Net–Migration in the Human Mortality Database and related databases

a proposal for the 5th HMD Symposium, May 13–14, 2019, Berlin

Jim Oeppen*[1]

[1]Interdisciplinary Centre on Population Dynamics, University of Southern Denmark

February 15, 2019

1 Population Estimation and Intercensal Survival in the HMD

For many countries in the Human Mortality Database (HMD), the population at risk to die is derived from periodic censuses, rather than from a continuous population register. In the English–speaking countries, these are typically 10 years apart and international migration is poorly recorded. Annual, inter–censal, age–specific cohort populations are estimated by subtracting the cumulation of inter-censal deaths from each annual cohort in a starting census, or new inter–censal birth cohort, and comparing the result with the same cohort in the next census. Any inconsistency, which may be due to net–migration and/or data errors, is removed by treating the quantity as net–migration in the population balancing equation.

As the HMD protocol points out: 'This inconsistency is caused by two factors: migration and error. Although both of these factors tend to be small relative to cohort size (at least for national populations), as a matter of principle they should not be ignored. The standard method consists of distributing implied migration/error uniformly over the parallelogram shown in Figure 3a. Then, estimates of cohort size for intercensal years are found by subtracting, from the initial census count, both the observed death counts and an estimate of net migration/error.' Figure 3b in the Protocol shows the similar procedure for a new inter–censal birth cohort.

One problem with this process concerns the population estimates projected for the years since the last census: the 'post–censal' period when no closing census is available. These figures are of the most interest, but may be heavily revised when a new census becomes available. It can be shown that for a country like the United States, the HMD method creates cohort artefacts that revert to age–effects when the next census becomes available.

Although the Lexis surface of 'net migration/error' for each country and sex is estimated by the HMD, these estimates are not provided to users in a convenient format. This paper speculates about the uses that could be made of these data.

2 Assessing the possibilities of Error

Discontinuities in the Lexis surface of net–migration by age, period, cohort, or sex may indicate errors. This issue will be explored with three–dimensional visualisation methods.

*email: joeppen@health.sdu.dk

3 The uniform distribution assumption

Is it reasonable to assume that each cohort's net–migration is independent, and uniformly distributed over the inter-censal interval? As an extreme example, imagine a cohort aged 10 in the initial census and aged 20 in the next. The evidence collected by Rogers and Castro (1981) as the basis for their migration model suggests that sharp changes in migration rates may occur over this age–range. The net effect in national populations may be small, but there has been a recent growth in sub–national databases of the HMD type: e.g Japan, Spain, Canada, and the USA. It is clear in the Japanese case that there are strong artefacts that appear to be associated with migration for Prefectures that are city–dominated. It may also be that the imprecision of the exposure estimates for young male adults may distort the interpretation of the accident hump in mortality.

The talk will explore two possibilities for non–uniform distribution of the net–migration balance: first, a non–parametric approach based on the work of Rizzi *et al.* (2015), and second, a parametric approach based on the Rogers and Castro (1981) model.

4 Prospects: the HMD as a source of information for migration modelling

If the signal–to–noise ratio is high in the net–migration estimates derived from the HMD and its related databases, they could be used to improve the modelling of migration in population projections. They could also be compared, and possibly integrated, with the results from individual–based surveys which are necessarily based on small samples.