A PLANNING-TOOL FOR LOGISTIC TWO-PHASE STUDIES

<u>W. SCHILL^{$\dagger 1$}</u>, P. Wild², I. Pigeot¹

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

 † E-mail: schill@bips.uni-bremen.de

Epidemiologic two-phase studies collect data in two phases. Phase one has information on disease status and (minimal) information on exposure. Phase two uses this information to recruit a subsample, stratified on disease status and exposure, for further data collection to obtain more detailed exposure/confounder information.

This talk illustrates the planning of such a case-control study. To this end, we developed a software that enables the organization and data-entry of the relevant entities, such as distributional parameters of the categorical variables and assumed odds ratios. The program computes the expected distributions of a two-phase study and outputs them to an Excelsheet.

We investigate the power of such a study. Is there, given the planned phase-one sample size and given a certain recruitment for the phase-two data, enough power to detect a relevant odds ratio? This and other questions can be answered by transferring the Excel-sheet to a statistics package like SAS or Stata, where two-phase data of the desired size can be generated and analyzed. The necessary software is also made available.

The procedure is demonstrated with two examples: a) planning of a study in occupational epidemiology, where phase one has a proxy of the relevant exposure and a confounder; b) planning of a GEI-study, where phase one has data on the environmental factor (E) and a surrogate of the gene.