Approaches for combining data from multiple probability samples

Loveness N. Dzikiti[1,3], Marcel D. T. Vieira[2,1] and Brendan Girdler-Brown[1]

[1] University of Pretoria, Pretoria, South Africa

[2] Federal University of Juiz de Fora, Juiz de Fora, Brazil

[3] Ross University School of Veterinary Medicine, St Kitts and Nevis

IASS Webinar 15

Outline

Approaches for combining data from multiple probability samples

- Motivation and introduction
- Methodology
- Simulation study
- Illustrative examples
- Some concluding remarks

Motivation

- It is sometimes necessary to combine data from different surveys either to increase the sample size and precision of estimates for sub-populations of interest or to improve coverage.
- Large surveys may not have enough sampling units for some subgroups of interest to allow any meaningful inference.
- Combining surveys can also provide a broader opportunity to investigate more research questions than the original individual surveys.

Introduction

- There is substantial literature on studies that pool survey data, but it is still not clear which are the most efficient methods and how sampling designs might affect the efficiency of combined estimators.
- Should estimates from the surveys involved be given equal weights in the calculation of the combined estimate?
- How should they be weighted and why?
- Pooled and separate approaches (*e.g.* Wannell and Thomas, 2009; Roberts and Binder, 2009).

Pooled and separate approaches

Pooled approach

- Individual records from the surveys are merged resulting in an increase in the sample size.
- Original sampling weights may be modified, and estimates calculated based on the new weights and the pooled sample.
- Micro-data from the individual surveys need to be available.
- After combined estimate is calculated, no need to go back to the individual survey data.
- Pooled approach requires more technical expertise.

Separate approach

- Combined estimate can be calculated from the individual survey estimates when micro-data is not available.
- May be cumbersome if the required estimates are not published or have been published without their variance estimates.
- If you need to calculate estimates separately for each survey, that will require sampling design variables, clusters, strata identifiers and sample design weights.

Examples

Pooled approach

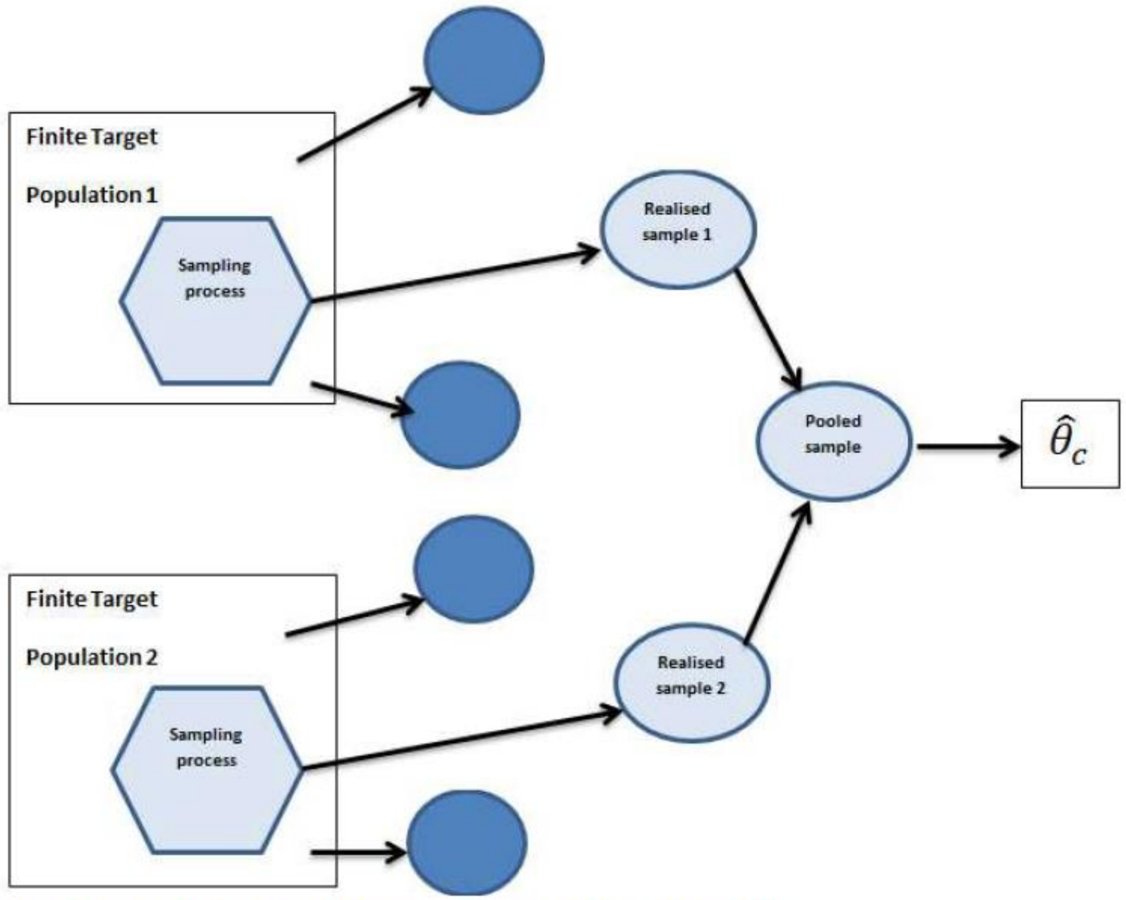
- Schenker and Raghunathan (2007) combined the National Health Interview Survey (NHIS) and the National Nursing Home Survey (NNHS) to estimate the prevalence of chronic conditions among elderly people in the US.
- The two surveys were treated as strata of the combined population.
- Complementary surveys, as one targeted the non-institutionalised elderly people and the other referred to the elderly people living in nursing homes.

Separate approach

- Latouche, Dufor and Merkouris (2000), considered the separate approach to produce cross-sectional estimates based on combining two longitudinal panels of the Survey of Labour and Income Dynamics (SLID), which consists of two overlapping panels with duration of six years each.

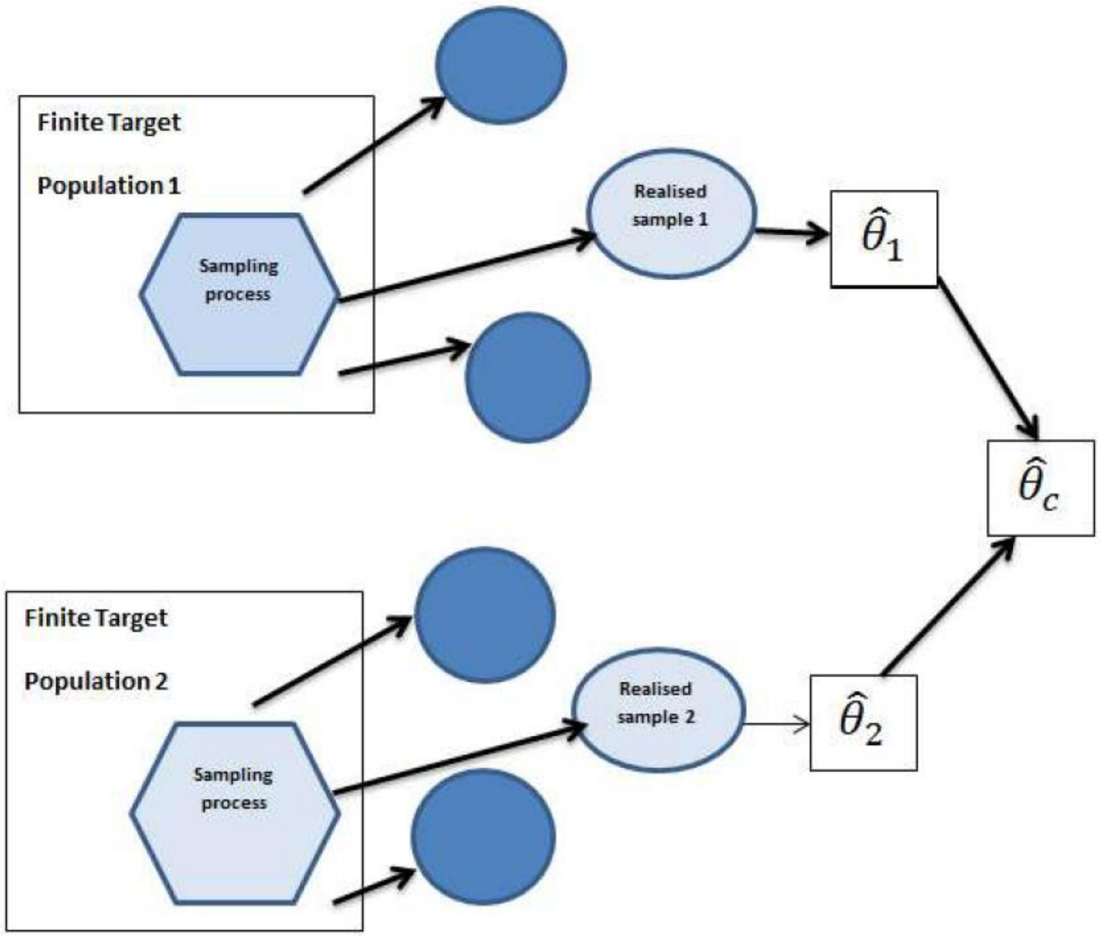
Pooled and separate approaches

Pooled approach



Pooled Analysis: Adapted from G. Roberts and D. Binder 2009.

Separate approach



Separate Approach: Adapted from G. Roberts and D. Binder 2009

Aims

- To evaluate different methods used to combine survey data under the separate approach are evaluated under different sampling designs.
- To propose alternative methods.
- Methods are evaluated through simulation, in the context of:
  - simple random sampling without replacement (srswor),
  - stratified random sampling (strs) and
  - two stage cluster random sampling (2cls)
  - from finite populations generated from alternative super-population models.

Methodology

- A combined estimate is calculated as a function of the separate estimates, commonly based on a linear combination. In the case of two surveys,

$$\hat{\theta}_c = \alpha\hat{\theta}_1 + (1-\alpha)\hat{\theta}_2 \, ,$$

where $\hat{\theta}_c$ is the combined estimator of the target parameter θ, $\hat{\theta}_d$, with d = 1, 2, is the estimator based on data from survey d, and α is a weight allocated to survey 1 (Kish, 1999; Roberts and Binder, 2009).

- Alternative choices of α can be adopted.

Combining surveys and meta-analysis

- Meta-analysis, which aims to improve the precision of the estimate of an effect size through pooling information across studies.
- Effect size may be defined as an impact of an intervention, a relationship between two variables or even the prevalence of a condition.
- Different studies may differ in their accuracy in estimating the true effect size.
- Meta-analysis may use the inverse of variance of the estimate calculated for each study to weight the estimates from the different studies.
- Less precise studies are given less weight in the estimation.
- We consider such weighting alternative for combining sample surveys estimates.

Some further review

- Fox (2010) compared the sample size, the inverse variance and average weighting methods for combining survey estimates.
- The inverse of variance has also been used by Maheswaran et al. (2015) to combine surveys in a repeated national cross-sectional study of self-reported health and socio-economic inequalities in England.
- Kish (1999) suggested using equal weights or sample sizes to weight samples when combining surveys.

Additional methods

- We explore alternative importance measures for the different surveys involved in the estimation of the combined estimate, including the inverses of
  - the coefficient of variation (cv) of the estimator, and
  - the misspecification effect (meff) (Skinner, 1986).
- The efficiency of tree alternative sampling designs is considered for the surveys estimates that are being combined –
  - srswor, strs, and 2cls.

The inverse of the coefficient of variation

- When combining estimates from different surveys, it is desirable to give more weight to surveys with greater precision.
- Hence it is important to consider variability.
- We propose to weight by the inverse of the *cv* when calculating combined estimates, $\hat{\theta}_c$,

$$\hat{\theta}_c = \sum_{d=1}^{D} \left(\frac{\frac{1}{cv(\hat{\theta}_d)} \times \hat{\theta}_d}{\sum_{d=1}^{D} \frac{1}{cv(\hat{\theta}_d)}} \right) \text{ with } cv(\hat{\theta}_d) = \frac{se(\hat{\theta}_d)}{\hat{\theta}_d} \times 100\%,$$

where $se(\hat{\theta}_d)$ estimates the standard error of $\hat{\theta}_d$, allowing for the sampling design of survey d.

Why the cv as an importance measure?

- The cv is independent of the scale of measurement.
- An extreme example: an estimator with an estimated standard error of 10 with a point estimate of 100 results in a cv = 10% whereas an estimator with an estimated se of 10 with an estimate of 1000 results in a cv = 1%.
  - Although the estimate of the se is the same in both cases, there is greater variability for an estimator with a cv = 10%.
- When combining estimates from different surveys, we consider the cv offers a more reasonable measure of variation for estimators than the variance and therefore the se (Sorensen, 2002).

The misspecification effect (I)

- The meff is defined as (Skinner, 1986)

$$meff(\hat{\theta}, var_0) = \frac{Var_{true}(\hat{\theta})}{E_{true}[var_0(\hat{\theta})]},$$

where $Var_{true}(\hat{\theta})$ is the true variance of $\hat{\theta}$, considering the sampling design adopted to select the sample while $var_0(\hat{\theta})$ is the variance estimator assuming srswor. The meff measures of how much $var_0(\hat{\theta})$ over- or underestimates $Var_{true}(\hat{\theta})$, and may be estimated by

$$\widehat{meff}(\hat{\theta}, var_0) = \frac{var_{true}(\hat{\theta})}{var_0(\hat{\theta})}$$

The misspecification effect (I)

- The meff is defined as (Skinner, 1986)

$$meff(\hat{\theta}, var_0) = \frac{Var_{true}(\hat{\theta})}{E_{true}[var_0(\hat{\theta})]},$$

where $Var_{true}(\hat{\theta})$ is the true variance of $\hat{\theta}$, considering the sampling design adopted to select the sample while $var_0(\hat{\theta})$ is the variance estimator assuming srswor. The meff measures of how much $var_0(\hat{\theta})$ over- or underestimates $Var_{true}(\hat{\theta})$, and may be estimated by

$$\widehat{meff}(\hat{\theta}, var_0) = \frac{var_{true}(\hat{\theta})}{var_0(\hat{\theta})}.$$

The misspecification effect (II)

- The estimated values for the meff may be interpreted as

  a. $\widehat{meff}\left(\hat{\theta}, var_0\right) < 1 \rightarrow Bias\left[var_0\left(\hat{\theta}\right)\right] > 0;$

  b. $\widehat{meff}\left(\hat{\theta}, var_0\right) = 1$ suggests correct estimation of $Var_{true}\left(\hat{\theta}\right)$; and

  c. $\widehat{meff}\left(\hat{\theta}, var_0\right) > 1 \rightarrow Bias\left[var_0\left(\hat{\theta}\right)\right] < 0.$

The inverse of the meff

- The inverse of the misspecification effect (meff) is also proposed here as weight when combining estimates from complex sampling surveys.

$$\hat{\theta}_c = \sum_{d=1}^{D} \left(\frac{\frac{1}{\widehat{meff}(\hat{\theta}_d)} \times \hat{\theta}_d}{\sum_{d=1}^{d} \frac{1}{\widehat{meff}(\hat{\theta}_d)}} \right)$$

A general combined estimator

- A general combined estimator is be given by

$$\hat{\theta}_c = \sum_{d=1}^{D}\left(\frac{\omega\times\widehat{\theta}_d}{\sum_{d=1}^{D}\omega}\right),$$

where ω is the relevant weight.

- The following particular cases are considered: $\omega = [cv(\hat{\theta}_d)]^{-1}$, $\omega = [\widehat{meff}(\hat{\theta}_d)]^{-1}$, and those studied by Fox (2010), *i.e.* $\omega = n$ and $\omega = [var(\hat{\theta})]^{-1}$.

Simulation Study

- For all the simulation scenarios considered, both sampling and combining process was repeated K = 1,000 times.

Simulation under simple random sampling without replacement (I)

- A normally distributed superpopulation with mean 100 and variance 25 was considered to generate the population.
- Finite populations of sizes N = 100,000, 1,000,000 and 5,000,000, were simulated.
- For each generated population, srswor was adopted to select D = 2, 5, 10, and 20 samples of sizes n = 1,000, 5,000 and 10,000.

Simulation under simple random sampling without replacement (II)

- Estimates of means were then calculated from each selected sample.
- The sample estimates were then combined using the alternative weighting methods.

Simulation under stratified sampling (I)

- A stratified finite population of size N = 1,000,000 was simulated from a super-population with a normal distribution.
- The population was generated with five strata of equal size.
- Parameters of the distributions within the strata were chosen in such a way that the combined population would have a normal distribution with mean 100 and variance 25.

Simulation under stratified sampling (II)

- Sampling fractions were chosen to yield the required sample sizes of 1,000, 5,000 and 10,000 with higher sampling fractions in the strata with greater variation.
- Within each stratum, the sample was selected by srswor.
- The selected stratified samples were then combined using different weighting methods.

Simulation under two-stage cluster sampling (I)

- A population of size 1,000,000 clustered by primary sampling units of size 250 was simulated according to the following super-population model,

$$Y_{ij} = \mu + u_i + v_{ij},$$

where μ is the overall mean of the observed variable Y_{ij}, u_i are the cluster random effects, v_{ij} are the individual random effects, unobservable individual-specific trait.

Simulation under two-stage cluster sampling (II)

- Cluster and individual random effects were simulated from normal distributions with mean of 0 and a variance of 5, and mean of 0 and a variance of 20, respectively.
- Samples of size 1,000, 5,000 and 10,000 were selected.

Simulation under two-stage cluster sampling (III)

- For the sample size of 1,000, 50 clusters of size 250 were selected by srswor, and 20 units were then selected also by srswor from each of the selected clusters.
- 20 units were selected from each of 250 clusters for the sample size 5,000, and 20 units were selected from each of the 500 clusters for the sample size 10,000.
- The selected cluster random samples were then combined using different weighting methods.

Combining D samples of size n = 10,000, from a population of size N = 1,000,000, distributed as N(100,25), under srswor.

Number of samples combined (D)	*Weighting strategy*	*Relative bias* ($\times 10^{-3}$)	*cv* ($\times 10^{-3}$)	*mse* ($\times 10^{-3}$)
only one sample	-	-2.420 %	1.574 %	2.484
2	n	-0.920 %	1.141 %	1.303
	$[var(\hat{\theta})]^{-1}$	-0.909 %	1.141 %	1.304
	$[cv(\hat{\theta}_d)]^{-1}$	-0.902 %	1.141 %	1.304
5	n	0.165 %	0.649 %	0.518
	$[var(\hat{\theta})]^{-1}$	0.180 %	0.649 %	0.517
	$[cv(\hat{\theta}_d)]^{-1}$	0.192 %	0.649 %	0.518
10	n	0.037 %	0.360 %	0.242
	$[var(\hat{\theta})]^{-1}$	0.040 %	0.359 %	0.242
	$[cv(\hat{\theta}_d)]^{-1}$	0.060 %	0.360 %	0.242
20	n	0.104 %	0.164 %	0.125
	$[var(\hat{\theta})]^{-1}$	0.107 %	0.164 %	0.125
	$[cv(\hat{\theta}_d)]^{-1}$	0.129 %	0.164 %	0.125

Simulation under srswor Results

- Combining simple random samples improved estimators' mse when compared to considering data from only one sample.
- Estimated cv and mse tend to decrease as more samples are combined.
- Weighting methods resulted in estimators with comparable estimated relative bias, cv and mse values.
- Proposed weighting strategy based on the cv performed very similarly to previously proposed strategies.

Combining D samples of size n = 10,000, from a population of size N = 1,000,000, distributed as N(100,25), under strs

Number of samples combined (D)	*Weighting strategy*	*Relative bias* ($\times 10^{-3}$)	*cv* ($\times 10^{-3}$)	*mse* ($\times 10^{-3}$)
only one sample	-	-0.251 %	1.086 %	1.179
2	n	1.047 %	0.778 %	0.606
	$[var(\hat{\theta})]^{-1}$	1.057 %	0.782 %	0.613
	$[cv(\hat{\theta}_d)]^{-1}$	1.074 %	0.782 %	0.613
	$[\widehat{meff}(\hat{\theta}_d)]^{-1}$	1.068 %	0.782 %	0.613
5	n	0.637 %	0.489 %	0.239
	$[var(\hat{\theta})]^{-1}$	0.635 %	0.489 %	0.240
	$[cv(\hat{\theta}_d)]^{-1}$	0.646 %	0.489 %	0.240
	$[\widehat{meff}(\hat{\theta}_d)]^{-1}$	0.627 %	0.490 %	0.240
10	n	0.499 %	0.348 %	0.121
	$[var(\hat{\theta})]^{-1}$	0.496 %	0.348 %	0.122
	$[cv(\hat{\theta}_d)]^{-1}$	0.508 %	0.348 %	0.122
	$[\widehat{meff}(\hat{\theta}_d)]^{-1}$	0.497 %	0.349 %	0.122
20	n	0.512 %	0.282 %	0.059
	$[var(\hat{\theta})]^{-1}$	0.513 %	0.283 %	0.059
	$[cv(\hat{\theta}_d)]^{-1}$	0.524 %	0.283 %	0.059
	$[\widehat{meff}(\hat{\theta}_d)]^{-1}$	0.475 %	0.283 %	0.060

Simulation under strs Results

- Weighting strategies performed similarly under strs and normally distributed data, including those proposed here.
- Combining stratified random samples improved estimators' cv and mse when compared to considering data from only one sample.
- Moreover, estimated cv and mse tend to decrease as more samples are combined also for stratified samples.

Combining D samples of size n = 10,000, from a population of size N = 1,000,000, distributed as N(100,25), under 2cls

Number of samples combined (D)	*Weighting strategy*	*Relative bias* ($\times 10^{-3}$)	*cv* ($\times 10^{-3}$)	*mse* ($\times 10^{-3}$)
'one sample'	-	-10.930 %	9.764 %	9.653
2	n	-14.760 %	6.104 %	3.944
	$[var(\hat{\theta})]^{-1}$	-3.640 %	6.167 %	3.816
	$[cv(\hat{\theta}_d)]^{-1}$	-4.030 %	6.170 %	3.823
	$[\widehat{meff}(\hat{\theta}_d)]^{-1}$	-3.810 %	6.162 %	3.812
5	n	-0.780 %	4.067 %	1.655
	$[var(\hat{\theta})]^{-1}$	-0.490 %	4.087 %	1.671
	$[cv(\hat{\theta}_d)]^{-1}$	-0.550 %	4.076 %	1.662
	$[\widehat{meff}(\hat{\theta}_d)]^{-1}$	-0.710 %	4.052 %	1.642
10	n	-3.400 %	2.808 %	0.800
	$[var(\hat{\theta})]^{-1}$	-3.200 %	2.822 %	0.807
	$[cv(\hat{\theta}_d)]^{-1}$	-3.200 %	2.808 %	0.802
	$[\widehat{meff}(\hat{\theta}_d)]^{-1}$	-2.900 %	2.814 %	0.809
20	n	1.370 %	2.829 %	0.395
	$[var(\hat{\theta})]^{-1}$	1.670 %	1.982 %	0.394
	$[cv(\hat{\theta}_d)]^{-1}$	1.620 %	1.978 %	0.394
	$[\widehat{meff}(\hat{\theta}_d)]^{-1}$	1.390 %	1.978 %	0.402

Simulation under 2cls Results

- Combining more samples and larger samples resulted in estimators with lower cv and mse values regardless of the weighting method used under the three sampling designs that have been evaluated.
- However, based on relative bias, cv and mse estimated through simulation, weighting strategies behaved somehow differently depending on the number of samples being combined under 2cls.

Mean square errors and efficiency ratios under a population of size N = 1,000,000, distributed as N(100,25).

$$r_1 = \frac{se_{strs}}{se_{srswor}}$$

$$r_2 = \frac{se_{2cls}}{se_{srswor}}$$

Number of samples combined (D)	*Weighting strategy*	*mse* ($\times 10^{-3}$)			r_1	r_2
		srswor	*strs*	*2cls*		
'*one sample*'	-	2.484	1.179	9.653	0.690	6.203
2	n	1.303	0.606	3.944	0.682	5.350
	$[var(\hat{\theta})]^{-1}$	1.304	0.613	3.816	0.685	5.405
	$[cv(\hat{\theta}_d)]^{-1}$	1.304	0.613	3.823	0.685	5.408
	$[\widehat{meff}(\hat{\theta}_d)]^{-1}$	-	0.613	3.812	-	-
5	n	0.518	0.239	1.655	0.753	6.267
	$[var(\hat{\theta})]^{-1}$	0.517	0.240	1.671	0.753	6.297
	$[cv(\hat{\theta}_d)]^{-1}$	0.518	0.240	1.662	0.753	6.280
	$[\widehat{meff}(\hat{\theta}_d)]^{-1}$	-	0.240	1.642	-	-
10	n	0.242	0.121	0.800	0.967	7.800
	$[var(\hat{\theta})]^{-1}$	0.242	0.122	0.807	0.969	7.861
	$[cv(\hat{\theta}_d)]^{-1}$	0.242	0.122	0.802	0.967	7.817
	$[\widehat{meff}(\hat{\theta}_d)]^{-1}$	-	0.122	0.809	-	-
20	n	0.125	0.059	0.395	1.482	12.085
	$[var(\hat{\theta})]^{-1}$	0.125	0.059	0.394	1.482	12.061
	$[cv(\hat{\theta}_d)]^{-1}$	0.125	0.059	0.394	1.482	12.061
	$[\widehat{meff}(\hat{\theta}_d)]^{-1}$	-	0.060	0.402	-	-

Remarks (I)

- strs was the most efficient sampling design considered for all the weighting methods and most of the numbers of samples being combined.
- 2cls was the least efficient sampling design when combining estimates from different samples under normality when compared to strs and srswor.
- Such results are consistent with the survey sampling literature for estimates produced from only one sample (e.g. Kish, 1995; and Lohr, 2010).

Remarks (II)

- Further simulation results, not included here, suggest that
  - (a) combining samples of larger sizes improve results when compared to combining smaller samples;
  - (b) mse increases as the superpopulation variance increases;
  - (c) superpopulation variance did not influence the choice of weighting method; but
  - (d) the choice of weighting method depends on the size of the population.

Illustrative Examples with Real Data

- We consider data from two large-scale South African national surveys:
- (a) South African Community Survey (CS) ; and
- (b) South African General Household Survey (GHS).

Illustrative Examples (I)

- Both surveys have adopted complex survey designs with stratification and cluster sampling.
- However, publicly available data for both surveys do not provide all the sampling design variables that would be necessary for combining estimates allowing for their sampling designs.
- Primary sampling unit identifiers are not publicly available with CS data and strata identifiers are not provided with GHS data.

Illustrative Examples (II)

- An application under two alternative sampling designs is illustrated considering the selection of two probability samples from the data sets of each survey, which are taken here as our study populations.
- Weighting strategies are illustrated under srswor selecting samples from the CS data and under strs selecting samples from the GHS.
- Variable of interest is the age of the head of the household, observed in both surveys. Because of HIV/AIDS, there is a high number of child-headed households in South Africa.

Combined CS estimates of mean age of the household head and sampling error under srswor

$$\theta = \bar{Y}_{CS} = 47{,}48$$

Sample size	*Weighting method*	*Combined estimate*	*Sampling error of the combined estimate*	*Standard error of a 'one sample' estimate*
500	n	47,75	-0,26	0,74
	$[var(\hat{\theta})]^{-1}$	47,71	-0,22	
	$[cv(\hat{\theta}_d)]^{-1}$	47,74	-0,26	
1200	n	47,37	0,12	0,47
	$[var(\hat{\theta})]^{-1}$	47,37	0,12	
	$[cv(\hat{\theta}_d)]^{-1}$	47,37	0,12	
2000	n	47,46	0,03	0,37
	$[var(\hat{\theta})]^{-1}$	47,45	0,03	
	$[cv(\hat{\theta}_d)]^{-1}$	47,46	0,03	
5000	n	47,44	0,04	0,23
	$[var(\hat{\theta})]^{-1}$	47,44	0,04	
	$[cv(\hat{\theta}_d)]^{-1}$	47,44	0,04	

Combined GHS estimates of mean age of the household head and sampling error under strs

$$\theta = \bar{Y}_{GHS} = 47{,}82$$

Sample size	Weighting method	Combined estimate	Sampling error of the combined estimate	Standard error of a 'one sample' estimate
500	n	47,80	0,02	0,19
	$[var(\hat{\theta})]^{-1}$	48,62	-0,80	
	$[cv(\hat{\theta}_d)]^{-1}$	47,80	0,02	
	$[\widehat{meff}(\hat{\theta}_d)]^{-1}$	47,83	-0,01	
1200	n	48,15	-0,32	0,11
	$[var(\hat{\theta})]^{-1}$	48,15	-0,32	
	$[cv(\hat{\theta}_d)]^{-1}$	48,14	-0,31	
	$[\widehat{meff}(\hat{\theta}_d)]^{-1}$	48,04	-0,22	
2000	n	47,84	-0,02	0,09
	$[var(\hat{\theta})]^{-1}$	47,80	0,02	
	$[cv(\hat{\theta}_d)]^{-1}$	47,84	-0,02	
	$[\widehat{meff}(\hat{\theta}_d)]^{-1}$	47,85	-0,03	
5000	n	47,84	-0,03	0,05
	$[var(\hat{\theta})]^{-1}$	47,85	-0,02	
	$[cv(\hat{\theta}_d)]^{-1}$	47,88	-0,06	
	$[\widehat{meff}(\hat{\theta}_d)]^{-1}$	48,14	-0,32	

Concluding remarks (I)

- The population variance did not influence the choice of weighting method, i.e. the weighting strategy with the lowest mse value remains the same when the superpopulation variance is increased.
- Combining more samples generally improved estimates regardless of the weighting method or sampling techniques used and combining samples of larger sizes improve results when compared to combining smaller samples.
- Combining more samples results in better estimates in terms of cv and mse.

Concluding remarks (II)

- Combining stratified random samples resulted in the lowest standard errors and mse, followed by combining simple random samples and then, combining two stage cluster random samples.
- Weighting methods were illustrated considering South African survey data: all the weighting methods resulted in mean estimates close to the population mean, yielding small sampling errors.
- Combining samples results mostly in better estimates when compared to using estimates obtained from one sample survey.

References

Borenstein, M., Hedges, L.V., Higgins, J.P.T., Rothstein, H.R. (2009) Introduction to meta-analysis. New York: John Wiley & Sons.
Fox, K. (2010) A Framework for the Meta-Analysis of Survey Data. In: JSM Proceedings, Survey Research Methods Section, 527–38.
Kish, L. (1995) Survey Sampling. Classics. New York: John Wiley & Sons.
Kish L. (1999) Cumulating / Combining Population Surveys. Surv. Methodol., 25(2):129–38.
Lohr, S.L. (2010) Sampling design and analysis. 2nd ed. Boston: Richard Stratton.
Maheswaran, H., Kupek, E., and Petrou, S. (2015) Self-reported health and socio-economic inequalities in England, 1996 e 2009: Repeated national cross-sectional study. Soc. Sci. Med. [online], 136:135–46. Available from: http://dx.doi.org/10.1016/j.socscimed.2015.05.026
Nguyen, N., Burke, O., and Murphy, P. A. (2016) A Simulation Study of Weighting Methods to Improve Labour Force Estimates of Immigrants in Ireland. J Off Stat 2016.
Roberts, G. and Binder, D. (2009) Analyses Based on Combining Similar Information from Multiple Surveys. In: JSM Proceedings, Survey Research Methods Section, 2138–47.
Schenker, N, and Raghunathan, T.E. (2007) Combining information from multiple surveys to enhance estimation of measures of health. Statist. Med., 26(8):1802–11.
Skinner, C. J. (1986) Design Effects of Two-Stage Sampling. Journal of the Royal Statistical Society – Series B, 48(1), 89-99.
Sorensen, J.B. (2002) The Use and Misuse of the Coefficient of Variation in Organizational Demography Research. Sociol. Meth. Res., 30(4):475–91.
Statistics South Africa (2016a). Community Survey 2016: Technical Report [online]. Available from: http://oemmndcbldboiebfnladdacbdfmadadm/http://cs2016.statssa.gov.za/wp-content/uploads/2016/06/CS-2016-Technical-report_Web.pdf
Statistics South Africa (2016b). Community survey 2016 in brief [online]. Available from: http://cs2016.statssa.gov.za/wp-content/uploads/2016/07/NT-30-06-2016-RELEASE-for-CS-2016-_Statistical-releas_1-July-2016.pdf
Statistics South Africa (2018). General Household Survey 2016. Available from: http://www.statssa.gov.za/publications/P0318/P03182018.pdf
Wannell, B. and Thomas, S. (2009) Combining cycles of the Canadian Community Health Survey. Heal Reports, 20(1).

Marcel D. T. Vieira

marcel.vieira@ice.ufjf.br

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Thank you!