

How Many Cases Are Missed When Screening Human Populations for Disease? A Regression Based Capture-Recapture Approach

Dankmar Böhning

Department of Mathematics and Statistics
School of Mathematical and Physical Sciences
University of Reading, Whiteknights
Reading, RG6 6BX, UK
email: d.a.w.bohning@reading.ac.uk

July 13, 2011

Abstract

This work has been motivated by the following situation. Human populations are frequently screened for specific diseases with the aim to detect the disease early when it is easier to treat and cure. However, only for those at risk (at risk being defined as having a positive screening test result) it is meaningful to verify the disease status. As a result a large portion of the screened population remains unverified in their disease status. We investigate techniques to estimate the amount of disease present in the unverified population. In particular, we consider a situation in which a specific screening test (here for the presence of bowel cancer) is applied several times and we focus on the count of times the test has been positive for each subject. A method is suggested to estimate the number of individuals with cancer but having a negative screening test result at all times.

In contrast to previously used methodology which focuses on modeling a zero-truncated count distribution using the zero-truncated binomial, the zero-truncated beta-binomial model, or a zero-truncated nonparametric binomial mixture, the suggested technique is based upon a linear model-approach for the ratios of neighboring probabilities of the number of times the test has been positive. As will be argued and demonstrated in the contribution the regression technique offers more flexibility than parametric count distributions as well as avoiding identifiability problems occurring in the nonparametric mixture approach. The technique is applied to publicly available data of a screening study on bowel cancer in Sydney and it is demonstrated that the method provides realistic estimates of the number of missed cancer cases and performs superior to other available techniques.

Some key words: capture-recapture, screening with partial verification of disease status, ratio estimator, zero-truncated model

References

1. Böhning, D., Baksh, M.F., Lerdsuwansri, R., Gallagher, J. (2011). Use of the Ratio Plot in Capture-Recapture Estimation. *Journal of Computational and Graphical Statistics* (to appear).
2. Rocchetti, I., Bunge, J., and Böhning, D. (2011). Population Size Estimation Based upon Ratios of Recapture Probabilities. *Annals of Applied Statistics* **5**, 1512-1533.