Neville Calleja and C. P. Farrington (*The Open University, Milton Keynes*)

The methods that are described in this paper are ideal for combining data sets arising from various unrelated sources, including both surveys and surveillance data. This methodology could prove useful for combining several historical health surveys and registers with the aim of producing overall estimates of health indicators. In Malta, we are aiming to combine a health examination survey (1984) with two health interview surveys (2002, 2008) and a subsequent health examination survey (2009) to investigate health inequalities. In addition, some detailed tables are available for a similar interview survey that was conducted in 1992. The following equation that was described in the paper provides a useful framework for explaining the discrepancy between self-reported health data, which are cheaper to collect, and those which are obtained following examination of individual patients:

$$M = \sum_{i=1}^{g} N\{\pi_g \rho_g \delta_g + \pi_g \rho_g (1 - \delta_g)\}.$$

There is always an underestimate of prevalence figures in health interview surveys and this may be represented by the proportions δ_g. The term $\pi_g \rho_g \delta_g$ would correspond to the proportion of the population reporting having the disease through a health interview survey. However, there would still be a proportion, $\pi_g \rho_g (1 - \delta_g)$, who are still unaware of their health condition. However, a health examination survey would identify $\pi_g \rho_g$. Therefore the difference between the two estimated rates would supply the proportion vector $1 - \delta_g$ which would allow correction of health interview survey estimates to reflect real disease prevalence, as measured by examination. The categories corresponding to subscripts g would be defined in terms of age–sex strata, usually.

However, a crucial additional variable to account for in our study is time. It is expected that prevalence figures would have changed over a span of 25 years.

Indeed this variable appears to be absent in the present paper, unlike a previous study on toxoplasmosis that was published by some of the same authors. Given the accelerated developments in management of human immunodeficiency virus infection, a parameter should be included to account for temporal changes, even though the studies that were considered were not very spaced out in time. In addition, it is understood that inconsistency parameters have been included to allow for named biases in each type of survey. Perhaps inclusion of further parameters identifying the type of source could have usefully absorbed any interstudy variability that is specific to each type.

In any case, this is a landmark paper for research projects such as mine where a number of *ad hoc* sources of a differing nature are to be pooled to investigate trends in health inequalities.

Sander Greenland (*University of California, Los Angeles*)

Goubar and colleagues expand their models to resolve *apparent* inconsistencies. I emphasize 'apparent' because (as they discuss) inferring inconsistency depends on non-identified constraints. For aetiologic research synthesis, I prefer non-identified starting models, in which each data set receives its own submodel to reflect study idiosyncrasies and has a few exchangeable parameters across the submodels. I wish to avoid models that are shaped by desire for identification or by exchangeability assumptions that are falsified on reading reports closely (as found in popular random-effects models for meta-analysis). The result is a hierarchical model for study parameters regressed on study characteristics (Greenland, 2005), for which identification (and hence learning from the data) hinges on further parameter constraints. Those constraints may reflect judgements about what seems reasonable given other information, or they may be for simplification, with little or no regard for subject matter.

In some descriptions of identifying constraints, there is a peculiarity of statistical culture (echoed in the discussion of this paper): if informed judgements are expressed honestly as fuzzy constraints, using probability distributions to express uncertainty, they are called 'subjective priors' and used with apologies. If instead judgements are expressed as dogmatic point constraints and we proceed as if they are correct, we may call our analysis 'objective'. We receive absolution for our dishonest use of certainty as long as we test the constraints, which we can do only by assuming other, unidentified constraints.

Apparent 'inconsistency of evidence' reflects combinations of

(a) misjudgements about the studies or about the background evidence and
(b) falsity of arbitrary constraints.

These are qualitatively different sources of inconsistency but can be hopelessly entangled, as reflected by the non-identification that is reintroduced when arbitrary constraints are withdrawn. If we prefer risking misjudgement to assuming outright falsehoods, we can replace point constraints with priors that are