluated. For example, for the traditional, the AIC, the weighted AIC, and the all-two-way interaction approaches with all studies, the upper 90th percentiles of the distributions are greater than the known values by 5%, 5%, 12%, and 49% respectively. (For the lower 90th percentile by contrast, the results are 53%, 54% 47% and 46% less than the known values respectively. Here the all-two-way-interaction approach performs slightly, albeit not markedly better.)

The mean log relative bias of the estimate derived from the all-two-source-interaction model was markedly and significantly different from each of the other methods ( $P < 0.0001$ ). The only comparison amongst the other methods of significance was the greater accuracy of estimate associated with weighted AIC

(mean log relative bias = -0.23) compared to that derived from the traditional method (mean log relative bias = -0.30;  $P=0.048$ ). The mean log relative bias of the estimate derived from the all-two-source-interaction model was not significantly different from zero ( $P = 0.18$ ), while the results for the other three methods were clearly biased to underestimates ( $P<0.0001$ ).

Figure 1 illustrates the observed distributions of log relative biases across all data sets shown as nonparametric kernel density estimates. There were two cases of outlying overestimation resulting in the slight right tail bumps in the fitted densities. With the exception of two outliers, the distributions suggest a spread surprisingly similar to a Gaussian distribution. The figure reinforces the clear advantage of estimates derived from the all-two-source-interaction model compared to those derived from the other three options.

## Conclusions

### *Discussion*

Our results in these data sets indicate the all-two-source interaction model is optimal for generating estimates. They are also consistent with a previous finding by two of us that even if one knew of or could postulate highly simple, plausible processes acting on a particular population that led to ascertainment by different sources, the all-two-source-interaction model, the most complex, is optimal (Regal and Hook, 1998). And as a reviewer has commented, the results are also noteworthy in indicating the “traditional” method and AIC, the methods used by many epidemiologists, result in significant underestimates of prevalence.

There are many other possible statistical methods, although none have had the same attractiveness for epidemiologists. Some statisticians have advocated using empirical Bayesian methods (Draper, 1995), full Bayesian methods (King et al., 2001, 2005; Madigan and York, 1997) or various classes of product multinomial, Poisson or logit models (Bishop et al., 1975; International Working Group for Disease Monitoring and Forecasting, 1995; Tilling and Sterne, 1999). The models of Darroch et al. (1993), used for example by the International Working Group (1995), are special in this regard in that they involve symmetry of the sources and fit a common resulting degree of dependency between the  $k$  sources. This sort of symmetry between sources as a result of combined source dependencies and capture heterogeneity we feel are likely more applicable to ecological studies employing repetitions of the same capture strategies over time.

If data on covariates are available, then methods taking these into account and using, for instance, explicit conditional logit models or Poisson models are very attractive (Tilling and Sterne, 1999; Alho et al., 1993; Tilling et al., 2001; Platt et al., 2004).

While we have analyzed “set-aside sources” or subsets so to speak extracted from data reported in the literature, they are, nevertheless, similar in type and nature to those used by epidemiologists in deriving capture recapture estimates of prevalence in an entire population. Certainly they appear useful in comparing performance of potential estimation methods. One possible limitation is that all analyses have been conditional upon there being cases in the population in at least the set aside source. It is striking that despite this, the most complex model available has clearly performed the best and the simpler models resulted in notable underestimates. It seems unlikely to us that any simpler model would be preferable to the most complex model for estimating prevalence in general.

Some standard assumptions in capture recapture analysis applied in epidemiology are that records of affected individuals have been matched appropriately and there has been no migration of affected individuals in or out of the population during the time interval studied (Hook and Regal, 1995). Sometimes these assumptions are “latent”. For instance, both recovery from a disease or death result, in essence, in “emigration” from the population in the time interval studied. A related assumption is that there is no “variable catchability” of individuals in the population by sources, even of individuals who remain affected and alive in the interval studied. (Hook and Regal, 1993). Violations of these assumptions will tend to result in deviation from independence of sources. Selection of more complex models in essence attempts to adjust for the consequences of violations of these underlying assumptions (Hook and Regal, 1995).

One important assumption is that any individual whose name appears on a list of cases has been correctly diagnosed. The investigator may find it particularly worthwhile to consider the plausibility of this assumption, confirm, to the extent possible, the diagnosis, and consider the likely biases if cases are included who are misdiagnosed. She or he may well conclude after doing this that no useful or reliable capture recapture estimate of prevalence is possible.

### *Conclusion*

What inferences may be drawn from our results for studies in which the investigator does not know the true population size, and what guidance may be offered? As emphasized elsewhere (Hook and Regal, 1999), in the spirit of W.E. Deming's approach to application of statistical methods, she or he must consider the choices for estimation in the light of intended use of the data.

We believe it reasonable, first, to review the likely accuracy of diagnoses of cases appearing on available lists and either to exclude those lists which appear unreliable or attempt to confirm diagnoses. Second, then to generate from apparently reliable lists the estimates and intervals by alternative methods, to

examine how these differ, to consider why they do, and what impact the use of different estimates and intervals might have on the investigator's intended goal of the prevalence estimates.

If there is no reason in any study not to do so, then on the basis of the results here, we believe it reasonable, with three sources, to choose the estimate generated by the all-two-source-interaction model and use its lower limit as a plausible lower boundary on the derived estimate.

Calculation of an appropriate upper limit of the estimate derived from the all-two-source-interaction model provokes difficulties. The penalty of its use is that the upper limit of the generated estimate is, on average, not only much higher than that generated by the other methods as exhibited in the table, but, in particular instances, the calculated limit may be unreasonably, indeed uselessly large. The upper limit derived from a simpler model, e.g. based on AIC, on average gives a closer bound on the actual population size, but the coverage of these intervals is worse. Yet rarely does an investigation require a precise estimate or interval. A range derived from or corresponding to alternative approaches would be worth considering.

## References

- Aaron DJ, Chang Y-F, Markovic N, LaPorte RE. Estimating the lesbian population: a capture-recapture approach. *Journal of Epidemiology and Community Health* 2003; 57 : 207-9.
- Alho JM, Mulry MH, Wunderman K, Kim J. Estimating heterogeneity in the probabilities of enumeration for the dual-system estimation. *Journal of the American Statistical Association* 1993; 88 : 1130-6.
- Armstrong B, Busby A, Dolk H. Using Capture-Recapture Methods to Ascertain Completeness of a Register: Case Study and Methodological Considerations. In: EUROCAT, ed. *European Surveillance of Congenital Anomalies*. London: London School of Hygiene and Tropical Medicine; University of Ulster, 2003.
- Bishop YMM, Fienberg SE, Holland PW. *Discrete multivariate analysis: theory and practice*. Cambridge, MA: MIT Press, 1975 : 229-56.
- Coffman CJ, Horner RD, Grambow SC, Lindquist J. Estimating the occurrence of amyotrophic lateral sclerosis among Gulf War (1990-1991) Veterans using capture-recapture methods - An assessment of case ascertainment bias. *Neuroepidemiology* 2005; 24 : 141-50.
- Darroch JN, Fienberg SE, Glonek GFV, Junker BW. A three-sample multiple-recapture approach to census population estimation with heterogeneous catchability. *Journal of the American Statistical Association* 1993; 88 : 1137-48.

- Draper D. Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society Series B* 1995; 57 : 45-97.
- Du XL. A community-based stroke register and case-control study of stroke and the quality of hypertension control in northwest England. Manchester, England: University of Manchester, 1997.
- Evans MA, Bonett DG. Bias reduction for multiple-recapture estimators of closed population-size. *Biometrics* 1994; 50 : 388-395.
- Hall HI, Song R, III JEG, Lee LM. Assessing the Completeness of Reporting of Human Immunodeficiency Virus Diagnoses in 2002-2003: Capture-Recapture Methods. *American Journal of Epidemiology* 2006; 164 : 391-397.
- Hay G. The selection from multiple data sources in epidemiological capture-recapture studies. *The Statistician* 1997; 46 : 515-520.
- Hay G. Capture-recapture estimates of drug misuse in urban and non-urban settings in the north east of Scotland. *Addiction* 2000; 95 : 1795-1803.
- Hickman M, Higgings V, Hope V, et al. Injecting drug use in Brighton, Liverpool, and London: best estimates of prevalence and coverage of public health indicators. *Journal of Epidemiology and Community Health* 2004; 58 : 766-71.
- Holland R, Vivancos R, Maskrey V, et al. The prevalence of problem drug misuse in a rural county of England. *Journal of Public Health* 2006; 28 : 88-95.
- Hook EB, Regal RR. Validity of Bernoulli census, log-linear, and truncated binomial models for correcting for underestimates in prevalence studies. *American Journal of Epidemiology* 1982; 116 : 168-176.
- Hook, E.B., Regal, R.R.: Effect of variation in probability of ascertainment by sources ("variable catchability") upon "capture-recapture" estimates of prevalence. *American Journal of Epidemiology* 1993 137: 1148-1166.
- Hook EB, Regal RR. Capture-Recapture Methods in Epidemiology: Methods and Limitations. *Epidemiologic Reviews* 1995; 17 : 243-264.
- Hook EB, Regal RR. Recommendations for presentation and evaluation of capture-recapture estimates in epidemiology. *Journal Of Clinical Epidemiology* 1999; 52 : 917-926.
- Hook EB, Regal RR. Validity of Methods for Model Selection, Weighting for Model Uncertainty, and Small Sample Adjustment in Capture-Recapture Estimation. *American Journal of Epidemiology* 1997; 145 : 1138-1144.
- Hook EB, Regal RR. Accuracy of Alternative Approaches to Capture-Recapture Estimates of Disease Frequency: Internal Validity Analysis of Data from Five Sources. *American Journal of Epidemiology* 2000; 152 : 771-779.
- International Working Group for Disease Monitoring and Forecasting. Capture-recapture and multiple record systems I: history and theoretical development. *American Journal of Epidemiology* 1995; 142 : 1047-57.

- King R, Bird SM, Brooks SP. On the Bayesian Analysis of Population Size. *Biometrika* 2001; 88 : 317-36.
- King R, Bird SM, Brooks SP, et al. Prior information in behavioral capture-recapture methods: Demographic influence on drug injectors' propensity to be listed in data sources and their drug-related mortality. *American Journal of Epidemiology* 2005; 162 : 694-703.
- Laing TJ, Gillespie BW, Toth MB, et al. Racial differences in Scleroderma among women in Michigan. *Arthritis & Rheumatism* 1997; 40 : 734-42.
- Madigan D, York JC. Bayesian methods for estimation of the size of a closed population. *Biometrika* 1997; 84 : 19-31.
- Orton H, Rickard R, Miller L. Using Active Medical Record Review and Capture-Recapture Methods to Investigate the Prevalence of Down Syndrome among Live-Born Infants in Colorado. *Teratology* 2001; 64 : S14-S19.
- Pezzotti P, Piovesa C, Michieletto F, et al. Estimating the cumulative number of human immunodeficiency virus diagnoses by cross-linking from four different sources. *International Journal of Epidemiology* 2003; 32 : 778-83.
- Platt L, Hickman M, Rhodes T, et al. The prevalence of injecting drug use in a Russian city: implications for harm reduction and coverage. *Addiction* 2004; 99 : 1430-38.
- Regal RR, Hook EB. Marginal vs conditional vs structural models: A rationale for alternatives to log-linear models for capture-recapture estimation. *Statistics in Medicine* 1998; 17 : 69-74.
- Rosenman KD, Kalush A, Reilly MJ, et al. How much work-related injury and illness is missed by the current national surveillance system? *Journal of Occupational and Environmental Medicine* 2006; 48 : 357-65.
- Rosenman K, Reilly MJ, Kalinowski DJ. 2003 Annual Report on Silicosis in Michigan. East Lansing; Lansing: Michigan State University, Department of Medicine; Michigan Department of Labor and Economic Growth, 2004 : 30.
- Tilling K, Sterne JAC. Capture-recapture models including covariate effects. *American Journal of Epidemiology* 1999; 149 : 392-400.
- Tilling K, Sterne JAC, Wolfe CDA. Estimation of the incidence of stroke using a capture-recapture model including covariates. *International Journal of Epidemiology* 2001; 30 : 1351-1359.
- Uusküla A, Abel K, Rajaleid K, et al. The Prevalence of Injecting Drug Use in Estonia, 2004. Tartu: University of Tartu, Imperial College London, 2005 : 42.
- Wittes JT. Estimation of population size: the Bernoulli census. Department of Statistics. Boston: Harvard University, 1970.