Discussion of *The Central Role of the Identifying Assumption in Population Size Estimation*

by Aleshin-Guendel, Sadinle and Wakefield

John Whitehead, Lancaster University, UK

The authors of this paper are to be congratulated on an impressive piece of work. Their evaluation of the MSE method for estimating the size of a population that is only partially observed goes to the heart of the matter: no estimation is possible without making a non-testable assumption about the nature of the data. Their meticulous examination of the process of inference highlights the careful distinction that must be made between this fundamental assumption, referred to as the *identifying assumption*, and the imposition of further testable assumptions that may follow during model building. The summary of the paper contains their key messages that "models with different identifying assumptions can produce arbitrarily different population size estimates" and consequently that "If an appropriate identifying assumption cannot be found for a data set, no estimate of the population size should be produced".

An identifying assumption that is commonly used in practice is that there is no highest-order interaction (NHOI). One natural way in which statisticians can understand this arises from representation of the data by a log-linear model in which counts of individuals appearing on each possible combination of lists are related to a sum of terms with an intercept, main effects for each list, two-way interactions between any two lists, three-way interactions between any three lists, and so on. In fitting linear models to quantitative (normally distributed) data, binary data or survival data (amongst others), it is widely accepted and uncontroversial to neglect higher-order interaction terms. The wisdom of such neglect can usually be assessed using the data available. Furthermore, the nature of such assumptions can generally be explained to and discussed with subject-area specialists even if they are unfamiliar with the technicalities of linear modelling. In MSE, the validity of the NHOI assumption cannot be assessed from the data *at all*, nor can it be easily explained to subject-area specialists who will consequently be unable to judge whether it is reasonable or to appreciate the limitations it imposes on the analysis.

In the Web Appendix E1, the authors explain that for 4 lists (as in the case of the Kosovo data), the NHOI assumption amounts to accepting that "the highest-order interaction for the first three lists, conditional on not being observed in list 4 ... is equal to the highest-order interaction for the first three lists, conditional on being observed in list 4". It is not only difficult for subject area specialists to grasp what this means, many statisticians (including myself) will be baffled by it. As the number of lists increases, the complexity of the assumption grows. Drawing inferences from data is a task to be undertaken by a partnership of subject-area specialists and statisticians, and within that team the assumptions implied by the models being fitted must be understood and accepted. In other contexts, notions such as "increasing the level of nitrogen should have the same effect on yield whatever the accompanying level of potassium", or "the pattern of the responses of patients lost to follow-up would have been the same as those of the patients observed throughout" can generally be communicated, and the statistician can have some confidence that they will be challenged if unreasonable. This is not true for the NHOI assumption, and being non-testable, the opinion of subject-specialists would form the only possible way of justifying it. For Bayesian analyses, the need to communicate effectively with subject-specialists can be even greater. The authors are fortunate enough to have access to data from previous studies of casualties in the Kosovo war and use these to construct an informative prior for their illustrative analysis of the data in Table 1. More often, a Bayesian analysis using an informative prior would have to rely on the prior opinions of the analysis team. It seems doubtful that opinions concerning the obscure parameters that feature in MSE could reliably be elicited.

The authors distinguish between non-testable and testable assumptions. Although this is a valid and important distinction, it must be borne in mind that in this context, "testable" means "testable with extremely low power". Furthermore, the effects of such testing must be allowed for in any subsequent frequentist analysis that is based on a model selected through such testing, perhaps by applying bootstrapping. It must always be remembered how few data points feature in many MSE analyses: the Kosovo data used to illustrate the paper comprise just 15 observations.

As an alternative to making the NHOI assumption, the authors investigate making a marginal NHOI assumption. For example, in the context of the Kosovo example, they provide a justification that the interaction between two of the lists involved, the ABA and HRW lists, might be assumed to be zero. As is made explicit in the Web Appendix B.1.2, frequentist analyses based on this marginal NHOI assumption use only the data from these two lists – data concerning the EXH and OSCE lists do not feature at all. This reduces the number of data points used from 15 to 3 and estimation is based on the simplest of capture-recapture analyses. The argument for assuming that the two-way interaction between the ABA and HTW lists is negligible is clearly presented and persuasive. It is harder to imagine how, for some other data set, the neglect of a three-way interaction might be justified.

The approach recommended by Aleshin-Guendel, Sadinle and Wakefield requires the researchers to have the courage to conclude, if appropriate, that no estimate of population size can be derived. It is likely that this will often be the most reasonable conclusion, especially in applications to hidden human populations as might occur in epidemiology, official statistics or human rights. The responsibility for making such decisions may well fall upon the project statisticians. In many settings, the implications of choosing not to compute an estimate could be substantial.

For cases where a valid analysis can be undertaken, Web Appendix B provides the necessary details. The frequentist approach under the NHOI assumption is straightforward to implement. It can be expressed as follows. Suppose that data from K lists are available, each data point being a count of those individuals who appear on some selection of H of the lists, and are absent from the other (K – H). Define D to be the product of the counts for which H is odd, divided by the product of the counts for which H is even. Let R denote the sum of the reciprocals of all of the observed counts. Denote the total number of observed individuals by n. Then the total population size is estimated by n + D, with standard error given by $\sqrt{D(1 + DR)}$. This is also the approach described in Section 6.5 of Bishop, Fienberg and Holland (1975) – but with different notation. The marginal NHOI assumption leads to the analysis of data from a reduced number of lists, conducted in the same way.

For example, Table 1 shows that 108 people are on lists ABA and HRW, 577 are not on ABA and on HRW, and 1420 on ABA and not on HRW: a total of n = 2105. Hence D = 577 x 1420/108 = 7586 and R = 1/108 + 1/577 + 1/1420 = 0.01170. It follows that N is estimated by 9691 with standard error 825 and 95% confidence interval (8074, 11308). Using the NHOI assumption instead, all of the counts in Table 1 are utilised. Here, n = 4400, D = 12542 and R = 0.2240, leading to a population size estimate of 16942 and a confidence interval of (5304, 28580). Up to rounding errors (in the paper, rounding down is implemented), these results are the same as given in Tables 2 and 3.

These details are made explicit here to underline the fact that most of the frequentist analyses are straightforward to compute, even though that might not be apparent from some alternative mathematical representations. Revealing the simplicity of the calculations can aid communication between the project statisticians and subject-area specialists and avoid what the latter might feel to be the mysterious workings of computer algorithms.

The recipe provided above breaks down when any of the counts involved are zero, as then D is zero, infinite or indeterminate and R is infinite. The simple replacement of zero counts by ½ allows estimation to proceed. Aleshin-Guendel, Sadinle and Wakefield suggest a switch to a Bayesian approach as an alternative way of overcoming this impasse. The "add ½" approach is clearly rough-and-ready, but like an old tool held together by bits of string its shortcomings are clear for everyone to see. There is a danger that a Bayesian analysis emerging from black-box computations might be accorded more credibility than it deserves by subject-area specialists who do not appreciate the equally ad-hoc assumptions made inside the box about features such as prior distributions of obscure parameters.

Consideration of the operational details of conducting MSE analyses is not the focus of this excellent paper. The authors' main warning is that in many cases it should not get that far: there will be no justification for making any identifying assumption that will allow an estimate to be computed. In applications where there might be pressures to find an estimate, and especially when it is clear that sponsors would like to see a large estimate or a small estimate in order to promote a particular message, MSE is a methodology to be applied and interpreted with great caution.