

Paul A. Smith's contribution to the Discussion of "A system of population estimates compiled from administrative data only" by John Dunne and Li-Chun Zhang

Paul A. Smith, University of Southampton, UK. p.a.smith@soton.ac.uk

I congratulate the authors on an interesting paper presenting a new toolkit of approaches for making population estimates from administrative data, and have some questions under three headings.

Generality

I wonder how generally applicable the methods are to any situation with administrative data only. The existence of the PPSN seems a boon in Ireland, an *accurate* implied population "spine" to which sources can be linked. If a spine must be constructed from linked administrative datasets without a common identifier, it risks especially overcoverage, and may make the pure undercoverage scenario, which works so well in Ireland, unrealistic.

I note that the population concept is based on a year of interactions. But it could be based on a different period. How narrowly (in time) could PECADO tools be focussed?

PECADO in Ireland

The paper reports that signs of life are *strong* evidence of residence, but I would like some supporting evidence. I can imagine people living abroad but paying tax in Ireland, or having pension paid to an Irish bank account. Are these kinds of cases identifiable? Do you have gold standard data to assess what proportion of cases with this evidence are truly resident? Similarly I wonder whether people living abroad nevertheless manage to renew Irish driving licenses using (perhaps) their parents address.

The population concept uses a 20 weeks work threshold for being counted; what prompts this choice? Is it a threshold in the administrative system, or just an arbitrary number? I also suspect that PECADO does not easily handle cross-border workers who may interact with administrative systems but not be resident, or may (in a wider system of administrative data censuses) meet the residence criteria in more than one country.

Methodology

The selective trimming for overcoverage decreases the bias but increases the variance of the population size estimates. Although I suspect that this is an advantage, it would be nice to see some evaluation of the mean squared error to show that you have a good bias-variance trade-off.

I was interested that you are able to use the QNHS as a coverage survey with apparently reasonable results. The usual assessment is that differential nonresponse in social surveys makes them unsuitable as a coverage survey, though in principle a poststratification by variables which explain the nonresponse may make such estimation practical. Can you say more about the quality of the estimates using a social survey as the coverage survey?