FLEXIBLE TWO-PHASE STUDIES IN EPIDEMIOLOGY

<u>P.Wild</u>^{$\dagger 1$}, W. Schill²

¹National Institute for Research and Safety, Vandoeuvre, France; ²Bremen Institute for Prevention Research and Social Medicine, Bremen, Germany

[†]E-mail: *pascal.wild@inrs.fr*

In standard two phase studies, as introduced in epidemiology by Breslow and Cain (1988), samples of diseased cases and non-diseased controls are drawn at random from the population at risk and some exposure variables are measured. In a second phase, after stratification on available exposure information and the disease status, stratum-specific subsamples are drawn in which a supplemental exposure or confounder information is collected. The numbers drawn from each stratum will be chosen in order to oversample sparse cells and thus increase the power. We propose to adapt this design by fixing a priori phase two stratum sizes from which the complete information is assessed and sampling the population at risk, while recording the strata information, until the desired phase two sample sizes are met. The statistical analysis of such data can be done using ML methods for two-phase data since binomial and negative binomial distribution correspond to the same likelihood. Two practical examples are presented :(1) a study focused on estimating a gene-environment interaction in which a given number of subjects with a familial disease history will be included and (2) a population-based case-control study focused on bladder cancer and exposure to PAH in which a given number of potentially exposed blue-collar controls are included. Asymptotic variance computations show that the gain in variance can be considerable under certain hypotheses.