

## Title: Transforming population statistics: From census to fractional counting

Conducting a census is the oldest method of producing *detailed* population statistics. However, the high cost and--hence--low frequency of census has become increasingly difficult to justify, so that many countries are currently developing alternative ways to produce *census-like* population statistics annually. Here we shall focus on the problem of *estimating* population counts (with basic demographics) at detailed locations.

In the first part of the talk, we summarise the two alternative frameworks for producing census-like population statistics, where direct counting from the Central Population Register is either unavailable or deemed unreliable. First, the *Register Survey* approach is based on combining available population registers and *large* coverage surveys. Replacing the census enumeration list by an integrated register compiled from existing data is the key to cost reduction, as is the case in Israel 2008 and 2021. The second *Fractional Counting* approach aims to estimate the population counts directly in a manner that can be characterised as register-based. Starting from an extended population register (EPR) with negligible under-coverage errors, each EPR person is assigned, successively, a probability of belonging to the target population, such as practised in Estonia and Latvia, and a vector of probabilities of living at one of the known addresses. Not only can it further reduce the cost, but the framework enables a conceptual shift away from population statistics envisaged as the results of pigeonhole classification. For instance, instead of 'how many people have a permanent address at a given place', one can make statistics about 'how many people can be expected at the given place and, in addition, where else they can be expected'.

For any statistics that require model-based methods of estimation, whether the model is trained real-time or fixed in advance, there is a question of valid uncertainty assessment. In the second part of the talk, we provide an introduction to *design-based inference methods for model-based statistics*. A relevant starting point in the present context is that the register-based census-like statistics have sufficient quality, such that they could *replace* sample survey or census altogether in producing statistics, as is the case in the Nordic and Baltic countries. Nevertheless, the register-based statistics are not without errors, which motivates the central concept of *audit sampling inference* (Zhang, 2021), "Wherever the goal of survey sampling is to produce a point estimate of some target parameter of a given finite population, audit sampling aims not to estimate the target parameter itself, but some chosen accuracy measure of any given estimator of the target parameter, which may be potentially biased due to failure of the underlying assumptions or other favourable conditions that are necessary." In particular, we shall explain how the *total and individual errors* of either completely register-based statistics or sample-based model estimators can be evaluated with respect to the known audit sampling distribution.