

Dr Gunky Kim
Centre for Statistical and Survey Methodology
University of Wollongong

Title: Regression Analysis Using Longitudinally Linked Data

Abstract: Probabilistic data linkage is an attractive data collection option when direct measurement is impossible or extremely costly. One important application is where different data sets relating to the same individuals at different points in time are linked to provide a 'synthetic' longitudinal data record for each individual. However, if the same unique identifier is not available in each of the linked data sets, there is always the possibility that linkage errors in the merged data could lead to such a longitudinal record being actually made up of data items from different individuals. This in turn could lead to bias and loss of efficiency in regression modelling using the linked data. Recent results on unbiased regression inference using longitudinally linked data in the presence of probabilistic linkage errors are described in this paper. They build on the inference framework described in Chambers (2009), and focus on the situation where the linked data are obtained by linking three separate data sources via two possibly dependent linkage operations. For example, these could represent different registers for the same population at different points in time or they could correspond to the situation where a survey sample is linked to two separate registers, one contemporaneous with the survey and the other containing historical information. In the first case one needs to adjust regression modelling to account for linkage errors as well as errors arising from incomplete linkage, while in the second case there is also the important issue of accounting for the impact of the complex survey design when using the linked survey data in regression modelling. Both these scenarios are considered here, and simulation-based results illustrating the gains from taking account of possible linkage errors in regression modelling using probabilistically linked longitudinal data are presented.