



Incorporating a Finite Population Correction into the Variance Estimation of a National Business Survey

Sadeq Chowdhury, AHRQ

David Kashihara, AHRQ

Matthew Thompson, U.S. Census Bureau

FCSM 2018

Disclaimer

The views expressed in this presentation are those of the authors and no official endorsement by the Department of Health and Human Services or the Agency for Healthcare Research and Quality or the U.S. Census Bureau is intended or should be inferred.

- Finite Population Correction (FPC)
- MEPS - Insurance Component (MEPS-IC)
- Variance Estimation in MEPS-IC
- Incorporating FPC in MEPS-IC
- Impact of FPC on MEPS-IC Variance Estimates
- Summary

Finite Population Correction (FPC) in Variance Estimation

Random sampling from an infinite population

$$Var(\hat{x}) = \frac{\sigma_x^2}{n}$$

with Finite Population Correction

$$Var(\hat{x}) = (1 - f) \frac{\sigma_x^2}{n} = \left(1 - \frac{n}{N}\right) \frac{\sigma_x^2}{n}$$

$$FPC = \left(1 - \frac{n}{N}\right)$$

with nonresponse $FPC = \left(1 - \frac{n_r}{N}\right)$

Relevance of Finite Population Correction (FPC)

- ▶ Required when sample size is large relative to the population size, i.e., sampling fraction is non-negligible
- ▶ Ignoring FPC overestimates variance in strata with large sampling fractions
- ▶ Some argue FPC is not required if the inference is intended for a larger population, not for the population in hand
- ▶ In MEPS-IC, since interest is about a particular year or monitoring year to year changes or evaluating the impact of policy changes, FPC is considered appropriate.



Medical Expenditure Panel Survey (MEPS)

- MEPS
 - ▶ A set of three large-scale surveys conducted by AHRQ since 1996
 - ▶ Most complete source of data on cost and use of healthcare and health insurance coverage

Three components:

- Household Component (HC)
 - ▶ Survey of households – families and individuals
- Medical Provider Component (MPC)
 - ▶ Survey of medical providers
- Insurance Component (IC)
 - ▶ Survey of businesses and state & local governments

MEPS Insurance Component (MEPS-IC)

- Annual Survey
 - ▶ Sponsored by AHRQ and conducted by the Census Bureau
 - ▶ To collect data on health insurance plans offered by employers, premiums, contributions, eligibility requirements, and employer characteristics
- Provides employer-based Estimates
 - ▶ Percent offer health insurance
 - ▶ Number of employees enrolled
 - ▶ Health insurance premiums, copays, and deductibles
- Sample ~ 45,000 businesses and state & local governments
 - ▶ ~ 42,000 private sector business establishments (single business entity or location)
 - ▶ ~ 3,000 state & local government agencies

MEPS Insurance Component (MEPS-IC)

- Sampling Frame - List based
 - ▶ Private Sector – Census Bureau’s Business Register
 - ▶ Government Sector – Governments Integrated Directory (GID)
- Sample Design
 - ▶ Stratified single stage sample of establishments/agencies
 - ▶ Within each stratum, mostly EPS (equal probability sampling) in private sector and PPS (probability proportional to size) sampling in government sector
 - ▶ Clustering of plans/sub-agency within an establishment/agency

MEPS-IC Stratification

Private	Certainty	5,000+ employees Railroads	
	Non certainty	<5,000 employees	State Firm Size Estab. Size
State and Local Govt.	Certainty	State Govt. Local Govt. with 5,000+ employees	
	Non certainty	Local Govt. with <5,000 employees	Census Division
		Local Govt. with Missing FTE	

- Produces employer-based estimates at the national & state levels
- No microdata/PUF is released
- Tables of estimates are published annually on the AHRQ web site (400 tables, 115,000 estimates)
- MEPSnet on the AHRQ web site
 - ▶ Uses published table cell estimates to produce
 - Estimates at different levels of aggregation
 - Data trends (1996-2016)
- Microdata (restricted access) at the Census Bureau Research Data Center in Suitland, MD and the University of Maryland Research Data Center in College Park, Maryland.

Variance Estimation in MEPS-IC

Random Group Method

- Random Group Method (used 1996 to 2013)

$$\text{var}(\hat{\theta}) = \frac{1}{k(k-1)} \sum_{\alpha}^k (\hat{\theta}_{\alpha} - \hat{\theta})^2$$

- ▶ Groups assigned sequentially at the time of sampling
- ▶ Estimates from each random group used to compute variance
- ▶ Used by other Census Bureau surveys, more often in the past
- ▶ FPC is not easy to incorporate because random groups consist of cases from multiple strata with different sampling rate

Variance Estimation in MEPS-IC

Taylor Series (TS) Method

- Taylor Series Method (used since 2014)

The variance of an estimator of total

$$\hat{\theta} = \sum_{h=1}^H \hat{\theta}_h = \sum_{h=1}^H \sum_{i=1}^{n_h} w_{hi} x_{hi} = \sum_{h=1}^H \sum_{i=1}^{n_h} \theta_{hi},$$

$$\text{var}(\hat{\theta}) = \sum_{h=1}^H \frac{n_h}{(n_h - 1)} \sum_{i=1}^{n_h} (\theta_{hi} - \bar{\theta}_h)^2$$

$$\text{with } \bar{\theta}_h = \frac{\sum_{i=1}^{n_h} \theta_{hi}}{n_h}$$

with $FPC = (1 - n_{hr}/N_h) = (1 - f_h)$

$$\text{var}(\hat{\theta}) = \sum_{h=1}^H (1 - f_h) \frac{n_h}{(n_h - 1)} \sum_{i=1}^{n_h} (\theta_{hi} - \bar{\theta}_h)^2$$

Effect of FPC in MEPS-IC Noncertainty Strata

- Not included prior to 2016 survey year
- Effect depends on sampling and response rates
- In noncertainty strata, FPC is expected to reduce the variance

Previously, $FPC=1$

Now, $FPC = (1 - n_{hr}/N_h) < 1.0$

- Reduction will be non-negligible if sampling rates in many strata are large

Effect of FPC in MEPS-IC Certainty Strata

- In certainty strata, previously variances were assumed zero
- That means FPC was implicitly used but ignoring nonresponse

$$FPC = (1 - n_h/N_h) = 0, \text{ where } n_h = N_h$$

- Now with nonresponse

$$0 < FPC = (1 - n_{hr}/N_h) \leq 1$$

- FPC will introduce a non-zero variance in certainty strata with nonresponse
- Because it will capture variance due to sample loss for non-response
- Opposite effect: FPC increases variance in certainty strata and decreases in noncertainty strata

Calculation of FPC

- FPC is calculated
 - ▶ using sample & pop counts of
 - sampling units if no clustering
 - PSUs if clustering
 - ▶ within each sampling stratum for noncertainties
 - ▶ within each NR cells for certainties

Private: Certainty and Noncertainty				
Estimation Level	PSU	USU	n_{hr}	N_h
Estab	-	Estab	# of resp. estabs	Pop # of estabs
Plan	Estab	Plans		

Calculation of FPC

Government: Noncertainty				
Estimation Level	PSU	USU	n_{hr}	N_h
Agency	Agency	Sub-agency	# of resp. agencies	Pop # of agencies
Plan	Agency	Plans within sub-agency		

Government: Certainty				
Agency	-	Sub-agency	# of resp. sub-agencies	Pop # of sub-agencies
Plan	Sub-agency	Plans		

Evaluating Effect of FPC on MEPS-IC Variance Estimates

- Analysis using 2015 MEPS-IC Estimates
- Variances for all published estimates
 - ▶ Private/Government, Establishment/Plan, National/State, Totals/Ratios/Percents
- Separate evaluation for certainty/noncertainty within private/government sectors
- Effect is opposite in certainty and noncertainty strata

- Variance estimates with FPC not available until next production cycle
- Used an indirect simple approach to decide if FPC to be incorporated or not
- Noncertainty Strata
 - ▶ Distribution of realized sampling rates
 - ▶ To assess the extent FPC is non-negligible
- Certainty Strata
 - ▶ Distribution of nonresponse rates by sector
 - ▶ To assess the effect of nonresponse on FPC/variance

Evaluation Result

Effect of FPC on Private Sector Noncertainties

Table 1. Distribution of Sampling Rates for Noncertainties in the Private Sector in 2015 MEPS-IC

Weighted Moments			
N	41819	Mean	0.0059
	Level	Quantile	
	100% Max	0.500	
	99%	0.044	
	95%	0.018	
	90%	0.011	
	75% Q3	0.005	
	50% Median	0.003	
	25% Q1	0.002	
	10%	0.002	
	0% Min	0.001	

- About 10% establishments belong to strata with a sampling rate >1% and about 1% belong to a sampling rate >4%
- FPC should make a reduction in variance to some strata

Evaluation Result

Effect of FPC on Government Sector Noncertainties

Table 2. Distribution of Sampling Rates for Noncertainties in the Government Sector in 2015 MEPS-IC

Weighted Moments			
N	2310	Mean	0.034
	Level	Quantile	
	100% Max	0.573	
	95%	0.135	
	90%	0.086	
	75% Q3	0.039	
	50% Median	0.015	
	25% Q1	0.006	
	10%	0.004	
	0% Min	0.002	

- About 50% estabs belong to strata with sampling rate >1% and 10% belong to strata with sampling rate >8%
- FPC will reduce variance in these strata

Evaluation Result

Effect of FPC on Private Sector Certainties

Table 3. Distribution of Nonresponse Rates of Certainties in the Privates Sector in 2015 MEPS-IC

Weighted Moments			
N	137	Mean	0.436
	Level	Quantile	
	100% Max	0.868	
	95%	0.750	
	75% Q3	0.589	
	50% Median	0.398	
	25% Q1	0.303	
	10%	0.221	
	5%	0.107	
	0% Min	0	

- 75% establishments have nonresponse rate > 30%, and 95% have nonresponse rate > 10%
- Therefore, FPC will introduce a non-zero variance in almost all private sector certainty strata

Evaluation Result

Effect of FPC on Government Sector Certainties

Table 4. Distribution of Nonresponse Rates of Certainties in the Government Sector in 2015 MEPS-IC

Weighted Moments			
N	423	Mean	0.196
	Level	Quantile	
	100% Max	0.667	
	95%	0.417	
	90%	0.417	
	75% Q3	0.247	
	50% Median	0.203	
	25% Q1	0.148	
	10%	0	
	0% Min	0	

- Almost 75% establishments have nonresponse rate > 14%
- FPC will introduce a non-zero variance in most strata

Summary

- FPC was not incorporated in MEPS-IC until 2016
- Incorporated FPC appropriately in different strata consistent with different sampling procedures
- Evaluation shows incorporating FPC will have a noticeable impact on the variances in many strata
- In many noncertainty strata, FPC will reduce variance significantly as sampling rate is high
- In certainty strata, previously variance assumed to be zero ignoring nonresponse
 - ▶ FPC is now used to decide if the variance is zero or non-zero based on nonresponse rate
 - ▶ FPC will add some variance in most certainty strata

Summary (cont.)

- However, since published estimates are generally based on contributions from both certainty and noncertainty strata with opposite effect of FPC, the resulting effect may be negligible
- Other recent improvements in variance estimation methodology may also have an impact on variance which are not considered as part of this evaluation
- More accurate assessment can be made when FPC is included in the next production cycle

References

- Chowdhury, S. (2017). “Incorporating FPC in MEPS-IC Variance Estimation”.
- Thompson, M. (2017). Internal U.S. Census Bureau documents.
- Wolter, K.M. (1985). Introduction to Variance Estimation. New York: Springer-Verlag.
- SAS Stat User’s Guide (2012)
- SUDAAN Technical Manual (1996).

Thank you!

Sadeq.chowdhury@ahrq.hhs.gov