

USE OF THE TWO-STAGE METHODS IN INDIVIDUAL-LEVEL POOLED DATA ANALYSES: ARE THEY ALWAYS THE RIGHT SOLUTION?

Valeria Edefonti¹ and Monica Ferraroni¹

¹*Department of Clinical Sciences and Community Health, University of Milan, Milan, Italy*

In pooled analyses, where covariates may not be uniformly defined and coded across studies, and occasionally not measured in all of them, a joint model is often not feasible, and a two-stage method (see [1] for instance) is a simple, valid and practical method for the analysis, lending itself to flexibility with respect to differences in design, confounders and data collection across studies and to a better control for confounding. Simulations indicate that when the individual studies are large, two-stage methods produce nearly unbiased exposure estimates and standard errors of the exposure estimates from a generalized linear mixed model (GLMM).

Based on these considerations, several existing cancer epidemiology consortia suggested in their protocols to apply two-stage methods instead of fitting GLMMs directly on the overall sample (see [2] for instance), although they still struggle to harmonize exposure variables of interest and potential confounders. However, it is unclear how well the two-stage method would perform if individual studies were smaller, especially when there are a few of them [3]. This may be a critical issue, especially as far as evidence has been accumulating on the major research questions a consortium was born for, and time is mature for more specific analyses on subgroups of studies.

The International Head and Neck Cancer Epidemiology (INHANCE) consortium was established in 2004 to contribute elucidating the aetiology of head and neck cancer by providing opportunities for pooled analyses of individual-level data on a large scale. In the current version, the consortium included 35 case-control studies, with questionnaire data on over 26,000 cases and 34,000 controls [4]. Only 10 studies in the consortium provided detailed information on nutrient intakes. The available studies differ in terms of number of subjects included, geographical region, assessment of dietary information and food composition tables.

This naturally provides a real-life situation of interest to understand how big is the difference, if any, in the effect estimates derived from the two-stage methods and the GLMM. In detail, we discuss this point in an application on vitamin C intake and head and neck cancer, after a preprocessing step improving comparability across studies.

This research question may have important fallouts in the refinement of the statistical methods supporting the recent global effort towards pooling individual-level data in consortia.

REFERENCES

[1] DerSimonian R, Laird, N, Meta-analysis in clinical trials. *Controlled Clinical Trials* 1986;7:177–188.

[2] Raimondi S, Gandini S et al., Melanocortin-1 receptor, skin cancer and phenotypic characteristics (M-SKIP) project: study design and methods for pooling results of genetic epidemiological studies. *BMC Med Res Methodol.* 2012;12:116.

[3] Stukel TA, Demidenko E, et al., Two-stage methods for the analysis of pooled data. *Stat Med* 2001;20(14):2115–2120.

[4] Conway DI, et al., Enhancing epidemiologic research on head and neck cancer: INHANCE - The international head and neck cancer epidemiology consortium. *Oral Oncol.* 2009;45(9):743-6.