

# Book Review: Capture-Recapture Methods for the Social and Medical Sciences. Böhning, Heijden and Bunge (eds.)”

Daniel Manrique-Vallier

November 13, 2023

Capture-Recapture (CR) is the generic name of a family of statistical methods for estimating the size of a finite population based on one or more incomplete samples. Although the first documented application of a similar method was to estimate France’s population (Laplace, 1786; Amorós, 2014), CR methods have primarily evolved in the ecological sciences for the estimation of animal abundance. They can also be used for estimating elusive or difficult to enumerate human populations. Applications abound: number of sex workers in an African country (Okiria et al., 2019), illegal drug users (Leclerc et al., 2014), casualties in armed conflicts (Manrique-Vallier et al., 2013), under-enumeration in censuses (Fienberg, 1992), *et cetera*.

*Capture-Recapture Methods for the Social and Medical Sciences*, edited by Böhning, Heijden and Bunge is the first book explicitly centered on applications of Capture-Recapture methods to human populations and medical problems. The book comprises 25 research articles plus an introduction. It is organized into eight sections, six centered around specific families of CR methods: ratio regression models, extensions of single source models, multiple sources, latent variable models, and Bayesian approaches. The other two sections contain articles on meta-analysis of CR studies, and on “miscellaneous topics”. Most of the articles follow the same structure: introduce a family of distributions that can be fit to observed data, develop methods for fitting the model to data and extrapolating to the unobserved individuals, and illustrate with application examples. The selection of methods is highly idiosyncratic, with a heavy bias towards frequentist single-list techniques.

The most interesting aspect of this book are the example applications. Most of them are taken from outside the classical setting of estimating animal abundance, and illustrate well the applicability of CR methods to diverse domains in social science and medicine. Through examples of estimation of cholera cases India, homeless population in Utrecht, hidden intravenous drug users in Los Angeles, *etc.*, readers can get a glimpse at the wide applicability of CR to whole new areas, and draw inspiration for finding new ones.

On the negative side, one major deficiency shared by most articles in this book is that they offer little guidance on why and how one would prefer a CR method over another. Most articles are limited to presenting methods as flexible candidate distributions for observed data (for example, generalizations of specific mixtures) from which it is possible to extrapolate to unobserved data. However, they do not discuss the observation mechanisms implied by the

models, nor their relationship to the actual real-life situations they want to model. This is a major omission in a book purported to introduce CR methods “for the Social and Medical Sciences”. Most of the example applications are indeed in these domains. However, with the exception of a few articles in Section 6 (latent variable models) I have not been able to find discussions on how or why the presented methods are appropriate for any particular data context, beyond the fact that they are flexible (for fitting the observed data). In fact, I believe that nothing would be lost in this exposition if all human-population datasets in the examples were replaced by datasets from any other application domain.

Another related limitation is the limited selection of topics. Most of the methods introduced here belong to the sub-category of single-list methods. This reflects the authors personal preferences and experience, but it does not make for a good overview. Many of the chapters discuss methods that seem to be just slight variations of one another. As an extensive methodological and applied literature over the last 40 years attests, there are many other CR approaches (specifically multi-list and Bayesian approaches) that have successfully been successfully applied to the estimation of human populations that should have been better represented in a collection such as this one.

*Capture-Recapture Methods for the Social and Medical Sciences* deserves praise for being the first of its class. It is a good catalog of single-list CR methods, and contains some limited information about multi-list methods. Despite its shortcomings, it can be useful for raising awareness among applied researchers of the applicability of capture-recapture methods for estimating human populations.

## References

- Amorós, J. (2014), “Recapturing Laplace,” *Significance*, 11, 38–39.
- Fienberg, S. E. (1992), “Bibliography on capture-recapture modelling with application to census undercount adjustment,” *Survey methodology*, 18, 143–154.
- Laplace, P. S. (1786), “Sur les naissances, les mariages et les morts,” in *Histoire de L’Académie Royale des Sciences*.
- Leclerc, P., Vandal, A. C., Fall, A., Bruneau, J., Élise Roy, Brissette, S., Archibald, C., Arruda, N., and Morissette, C. (2014), “Estimating the size of the population of persons who inject drugs in the island of Montréal, Canada, using a six-source capture–recapture model,” *Drug and Alcohol Dependence*, 174–180.
- Manrique-Vallier, D., Price, M., and Gohdes, A. (2013), “Multiple Systems Estimation Techniques for Estimating Casualties in Armed Conflicts,” in *Counting Civilian Casualties: An Introduction to Recording and Estimating Nonmilitary Deaths in Conflicts*, eds. Seybolt, T. B., Aronson, J. D., and Fischhoff, B., US: Oxford University Press, Studies in Strategic Peacebuilding, chap. 9, pp. 165–182.
- Okiria, A. G., Bolo, A., Achut, V., Arkangelo, G. C., Michael, A. T. I., Katoro, J. S., Wesson, J., Gutreuter, S., Hundley, L., and Hakim, A. (2019), “Novel Approaches for Estimating Female Sex Worker Population Size in Conflict-Affected South Sudan,” *JMIR Public Health Surveill*, 5, e11576.