Authors' reply to the Discussion of 'A system of population estimates compiled from administrative data only' by John Dunne and Li-Chun Zhang

John Dunne
Central Statistics Office (CSO), Cork, Ireland.
E-mail: John.Dunne@cso.ie
Li-chun Zhang
University of Southampton, Southampton, United Kingdom.

We are grateful to all those who submitted comments on the paper. Their comments give us an opportunity to revisit some of the key points in our paper, as well as to elaborate on some important issues which we could not cover in sufficient detail in the paper.

Specific assumptions and general applicability

The three assumptions related to the potential DSE bias are concerned with linkage errors, erroneous records, and heterogeneous list capture probabilities, as commented by Baffour, Chai, Eleanor, Shlomo and Smith. While the first two types of error seem to have their own distinct causes, heterogeneous capture probabilities may be confounded with list dependence. Nevertheless, since the consequence of any list dependence is just heterogeneous capture probabilities, one only needs to focus on whether the capture probability of at least one list is sufficiently constant *given* the relevant covariates, such as by post-stratification, rather than deliberating whether independence can or cannot be assumed for different register sources. Diagnostics and tests for homogeneous B-list capture probability are useful tools in practice.

In reality, where question marks cannot be entirely eradicated for any of the error sources, there are only *two* logical steps one can take.

The first step is to define a strategy, by which some types of error can be reduced to a negligible extent compared to the remaining ones. In the PECADO system, decisions are made to make A-list undercounting the predominant problem that remains, which can be solved by the traditional wisdom of DSE implementation. Whereas the system described by Bernardi et al. sets up EPR overcounting as the main issue to be solved. This shows how different practical approaches are possible following the same strategy of 'reduce-and-conquer', given the data situations in different countries.

The second step is to develop adequate methods to support the chosen strategy. For instance, trimming source-registers instead of individual records is found to be useful for the PECADO system. But it may not be feasible if the statistical agency has only limited access to various signs-of-life data sources. Chipperfield et al. (2023) investigate capture-recapture estimation given both non-negligible linkage errors and erroneous records. In particular, a record-level score for trimming can be modelled on linking census and administrative spine, such that the trimmed spine containing negligible erroneous records can be used as a list for capture-recapture estimation.

Methodological extensions

Several needs and possibilities for methodological extensions are mentioned by Abbott, Allin, Eleanor, Henry and O'Farrell, Pfeffermann and Shlomo, which pertain to potential uses of more sophisticated models, disaggregating statistics to more detailed levels including special subpopulations, and producing attribute statistics by similar methods.

First, multiple system estimation (MuSE) based on 2+lists are possible using log-linear models or latent variable models. A key motivation traditionally is to accommodate heterogeneous capture probabilities through the list-interaction terms, which are possible to introduce as the degree-of-freedom increases with the number of lists. More recently, the models have been extended to allow for erroneous list enumerations in addition; see e.g. Zhang (2015, 2019), Di Cecco et al. (2018), Ballerini (2021). The implementation of such models can however raise additional challenges. For one thing, model selection which is hardly an issue given only 2 lists becomes more critical as the number of lists increases, not least due to the non-identification problem inherent of capture-recapture data; see e.g. Fienberg (1972), Zhang (2019, 2023), Aleshin-Guendel et al. (2023). Moreover, MuSE can aggravate linkage errors in the absence of a unique key for matching. The MuSE models are nevertheless useful since they can provide evidence that suitable DSE methods may be good enough for practical purposes, even when the MuSe models cannot be established as viable alternatives themselves.

Pfeffermann rightly pointed out that trimming is also possible when both the lists contain erroneous records. The setting was considered by Zhang and Dunne (2018, Chapter 17.2.4). It can be shown that it is possible for the naïve DSE to be unbiased despite the erroneous records and, otherwise, trimming both the lists can reduce the DSE bias under certain conditions. This echos the discussion of MuSE models in the sense that a more sophisticated method may be beneficial although it can have its own implementation costs.

Next, disaggregation is always desirable by geography, demography or other attributes such as activity status or education level, whereas any attribute statistics can equally be viewed as population size statistics where the target population has the specific attributes, such as those who are employed and have the highest level of education. Insofar as the data can be organised as a contingency table, the MuSE models should remain applicable.

However, an extra challenge arises if the variables needed for disaggregation are missing in a list or subject to measurement errors. Van der Heijden et al. (2022) study the estimation of Māori population in New Zealand, where Māori status may differ between the lists for the same individual and the ethnicity status may be missing for some people. We mentioned Calibrated DSE as a methodological extension for disaggregation, given a person's registered locality may be subject to measurement errors. While this can work at the geographical county level (altogether 26 in Ireland), we plan to investigate a different approach for the more detailed levels, whereby individuals are 'anchored' via households and integration with the household statistics is handled at the same time.

Abbott mentioned the special populations, such as those living in managed accommodation. While formally a matter of disaggregation or 'attribution', it may be difficult to handle a special population by the general method if the required information is altogether lacking or heavily under-recorded. Take an example from Norway, where it is not mandatory for an elderly person residing at a care home to 'move' her/his address in the Population Register away from the spouse living at their private address. For register-based population statistics of such persons, data will need to be collected from the relevant institutions or municipalities directly.

Error evaluation

Error evaluation is fundamental to the quality of official statistics. Abbott, Bernardi et al., Chai, Eleanor, Henry and O'Farrell, Smith and Shlomo have all touched on many relevant aspects, where the perspective differs for (i) quality improvement and assurance, and (ii) end-outputs.

The total-error approach has a longstanding tradition and a large literature in official statistics. As Zhang (2018) emphasises in the discussion of Hand (2018), combining multiple data sources is generally required when using administrative data for making statistics directly. A two-phase total-error model for data integration has been developed, see Zhang (2012), Reid et al. (2017) and Rocci et al. (2022), which covers all the potential problems in the list data for capture-recapture estimation. The total-error perspective requires one to analyse and identify the most important error sources in a given context, in order to improve the source data and the processing steps accordingly.

When it comes to the end-estimates, it is clear that one cannot establish the validity of any model-based estimators of a *given finite population* only based on the data that have been used for estimation. External data are required. Our application of combing an administrative list and the census enumerations illustrates as well how the errors of the PECADO estimates can be audited using a census survey, although we mentioned that ideally a smaller sample survey would suffice. Bernardi et al. have related in their comments the pioneering development of an audit survey for the register-based population statistics in Italy. As Zhang (2021) puts it, "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."

To take a historic example in the present context, the US Census Bureau has published census coverage error estimates for over four decades, using a post-enumeration survey (PES) and implementing the DSE methodology by combining the census and the PES. The decisions not to adjust the 1980 and 1990 censuses, i.e. not to accept the DSE results as the official population estimates, were upheld by the legal system. In January 1999, the U.S. Supreme Court ruled that the DSE adjusted numbers may not be used for the allocation of congressional seats to states. In retrospect, this finally sealed the PES as audit sampling for the given census population estimates, rather than survey sampling for producing the population statistics directly.

Audit sampling inference, not quite a census as suggested by Chind, should provide the necessary assurance whether the PECADO population estimates can be fit-for-purpose over time.

Finally, we fully agree with Allin, Shlomo and Woolford on the related issues mentioned in their comments, such as collaborating on the international standard of population concepts, harmonising the estimates of stocks and flows, and engaging with the users for their acceptance and satisfaction.

References

Aleshin-Guendel, S., Sadinle, M. and Wakefield, J. (2023). The Central Role of the Identifying Assumption in Population Size Estimation (with discussions). Biometrics, to appear. `arXiv:2101.09304v2`

Ballerini, V. (2021). *The Fisher's Noncentral Hypergeometric distribution and population size estimation problems*. Unpublished PhD Thesis.

Chipperfield, J., Zhang, L.-C. and Baffour, B. (2023). Robust Statistical Estimation for Capture-Recapture using Administrative Data. *submitted*.

Di Cecco, D., M. Di Zio, D. Filipponi, and I. Rocchetti (2018). Population Size Estimation Using Multiple Incomplete Lists with Overcoverage. *Journal of Official Statistics*, 34: 557-572.

Fienberg, S.E. (1972). The multiple recapture census for closed populations and incomplete 2^k contingency tables. *Biometrika*, 59:591-603.

Van der Heijden, P.G.M., Cruyff, M., Smith, P.A., Bycroft, C., Graham, P. and Matheson-Dunning, N. (2022). Multiple System Estimation using Covariates having Missing Values and Measurement Error: Estimating the Size of the Māori Population in New Zealand. *Journal of the Royal Statistical Society, Series A*, 185:156-177.

Reid, G., Zabala, F. and Holmberg, A. (2017). Extending TSE to administrative data: A quality framework and case studies from Stats NZ. *Journal of Official Statistics*, vol. **35**, pp. 477-511.

Rocci, F., Varriale, R. and Luzi, O. (2022). Total Process Error: An Approach for Assessing and Monitoring the Quality of Multisource Processes. *Journal of Official Statistics*, 38:533-556. `10.2478/jos-2022-0025`

Zhang, L.-C. (2023). Discussion on "The central role of the identifying assumption in population size estimation" by Serge Aleshin-Guendel, Mauricio Sadinle, and Jon Wakefield. Biometrics, to appear.

Zhang, L.-C. (2021). Proxy expenditure weights for Consumer Price Index: Audit sampling inference for big-data statistics. *Journal of the Royal Statistical Society, Series A*, **184**, 571-588.

Zhang, L.-C. (2019). Log-linear models of erroneous list data. In *Analysis of Integrated Data*, eds. L.-C. Zhang and R.L. Chambers. Chapter 9, pp. 197-218. Chapman & Hall/CRC.

Zhang, L.-C. (2018). Discussion of "Statistical challenges of administrative and transaction data" by D. Hand. *Journal of the Royal Statistical Society, Series A*, **181**, 579-580.

Zhang, L.-C. (2015). On modelling register coverage errors. *Journal of Official Statistics*, 31:381-396.

Zhang, L.-C. (2012). Topics of statistical theory for register-based statistics and data integration. *Statistica Neerlandica*, vol. **66**, pp. 41-63.