

## Statistical considerations for analysing data derived from longitudinal cohort studies

*Rocío Fernández-Iglesias<sup>1</sup>, Pablo Martínez-Cambor<sup>2</sup>, Ana Fernández-Somoano<sup>3</sup>*

<sup>1</sup>rocio.fdez.iglesias@gmail.com; Spanish Consortium for Research on Epidemiology and Public Health (CIBERESP), Madrid, Spain; University Institute of Oncology of the Principality of Asturias (IUOPA), University of Oviedo, Asturias, Spain; Health Research Institute of the Principality of Asturias (ISPA), Asturias, Spain.

<sup>2</sup>pablo.martinez-cambor@hitchcock.org; Biomedical Data Science Department, Geisel School of Medicine at Dartmouth, Lebanon, NH, USA; Faculty of Health Sciences, Universidad Autónoma de Chile, Providencia, Chile.

<sup>3</sup>fernandezsana@uniovi.es; Spanish Consortium for Research on Epidemiology and Public Health (CIBERESP), Madrid, Spain; University Institute of Oncology of the Principality of Asturias (IUOPA), University of Oviedo, Asturias, Spain; Health Research Institute of the Principality of Asturias (ISPA), Asturias, Spain.

Modern science is frequently based on the exploitation of large volumes of information storage in data sets, and involving complex computational architectures. The statistical analyses of these datasets have to cope with specific challenges, and frequently involve making more or less arbitrary decisions. Epidemiological papers have to be concise and use to be focused on the underlying clinical or epidemiological results, not paying too much attention to some relevant methodological details. In this work, we use an analysis of the cardiovascular-related measures tracking in 4-8 years children using data from the INMA-Asturias project for illustrating how the making-decision process was performed. We focus on two particular aspects of the problem, how to deal with missing data and which regression model to apply to evaluate tracking when there are no defined thresholds to categorize variables into risk groups. As spoiler, we discuss what were the results of our multiple imputation analyses and which were the advantage of using quantile regression models to analyse the tracking.

**Keywords:** Missing data; quantile regression, tracking.