

International Journal of Population Data Science

Journal Website: www.ijpds.org



The Illusion of an Automated German Register Census

Rainer Schnell¹ and Severin V. Weiland¹

¹University of Duisburg-Essen, Duisburg, Germany

Currently, the German Federal Statistical Office is working on the transition from a register-enhanced to a fully register-based census. Deduplicating and linking multiple registers using record-linkage methods for census purposes is challenging because no universal national identifier is available. To evaluate the expected results, we simulated every procedural step of the register census. The simulation is based on a synthetic base population of about 80 million records that reflects the joint distribution of personal attributes such as household structure, region, and age. Furthermore, we conditioned the distributions of last names and first names on age, nationality, sex, and region to account for cohort, ethnic, and regional variation. All previous studies neglected these dependencies, leading to an overestimation of the performance of automatic record-linkage

methods. The most important databases of the census are the about 9500 resident registers, which may contain duplicate records. Therefore, we estimate the deduplication rate of the resident registers by adding noise (errors in identifiers and population dynamics, such as regional mobility or demographic processes) based on previous research on errors in administrative data, and then deduplicating the registers using various record-linkage methods, such as probabilistic matching and multiple matchkeys. The results indicate that only a small percentage of the duplicates can be detected automatically. This fact is due to the limited number of available identifiers, not to algorithmic problems. Given the current set of identifiers, a register-based census will require an unreasonable additional amount of clerical work.

