

## Recent developments in life and social science applications of capture–recapture methods

Dankmar Böhning · Peter van der Heijden

Published online: 2 December 2008  
© Springer-Verlag 2008

**Keywords** Dual-system estimation · Estimation using several sources · Heterogeneity · Horvitz–Thompson estimator · Lincoln–Petersen estimator · Log-linear models · Mark-recapture · Mixture models · Population size estimation

Over the last 20 years capture–recapture methods have experienced important developments, in particular in their applications in the life and social sciences. It appears appropriate to take a closer look at some of these developments. A recent conference entitled *Recent Developments in Capture–Recapture Methods and their Applications* was held in 2007 at The University of Reading. A special issue focusing on applications mainly in the Biological Sciences appeared elsewhere (Böhning 2008), whereas in this special topic we would like to focus more on life and social science applications.

The capture–recapture or mark-and-recapture methodology goes back to the Biological/Ecological Sciences with the work of Lincoln and Petersen. About one hundred and ten years ago Petersen (1896) published his landmark paper suggesting what later became known as the Lincoln–Petersen estimator, since it was also independently developed by Lincoln (1930). This estimator and its stabilized version developed by Chapman (1951) is still in use by numerous practitioners.

---

D. Böhning (✉)  
Quantitative Biology and Applied Statistics, School of Biological Sciences, Whiteknights, Reading,  
RG6 6BX, UK  
e-mail: [d.a.w.bohning@reading.ac.uk](mailto:d.a.w.bohning@reading.ac.uk)

P. van der Heijden  
Department of Methodology and Statistics, Utrecht University, P.O. Box 80.140, 3508 TC, Utrecht,  
The Netherlands  
e-mail: [p.g.m.vanderheijden@uu.nl](mailto:p.g.m.vanderheijden@uu.nl)

In the biological sciences the classical book by Seber (1982) characterizes the various biological application fields which has recently been supplemented with the monograph by Borchers et al. (2002). The landmark paper by Sekar and Deming (1949) applied capture–recapture methods to estimate the number of births and deaths in an area near Calcutta and is probably one of the earliest applications of the Lincoln–Petersen estimator to demography. The paper started a development of the capture–recapture methodology in demography and social sciences in general and was further fueled and supported by the appearance of the book by Bishop et al. (1975) on multivariate discrete analysis in which the entire Chap. 6 was devoted to capture–recapture methods including many applications of multiple sources/list estimators to demographic and social sciences problems. Yet, another development was initiated with the establishment of the methodology in medicine and public health. Although occasionally papers appeared in the area before the 1990s including Wittes (1972) and Goldberg and Wittes (1978), the substantial development was initiated by the landmark papers by Hook and Regal (1995) and the two papers by the International Working Group for Disease Monitoring and Forecasting (IWGDMF 1995a, 1995b). In the early phases the method was used to develop a more complete picture of disease occurrence. Nowadays, the methodology is used in screening and preventive studies and is establishing its place in clinical settings.

The special issue at hand presents four papers. In the first paper, *van der Heijden, Zwane, and Hessen* discuss an extension of the conventional two- or multiple source situation for two cases where modifications of the standard method are required. The first problem is that the lists refer to different but overlapping populations. The second problem is that each list has a set of covariates and the sets of covariates are not identical. By considering both problems as missing data problems, they suggest a solution obtained through the EM algorithm. Both problems are illustrated with data sets from epidemiology and demography.

The second paper by *Brittain and Böhning* consider the classical two-source situation and suggest as an alternative to the conventional Chapman estimator the estimator suggested by Chao (1987), which is modified here for a binomial density with trial parameter 2. In order to evaluate these estimators, they are applied to real, three-source, capture–recapture data available from the literature. By conditioning on each of the sources of three-source data, they have been able to compare the estimators with the true value that they are estimating. In addition, a simulation study was executed providing evidence that the Chao estimator provides some benefit over the Chapman estimator.

In a third paper, *Fienberg and Manrique-Vallier* make the interesting observation that there are now three essentially separate literatures on the topics of multiple systems estimation, record linkage, and missing data. In practice, however, the three are intimately intertwined. For example, record linkage involving multiple data sources for human populations is often carried out with expressed goal of developing a merged database for multiple system estimation (MSE). Similarly, one way to view both the record linkage and MSE problems is as ones involving the estimation of missing data. Fienberg and Manrique-Valier work out in detail how these areas are connected. They also provide numerous examples including estimating the killings in Kosovo or estimating the size of the WWW.

The fourth paper by *Kuhnert and Böhning* discusses the nonparametric mixture approach for marginal modeling in capture–recapture. The background of mixtures and their flexibility in coping with population heterogeneity are highlighted. The focus on the paper is on algorithmic approaches, and a software tool is also presented for performing various tasks. Examples illustrate theory, algorithms, and software tool.

**Acknowledgement** The authors would like to thank the Editor of the journal *AStA Advances in Statistical Analysis*, Prof. Dr. Wilfried Seidel, for the possibility to publish this work.

## References

- Bishop, Y.M.M., Fienberg, S.E., Holland, P.W.: *Discrete Multivariate Analysis: Theory and Practice*. MIT Press, Cambridge (1975)
- Böhning, D.: Editorial—Recent developments in capture–recapture methods and their applications. *Biom. J.* **50** (2008)
- Borchers, D.L., Buckland, S.T., Zucchini, W.: *Estimating Animal Abundance: Closed Populations*. Springer, Berlin (2002)
- Chao, A.: Estimating the population size for capture–recapture data with unequal catchability. *Biometrics* **43**, 783–791 (1987)
- Chapman, D.G.: Some properties of the hypergeometric distribution with applications to zoological censuses. *Univ. Calif. Publ. Stat.* **1**, 131–160 (1951)
- Goldberg, J.D., Wittes, J.T.: The estimation of false negatives in medical screening. *Biometrics* **34**, 77–86 (1978)
- Hook, E.B., Regal, R.: Capture–recapture methods in epidemiology: methods and limitations. *Epidemiol. Rev.* **17**, 243–264 (1995)
- International Working Group for Disease Monitoring Forecasting (IWGDMF): Capture–recapture and multiple record systems estimation I. History and theoretical development. *Am. J. Epidemiol.* **142**, 1047–1058 (1995a)
- International Working Group for Disease Monitoring Forecasting (IWGDMF): Capture–recapture and multiple record systems estimation II. Application in human diseases. *Am. J. Epidemiol.* **142**, 1059–1068 (1995b)
- Lincoln, F.C.: Calculating waterfowl abundance on the basis of banding returns. *USDA Circ.* **118**, 1–4 (1930)
- Petersen, C.G.J.: The yearly immigration of young plaice into the Limfjord from the German sea. *Rep. Dan. Biol. St.* **6**, 5–84 (1896)
- Seber, G.A.F.: *The estimation of animal abundance*, 2nd edn. Griffin, London (1982)
- Sekar, C., Deming, W.E.: On a method of estimating birth and death rates and the extent of registration. *J. Am. Stat. Assoc.* **44**, 101–115 (1949)
- Wittes, J.T.: On the bias and estimated variance of Chapman’s two-sample capture–recapture population estimate. *Biometrics* **28**, 592–597 (1972)