A Framework for the Meta-Analysis of Survey Data

Karla Fox, Statistics Canada

Abstract
This research presents a framework for the meta-analysis of complex survey data using the sample design and a superpopulation model to quantitatively summarize different surveys that represent the same underlying population. Additionally, it will compare the classical design based methods using a superpopulation model approach.

Keywords: Meta-analysis, observational data, surveys

1. Introduction

The recent increase in straightforward access to data has lead to a proliferation of analysis that groups or pools multiple study results. This ready access, combined with user-friendly computer software, has lead researchers to try to borrow strength from the many different observational and non-randomized studies in an attempt to either improve precision or address research questions for which the original studies were not designed. Researchers have started to employ many different techniques including meta-analysis; however, the analysis is often done without reference to a generalized framework or a systematic review and often without an understanding of the methodological differences between surveys and experiments.

Although meta-analysis was first used in 1976 for the combination of educational studies, it is now common in the medical literature (Glass). This integration of studies has expanded from combining randomized control trials and other experimental studies to blending information from epidemiological quasi-experimental studies (cohort and case-control), and diagnostic tests and bioassays and now survey data.

This paper begins with an overview of the differences in the randomization processes in experiments and surveys. Next, we describe model and design based inference and estimation in surveys, followed by an overview of fixed and random effect models in meta-analysis. We then describe the simulations used to illustrate meta-analysis in the survey context, highlighting issues with convergence, inference and small sample properties.

2. Randomization: Allocation compared to Selection

It is important to address the differences of the randomization frameworks in experimental design, and that of survey design. What is random in experimental data is the assignment of individuals. With experimental data, we assign a treatment(s) and a control to individuals in a random fashion to create two (or more) probabilistically equivalent groups. Altman (1985) points out, however, that randomization does not guarantee that the treatment groups are comparable with respect to their baseline characteristics, and suggests that comparability should be assessed prior to analysis. We are thus using randomization to reduce confounding and to ensure that the results of the experiment are internally valid.

When we combine experiments we make several assumptions about the treatment effect in the different experiments, based on the fact that we are combing estimates that have been generated from a process with random allocation. We assume that the treatment effect follows a distribution and we either assume that the effects we are combining are identical and independent from this distribution or that the differences between the treatment effects from the different

experiments can be modeled in an analogous fashion to random effect ANOVA models. We make no assumptions on the population of individuals.

On the other hand, in the case of classical design-based survey methods, what is random is the selection of individuals. The purpose of random selection is to generalize to the entire finite population from which we are drawing our sample. Finite population parameters describe the finite population *U*, from which we draw our sample. Estimation is then about characteristics of this population. These quantities always portray the population at a given time point (the time of sampling), and are purely descriptive in nature. A finite population parameter has the property that in a census (with no non-response and no measurement error) its value can be established exactly. If the researcher was able to collect information from the entire population he would do so. However, since the researcher rarely can take a census of his population of interest and probability sampling is done to produce estimates of the finite population parameter of a sufficient precision at a minimum cost.

In probability sampling we select samples that satisfy the following conditions (Sarndal 2003):

1. We can define a distinct set of samples, S_1, S_2, …, S_v, which the sampling procedure is capable of selecting if applied to a specific population. This means we can say precisely which units belong to S_1, to S_2, and so on.
2. Each possible sample S_i has assigned to it a known probability of selection $P(S_i)$, and the probabilities of the possible samples sums to 1.
3. This procedure gives every unit a known non-zero probability of selection.
4. We select one of the S_i by a random process in which each sample has its appropriate probability of selection $P(S_i)$.

3. The Model and the Design

Though classical sampling theory is concerned with inference for finite population parameters we can extend inference to a superpopulation. Superpopulation modeling occurs when the researcher specifies or assumes a random mechanism that generated the finite population (Hartley and Sielken, 1975). We now have what is called model-based as opposed to design-based inference, where the model is the statistical conceptualization of a superpopulation.

For example if we assume that the vector y of responses in a survey are a realization of a random variable $Y = (Y_1, Y_2, ..., Y_N)$, then the finite population vector y is itself a sample from a hypothetical superpopulation model, ξ. Here then the model-builder is not interested in the finite population U at particular time, but the causal system described in the model. So in the case where we have a census we still would not be estimating these model parameters exactly.

Figure 1 illustrates the relationship of the superpopulation with the finite population with the sample. If we let θ be the parameter from a superpopulation model, then we can let θ_N be the estimator from the entire finite population and then $\hat{\theta}^{F(F)}$ is the estimator for the finite population based on the sample.

Figure 1
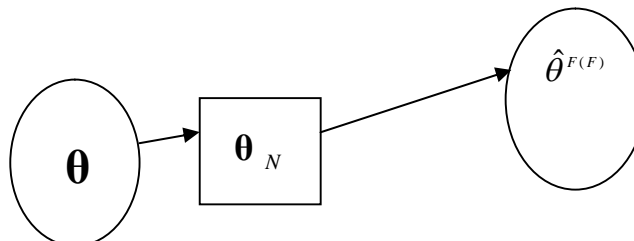
We now note that we can approach estimation in either a finite sampling or a superpopulation framework. If we cross the estimation approach used with the level of inference we have the following table (Hartley and Sielken, 1975):

Estimation Framework

	Sampling from a fixed or finite population	Two-stage sampling from a superpopulation
Parameter of interest		
Finite population parameters or descriptive parameters	Classical finite population sampling $\hat{\theta}^{F(F)}$	Superpopulation theory for the finite population $\hat{\theta}^{F(S)}$
Infinite superpopulation parameters or model parameters	Infeasible	Inference about superpopulation parameters from two-stage sampling $\hat{\theta}^{S(S)}$

For clarity the following notation will be used to represent the level of inference and the methodology used to estimate a particular parameter. $\hat{\theta}^{S(S)}$ denotes an estimator of the parameter of the superpopulation where the first superscript S denotes that the parameter of interest is from the superpopulation the second enclosed in parenthesis (S) denotes the two-stage superpopulation framework used to estimate the parameter. Hartely and Seilken and later Sarndal refer to this as case 3. Likewise $\hat{\theta}^{F(S)}$ is the finite population parameter estimate under a two-stage superpopulation framework (case 2 Hartley Seilken) and $\hat{\theta}^{F(F)}$ is the finite population parameter estimate under classical finite population sampling (case 1 Hartley Sielken). As the estimation of a superpopulation parameter under finite population sampling is not possible the superscript S(F) is not used.

As it is not logical to combine the estimates from different studies that are inferentially different or calculated under different estimation methods (without accounting for the differences) we only consider cases where we are combining like estimators.

3.1 Combining estimates of the finite population parameter using sampling from a fixed population - $\hat{\theta}^{F(F)}$

With this first type of inference the simplest approach is to use design-based estimation such as a Horvitz-Thompson estimator $\hat{\theta}^{F(F)}$. Here inference is based on repeated sampling of the population of the same design. We may have an analytical model, but we do not use it in any way. Survey methodologists most frequently encounter this case.

When we consider combining estimates of $\hat{\theta}^{F(F)}$ from different surveys we can see that several cases commonly discussed in the literature fall under this category. Pooling or cumulating cases as suggested by Kish (1999) is a method of taking different estimates of a finite population

characteristic and combining them. Kish uses the term combining when he estimates what he describes as multipopulation or mulitdomains, and cumulating or pooling for periodic or rolling samples. In both these cases we need not invoke a superpopulation and inference is for the 'finite' population of interest. Repeated sampling and panel samples of the same population are also a method to combine $\hat{\theta}^{F(F)}$ (Cochran 1977). Wu (2004) suggests a method to combine information from different surveys using an empirical likelihood method similar to the approach of Zieschang (1990) and Rensessn and Nieuwenbroek (1997), the purpose of which is to improve estimates of the census parameter. Merkouis (2004) suggests a regression approach to calculate design-based direct estimators using multiple surveys and contrasts it with the typical small-area estimation indirect approach which generally uses one sample. We can also consider combing samples from non-overlapping surveys and overlapping frames as special cases of combining estimates of $\hat{\theta}^{F(F)}$. In all these cases inference is about the census parameter θ_N and not the superpopulation parameter.

3.2 Combining estimates of the finite population parameter using superpopulation theory for a finite population parameter - $\hat{\theta}^{F(S)}$

When we invoke a superpopulation model we would want the estimator of θ_N to take into account both the model and the sampling design. Therefore, with inference of the finite population parameter under a superpopulation model we would select an estimator where $E_\xi E_p\left[\left(\hat{\theta}_S^{F(S)} - \theta_N\right)^2\right]$ is minimized. Here, the subscript ξ implies expectation taken over the model and the subscript p indicates expectation taken over the sampling plan.

Fuller (1973) presents theory and estimation for the case where we assume that the finite population is a simple random sample with finite fourth moments from the superpopulation and the sample is drawn from this finite population. Hartley and Seilken provide more on the inference of the finite population parameter for regression parameters. Isaki and Fuller (1982) expand these ideas to construct model-unbiased and design-consistent regression predictors.

Here under the approach of Isaki and Fuller $\hat{\theta}^{F(S)}$ is called a predictor. A predictor is conditionally model unbiased for the finite population parameter θ_N if given the sample S if we have $E_\xi\left[\left(\theta^{F(S)} - \theta_N\right)|S\right] = 0$. It is important to note that the conditioning is not on a characteristic of the population. The predictor is unbiased with respect to the design P(S) if $E_p\left[\hat{\theta}^{F(S)}\right] = \theta_N$. They then define the best model-unbiased predictor of θ_N as the one with minimum model variance.

Although a simple random sample of a simple random sample is also a simple random sample, most surveys have a more complex design. If inclusion probabilities depend on y through more than the covariates then we will make erroneous conclusions if we do not consider the design (Pffeferman 1993). So we approach this estimation problem by assuming that the sampling design P(S) plays no role (Royall 1973), we are stating that the relationship in the model holds for every unit in the population U and the sample design should have no effect on estimation. As noted above, this can be the case as long as the probability of selection depends on y only through the x's. If the design and inclusion probabilities play no role in estimation then design

unbiasedness is unimportant. This is often referred to as a frequentist approach and or a pure model-based inference.

This frequentist approach is related to the idea of ignorability. We consider a design to be non-informative if the sampling mechanism does not depend on Y and informative if it does. More formally, let the sampling distribution related to a design be $f(i_U)$. This distribution is a function of the indicator variable I_U and the finite population U. Where I denotes the (N x 1) sample indicator variable such that $I_j = 1$ if the unit $j \in U$ is selected to the sample 0 otherwise. In the case of probability sampling $\Pr(I_j = 1) > 0$ for all $j \in U$. Let $Z = [Z_1, Z_2, ..., Z_N]'$ define a (N x q) matrix of population values of q design variables $Z(1),\ldots, Z(q)$ (different from Y). For informative sampling plans we have the case that the sampling plan is dependent on Y, that is $f(i_U | Y_U = y_U)$. Then the joint distribution of the realized variable y and the design is $f(i_U | Y_U = y_U) f(y_U; \theta)$. If we can state that the variables I_U and the values Y_U are conditionally independent given other design variables Z_U, we can also have non-informative design where $f(I_U | Z_U = z_U ; \phi)$ and the distribution of Y_U given $Z_U = z_U$ as $f(Y_U | Z_U = z_U)$ (Pffeferman 2003).

Binder and Roberts (2001) state, that a design is ignorable if for a particular analysis if the inference based on all the known information is equivalent to the inference based on the same information excluding the outcomes of the random variables related to the design. They note that ignorability is a concept related to a particular analysis and one can have ignorability with both informative and non-informative designs. All non-informative designs are ignorable but not all ignorable analyses come from a non-informative design.

In practice the analyst does not always have the exact sampling plan and should not assume that the sampling plan is independent of y. And like any classical procedure model misspecification can lead to incorrect conclusions, particularly for large samples. As Sharon Lohr states 'theoretically derived models known to hold for all observations do not often exist for survey situations'.

Thus if we consider combining estimates of this type we can see that the variance will no longer just be a function of the sampling, but it now includes a part for the superpopulation, even though the point estimate is of the finite population parameter. If we naively tried to combine $\hat{\theta}^{F(S)}$ we need to ensure that we take into account this extra variance. Additionally, we note that combining $\hat{\theta}^{F(F)}$ and $\hat{\theta}^{F(S)}$ type estimators would be problematic as the variance of the first would always be smaller than the second even for exactly the same sample.

3.3 Combining estimates of the superpopulation parameter using a two-stage sampling approach - $\hat{\theta}^{S(S)}$

If we are interested in the inferences from this model, and we wish to extend our inference to the superpopulation and its parameters, Molina et al (2001) suggest that inference should be made using the joint process of model and design. They propose a joint model-design process satisfying the following: For all the samples S that are possible under p

$$E_\xi(Y|I_s = I_s) = E_\xi(Y),$$

$$E_\xi(YY^T|I_s = I_s) = E_\xi(YY^T)$$

These conditions are satisfied when I and Y are independent. If this condition does not hold then they suggest modeling the mean and variance conditional not on the observed sample but on all possible samples under p.

Estimation of a superpopulation characteristic θ can be the objective of a meta-analysis. In this case we need to be combining estimates of $\hat{\theta}^{S(S)}$ that we felt were from different finite populations that were realizations of the same superpopulation as in Figure 2 in order to converge to the superpopulation parameter.

Figure 2
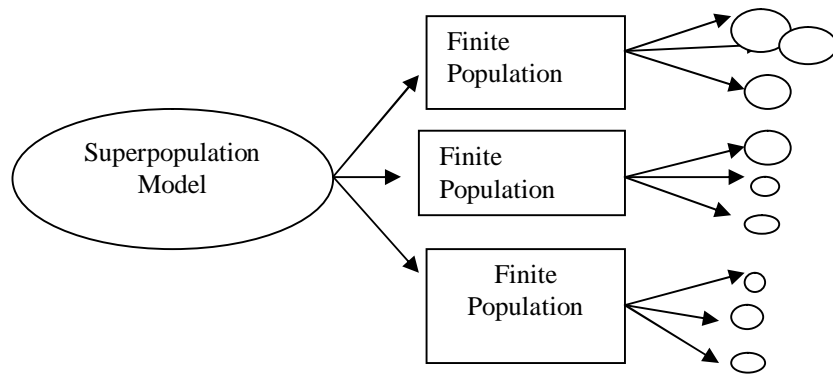
4. Meta-Analysis: Fixed and Random Effect Models

In traditional meta-analysis when we combine estimates from an experiment we assume effect size observed in a study estimates the corresponding 'population' effect with a random error that stems only from the chance associated with the individual level 'sampling error' (not sampling in the case of probability sampling but randomness due to the distribution). Differences in the studies are in the power to detect the outcome of interest and each experiment is independent of other experiments. Here we assume asymptotically that the estimator of the treatment effect θ_i is considered a random variable and has a normal distribution with mean θ_i and variance v_i. We also assume that there is homogeneity of treatment effects across all n studies, that is $\theta_1 = \theta_2 = ... = \theta_n = \theta$.

We can see that we can calculate linear combinations of unbiased estimates that are themselves unbiased estimates and then the simplest approach is to calculate a weighted average from all the studies. The fixed effect model (the weighted average) takes the general form

$$g(\hat{\theta}) = \frac{\sum_{i=1}^{n} w_i \hat{\theta}_i}{\sum_{i=1}^{n} w_i}$$

where $\hat{\theta}$ is the estimator and the weights w_i are chosen to minimize some constraint, such as $g(\hat{\theta})$ having minimum variance. Here the weights themselves are generally not known, but are estimated from the data. Hardy (2003) shows that there is no systematic bias caused by the estimation of the weights in the overall fixed effect estimate.

Many researchers have addressed the optimum choice of weights. In general we wish to give the greatest weight to studies that have the most precise estimates. If we want to minimize the variance then the minimum variance estimate is the one obtained by taking the weight for the ith study to be the inverse of the variance.

Fixed effect meta-analysis methods do not provide unbiased results when the heterogeneity of the estimates is not explained by known differences in the studies. Here the assumptions are that the true treatment effect of each trial is normally distributed with mean θ and a between-study variance σ_B^2. Inference is based on the assumption that effects are independently and identically sampled from an overall population of all such effects.

The infinite population is an important assumption in the classical parametric approach. The assumption that the effects are random implies that we must include a between-study as well as a within study component of variation when estimating an effect size and a confidence interval (Demets). Then we would assume that the variability beyond the individual 'sampling error' is random, that it stems from random differences among studies whose sources cannot be identified. This will normally mean that the confidence interval is at least as wide as that from a fixed effect model.

We suppose that each trial estimate has a normal distribution θ_i and variance v_i. Here let $\hat{\theta}$ be a pooled estimate of effect size (with weights chosen to be the inverse of the estimated variances). Then to calculate the random-effects estimator (DerSimonian-Laird) we use

$$\widehat{\theta}_{DL} = \frac{\sum_{i=1}^{k} \hat{\theta}_i /(\hat{\tau}^2 + v_i^2)}{\sum_{i=1}^{k} 1/(\hat{\tau}^2 + v_i^2)}$$

where τ^2 is a point estimate of the heterogeneity variance. This is a generalization of the fixed effect approach which occurs when $\tau^2 = 0$. However, it should be noted that the weights are themselves estimates and that will lead to an overestimation of the between study variance (Hardy year).

5. Simulations

We illustrate meta-analysis on the most common case where we have a single finite population. Using this framework we ran two simulations.

The first simulation was of a finite population of 100000 individuals from a simple superpopulation model. The outcome of interest Y was generated as a normal random variable with mean 500 and variance of 10000. The stratification variable used in this example had three levels and was generated to be strongly correlated with y ($\rho = .7$). 1000 replicated samples of

1000 were then selected from this large finite population using SAS using both simple random and stratified sampling. For the stratified case a simple random sample was selected within each stratum proportional to the size of the stratum. Design unbiased estimates were calculated for each of the 1000 replicates. Fixed model using the inverse of the design variance as the weight were then calculated and random effects estimates using the variance between samples as an estimate of τ^2 and compared for convergence to the superpopulation estimate.

The second simulation involved a finite population of 10 million individuals from a simple superpopulation model. The outcome of interest was generated as a normal random variable with mean 500 and variance of 100. The stratification variable used in this example had three levels and was generated to be strongly correlated with y ($\rho = .7$). 1000 replicated samples of 100 were then selected from this large finite population using procsurvey select stratified sampling with a simple random sample within each stratum proportional to the size of the stratum. Design based and purely model based estimates were calculated for each of the 1000 replicates. Fixed model estimates using the inverse of the design variance as the weight were then calculated and compared for convergence of the two methods.

5.1 Convergence

We can empirically show that when we are combining design unbiased estimates of the finite population parameter we will converge to the finite population parameter. Here the superpopulation has a mean of 500, but the finite population had a mean of 512. The convergence of the first 500 samples of the first simulation to the finite population mean is shown in Figure 3. Figure 3
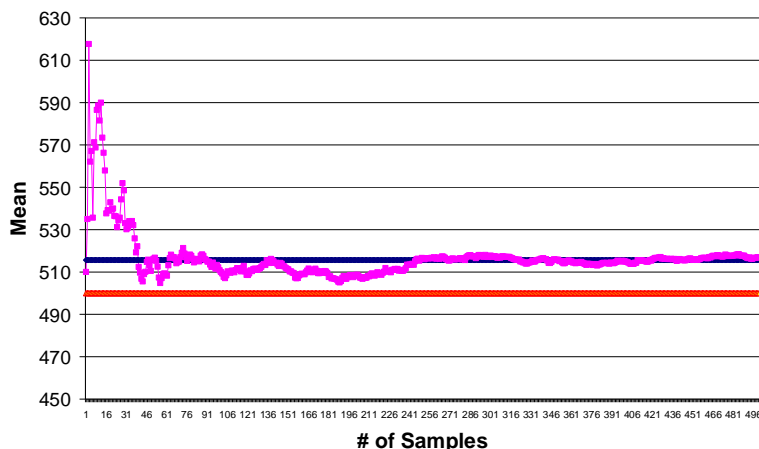

This should make intuitive sense to the reader though when we combine information under a superpopulation framework we often neglect this. If we wish to converge to the superpopulation parameter we must in fact be combining different realizations of the superpopulation as in Figure 2, not different samples of the same finite population.

5.2 Design or Model based

If we have a very large finite sample we know that the census estimate will be closer to the superpopulation estimate. Using a large finite population where sample mean is 499, we illustrate that purely model based estimates can be biased when the sampling design is not ignorable. Here

the stratification variable was highly correlated and if we look at the first 10 replicates, in Figure 4, the design based estimates in blue are not biased whereas the purely model based is.
Figure 4
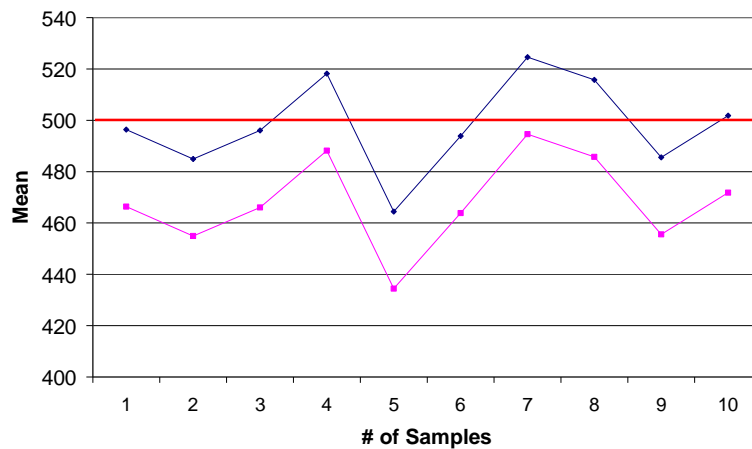
5.3 Small Sample Properties

We used different seeds to discover the behaviour of these estimates when the number of surveys we would combine is small. For many of the simulations these estimation techniques are not particularly robust when the number of samples being combined is small. Figure 5 illustrates that only two of the estimates were within 5% of the true value. This is consistent with findings combining experimental data.
Figure 5
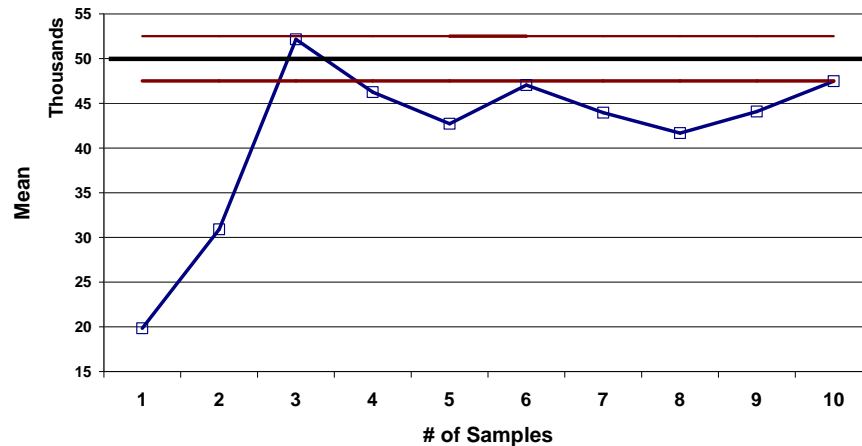
6. Practical Considerations

Systematic reviews and subsequent meta-analysis are a form of research that use other primary research studies as their subjects or units of observation. Like other forms of research they have many steps from planning, through implementation to dissemination. It has been widely recognized that as Naylor states 'methodology is less important than determining which results are to be aggregated' when combining experimental data. No two studies are identical and although it is often reported that research experiments should be reproducible, the reality is that research evolves. Individual research is designed based on previous results and findings not to just reproduce findings but to advance knowledge (Peto 87).

Establishing the eligibility criteria for inclusion and exclusion is a fundamental step in this process. There has been considerable research on the affect of data quality and publication bias on the results of a meta-analysis of a clinical trial (Schulz, 1996). In the case of survey data, as with other systematic reviews, the meta-analysis is not simply the combining of simple random sets data. Data sets need to be chosen using the same rigorous standards of other systematic reviews in order to minimize the publication and other sources of bias.

The selection of surveys should be a clearly defined process. Whether we use a research-based electronic facility (i.e. Pub-med, the Canadian Research Index, the Social Sciences Citation Index, POPLINE, PapersFirst or J-stor) or web-based electronic search engines including Google and Google Scholar, or contact a statistical agency directly, the researcher should set up and document the protocol use to select surveys. These protocols should include the key words, search criteria, the years under study and a description of the databases that were searched. Then the researcher should apply clear criteria for inclusion and exclusion to all studies found in the review. These criteria should be based on the comparability and the quality of the studies.

Data quality has many dimensions; accessibility, interpretability, coherence, relevance, unbiasedness, accuracy, precision, timeliness, coverage, completeness, consistency, and some would add methodological soundness. Researchers need to be able to determine the extent to which the accuracy and other quality factors are consistent with their interpretation. They can do this by verifying the conceptual framework and definitions used in the survey design, collection and processing of the data. Without this verification the researcher can not be certain that the data they analyze meets their needs.

7. Conclusions

As Schenker et al (2007) points out when combining complementary survey data the researcher needs to know 'to what extent are the target populations of the surveys mutually exclusive and exhaustive subsets of the overall domain of interest'. Following Schenker's lead we expand this to take into account that estimation can be done for different levels of inference using different estimation frameworks. So not only would want to know what is the target population and if the surveys differentially cover their target population we would also need to know why and how these survey estimates were calculated; adding into the framework the inference and the estimation procedure used and then assessing if these definitions overlap.

Additionally we would also want to understand the possible biases associated with combining estimates that are purely model based and recommend that whenever possible design based estimates are used. And like any estimation technique we would want to not only have convergence to what we feel we are estimating (be it the superpopulation or the finite population parameter), but also that the small sample properties behave in a way which satisfies our estimation needs.

References:

Glass, VG (1976). Primary, secondary, and meta-analysis of research. International Journal of Psychiatry, 1, 97-178.

Altman D, (1985) Comparability of randomised groups. Statistician 1985; 34: 125–36.

Sarndal, C. E., Swensson, B., and Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer-Verlag.

Hartley HO, Sielken RL Jr., (1975) A 'super-population viewpoint' for finite population sampling. Biometrics. Jun;31(2):411-22.

Kish, L. (1999). Cumulating/combining population surveys. *Survey Methodology* 25, 129–138. Cochran (1977).
Cochran,W.G. (1977). *Sampling Techniques*, 3rd ed. JohnWiley & Sons, New York.

Wu (2004). Combining information from multiple surveys through the empirical likelihood method. *Canadian Journal of Statistics*, 32 (1), 15-26.

Zieschang (1990). Sample Weighting Methods and Estimation of Totals in the Consumer Expenditure Survey. *Journal of the American Statistical Association*, 85, 986-1001

Rensessn and Nieuwenbroek (1997), Aligning estimates for common variables in two or more sample surveys. *Journal of the American Statistical Association*, 92, 368374.

Fuller (1973) Fuller,W.A. (1975). Regression analysis for sample survey. *Sankhȳ a Series C* **37**, 117–132.

Isaki and Fuller (1982) Isaki, C.T., Fuller, W.A. (1982). Survey design under the regression super-population model. *Journal of the American Statistical Association* **77**, 89–96.

Merkouris, T. (2004). Combining independent regression estimators from multiple surveys. *Journal of the American Statistical Association*, 99, 11311139.

Graubard, Barry I. and Korn, Edward L. (2002), "Inference for Superpopulation Parameters Using Sample Surveys," *Statistical Science, 17,* pp. 73- 96.

Pfeffermann, D. (1993). The role of sampling weights when modeling survey data. *International Statistical Review/Revue Internationale de Statistique* 61(2), 317–337.

Royall, R.M. (1970). On finite population sampling theory under certain regression models. *Biometrika* **57**,377–387.

Binder, David A. and Roberts, Georgia R. (2001), "Can Informative Designs be Ignorable?" *Newsletter of the Survey Research Methods Section, Issue 12,* American Statistical Association.

Molina, EA, Smith, TMF and Sugden, RA (2001) Modelling overdispersion for complex survey data. International Statistical Review, 69, (3), 373-384.

Naylor, D (1997)Meta-analysis and the meta-epidemiology of clinical research: Meta-analysis is an important contribution to research and practice but it's not a panacea, Editorial, *BMJ.* ; 315 : 617

Peto, R (1987). Why do we need systematic overviews of randomized trials? *Stat Med.* Apr-May;6(3):233-44.

Schulz KF. Randomised trials, human nature, and reporting guidelines. *Lancet* 1996; 348**:** 596–98.

Schenker, N. and T.E. Raghunathan, (2007), "Combining Information from Multiple Surveys to Enhance Estimation of Measures of Health," *Statistics in Medicine, 26*, 1802-1811.