

## Partially Identified Prevalence Estimation under Misclassification using the Kappa-Coefficient

**Anne Kunz**

Munich Cancer Registry  
IBE, Biometry and Bioinformatics  
University of Munich (LMU)  
anne.kunz@ibe.med.uni-muenchen

**Thomas Augustin**

Department of Statistics  
University of Munich (LMU)  
thomas@stat.uni-muenchen.de

**Helmut Küchenhoff**

Department of Statistics  
University of Munich (LMU)  
kuechenhoff@stat.uni-muenchen.de

### Abstract

Many empirical studies are confronted with severe misclassification error. While interest is in the realizations of a variable  $Y$  (e.g. disease yes/no), only a related variable  $Y^*$  (e.g. the outcome of a diagnostic test) can be observed. Even in the case of simple prevalence estimation, naive estimation, neglecting the fundamental difference between  $Y$  and  $Y^*$ , may be considerably biased ([1], [2]).

If sensitivity and specificity of the measurement instrument, and therefore the misclassification probabilities, are known then powerful correction method granting consistent estimation exist. On the other hand, without any knowledge on the misclassification process, the results simply remain vacuous.

We present a corrected prevalence estimator when some knowledge on the misclassification process is available from repeated measurements. In that setting we consider the situation, where, as is quite common in medical studies, kappa, the coefficient of inter-rater agreement is given. Then the problem is still unidentified from the viewpoint of classical statistics, but relying on the concepts of *partial identification* and *systematic sensitivity analysis* in the spirit of [3] and [4] gives nevertheless valuable insight into prevalence estimation. We derive tight identification regions as interval-valued point estimators for the true prevalence, and then derive appropriate confidence regions relying on arguments from [4] and [5].

Our results are illustrated by several theoretical scenarios and with data from an oral health study on prevalence of caries in children in Belgium.

**Keywords.** Partial identification, prevalence estimation, kappa-coefficient, confidence intervals, inter-rater agreement, misclassification.

### References

- [1] Carroll, R. J., Ruppert, D., Stefanski, L. A. and Crainiceanu, C.M. *Measurement Error in Nonlinear Models. A Modern Perspective*. Chapman & Hall, Boca Raton. 2nd edition, 2006.
- [2] Gustafson P., *Measurement Error and Misclassification in Statistics and Epidemiology: Impacts and Bayesian Adjustments*, CRC Press, 2004.
- [3] Manski, C.F. *Partial Identification of Probability Distributions*. Springer, New York, 2003.
- [4] Vansteelandt, S., Goetghebeur, E., Kenward, M.G. and Molenberghs, G. Ignorance and Uncertainty Regions as Inferential Tools in a Sensitivity Analysis, *Statistica Sinica*, 16: 953–979, 2006.
- [5] Stoye, J. More on Confidence Regions for Partially Identified Parameters. To appear in: *Econometrica*, 2009.