

SYSTEMATIC HANDLING OF MISSING DATA IN COMPLEX STUDY DESIGNS - EXPERIENCES FROM THE HEALTH 2000 AND 2011 SURVEYS

Tommi Härkänen

National Institute for Health and Welfare, Finland, tommi.harkanen@thl.fi

Juha Karvanen

University of Jyväskylä, Finland, juha.karvanen@jyu.fi

Hanna Tolonen

National Institute for Health and Welfare, Finland, hanna.tolonen@thl.fi

Risto Lehtonen

University of Helsinki, Finland, risto.lehtonen@helsinki.fi

Kari Djerf

Statistics Finland, Finland, kari.djerf@stat.fi

Teppo Juntunen

National Institute for Health and Welfare, Finland, teppo.juntunen@thl.fi

Seppo Koskinen

National Institute for Health and Welfare, Finland, seppo.koskinen@thl.fi

We present a systematic approach to the practical and comprehensive handling of missing data motivated by our experiences of analyzing longitudinal survey data. We consider the Health 2000 and 2011 Surveys (BRIF8901) where increased non-response and non-participation from 2000 to 2011 was a major issue. The model assumptions involved in the complex sampling design, repeated measurements design, non-participation mechanisms and associations are presented graphically using methodology previously defined as a causal model with design i.e. a functional causal model extended with the study design. This tool forces the statistician to make the study design and the missing data mechanism explicit. Using the systematic approach, the sampling probabilities and the participation probabilities can be considered separately. This is beneficial when the performance of missing data methods are to be compared. Using data from Health 2000 and 2011 Surveys and from national registries, it was found that multiple imputation removed almost all differences between full sample and estimated prevalences. The inverse probability weighting removed more than half and the doubly robust method 60% of the differences. These findings are encouraging since decreasing participation rates are a major problem in population surveys worldwide.