

English summary

Data quality has long been a contentious issue in population studies. Even today, contemporary data may present real challenges depending on the phenomena under study, especially in those regions with poor data registration. With historical databases the problems of data quality increase, given that such data sets are usually unique sources that cannot be improved with the collection of further data or sample checks.

In this context, the main contribution of this dissertation is the development of methods to analyse demographic data that, due to their complexity or incompleteness, could not have been analysed with other more standard tools. Four topics are addressed, one for each manuscript: (1) record linkage; (2) Bayesian inference; (3) mathematical demography; and (4) agent-based modelling. The thesis provides an overview of the theoretical foundations behind each of these four methodological approaches, and develops specific applications for the solution of particular data problems.

Manuscript [I](#) presents a record linkage algorithm for the reconstruction of individual lifespans using historical marriage records of Barcelona (Catalonia, Spain) from the sixteenth and seventeenth centuries. Given the absence of unique identifiers, the procedure uses nominal information such as first names and surnames to identify individuals across several marriage records. The structure of these reconstructed lifespans is the motivation underlying the Bayesian model developed in Manuscript [II](#) for the estimation of adult age-specific mortality and life expectancy from data with unknown ages. The model is based on an R package that was originally designed to study the survival of wild animals with unknown ages, which represents a novel approach to the analysis of human historical populations. Using simulated data, the results show that, despite the lack of information, the model is able to appropriately estimate the mortality parameters and the remaining life expectancy above age 15, as well as the corresponding survival curves. Manuscript [III](#)

formally proves the existence of a symmetry between the age composition and the structure of the remaining lifetimes in finite stationary populations. This interesting property can be used to estimate the age structure of a population with unknown ages, when individuals are followed until death and the assumption of stationarity is acceptable. Finally, Manuscript [IV](#) introduces an agent-based model of sex ratio at birth distortions that has been applied to South Korea and India. The aim is to analyse the dynamics of son preference, technology diffusion, and fertility decline underlying the emergence of distorted sex ratios at birth, and for which empirical data are not available.