Institute of Education Sciences
National Center for Education Statistics

NATIONAL INSTITUTE OF STATISTICAL SCIENCES
TECHNICAL FORUM

COORDINATING DESIGNS FOR MULTIPLE SURVEYS

TABLE OF CONTENTS

NATIONAL INSTITUTE OF STATISTICAL SCIENCES

COORDINATED SAMPLE DESIGN FOR MULTIPLE NCES SURVEYS

EXECUTIVE SUMMARY

The National Center for Education Statistics (NCES) and other Centers within the Institute of Education Sciences (IES) collect data on a national scale from districts, schools, and individual administrators, teachers, students and parents. The surveys, assessments and other studies vary widely in purpose and scope, but all contribute to the information available about individual schools, districts and states. In addition to creating databases for individual data collections, this compiled information is used in turn for designing future samples.

In 2020, NCES commissioned the National Institute of Statistical Sciences (NISS) to assemble a pair of panels of technical experts, *Post-COVID Surveys* and *Setting Priorities for Federal Data Access to Expand the Context for Education Data* to consider opportunities for changing the sampling paradigm and process. From the NCES point of view dual goals were to address the rising nonresponse and lack of participation and to enrich the information base by linking data collections. From the points of view of school districts, schools and participants, the need was to understand and alleviate the burden of participation. These two panels became the first in a series to examine the principal issues in greater detail. The next pair of panels, *Connecting the Dots, I & II* examined the technical design and the implementation issues of coordinating the sampling across multiple data collections.

This *FORUM* responds to the findings and recommendations of those four panels. The broad recommendations of the first two panels were to develop a combined approach for data collections during each academic year and to expand the information base while gaining efficiency by linking data and eliminating redundant requests. The second pair of panels focused on identifying specific steps required to achieve the original goals for NCES and for the education community.

The goal of NCES-sponsored research (through NISS) to further advance innovations required by a coordinated sampling approach was to define the statistical challenges, identify the critical points for technical solution, and examine the several posited strategies for solution. This *FORUM* brought together statistical experts and NCES staff for presentations of the preliminary work on fleshing out the most promising sampling paradigms from *Connecting the Dots, I & II*. Following those presentations, a technical working session engaged additional statistical experts and NCES staff in formal and informal discussion.

The first *FORUM* presentation set the context with a brief outline of the progression of panels leading to this point. The two presentations that followed outlined different strategies for developing a new class of design approaches for coordinating multiple surveys with varied objectives.

Statistics: The Problem Space

A sequence of expert panels on design innovation, technical issues and the challenges to implementation led to the research into new design strategies that are the focus of this FORUM. These panels set the context, *i.e.*, motivation, rationale and constraints, that define statistical challenges for coordinated sampling designs. Several specific strategies were proposed for further development and consideration.

Some Topics for Coordinated Sampling of Schools over Time

Sampling design for longitudinal surveys is tricky as we wish to balance the trade-off between the response burden and the statistical efficiency. To reduce the response burden, it is desirable to develop a negatively coordinated sampling design. We propose a modification of the Swiss method to allow for differential coordination for each sampling unit. In addition, statistical methods for imposing balancing conditions in planned missingness and adaptive sample size allocation for stratified random sampling are discussed.

Coordinating Sample Design for Multiple NCES Surveys

We have developed four strategies for a coordinated sampling process to potentially reduce response burdens: 1) Independently select schools for each survey, compute the burden for each selected school, and randomly substitute schools from the same stratum. 2) Independently select schools for each survey, compute the burden for each selected school, and reject samples that exceed the burden, 3) Sequentially sample schools based on a random survey order and decrease the selection probability for schools selected in previous surveys, and 4) Use matrix sampling of to assign surveys to schools using a probabilistic mechanism, *i.e.*, create replicates.

Discussion of technical points opened with formal comments from invited discussants. This working session continued engaging NCES staff, presenters and discussants in open discussion of technical points, specific applicability to NCES studies, and requirements for moving forward. The session culminated in five recommendations for immediate next steps.

Recommendations

- Pursue multiple promising strategies simultaneously, developing each selected methodology sufficiently to permit validation and feasibility testing of implementation, and to allow (preliminary) estimation of precision and bounds on bias. Strategies deserving of consideration for development include strategies derived from block design of experiment from a common sampling frame, also Bayesian design incorporating prior or non-probability information, a modified Swiss method with balance, and matrix sampling that creates separate sampling frames.
- Create both simulation and realistic test files and resources for evaluation of strengths, vulnerabilities and capacity for scaling up with extension to additional IES/Department of Education studies and surveys.
- Initiate consolidation of stratum definitions and items to enable data sharing for sampling design, reduce redundant data requests, and facilitate efficient, balanced sampling designs.
- Expand the detail available for CCD schools; for sampled schools collect data on participation history (e.g., agreement/refusal with rationale at each level of decision-making) as a basis for designing samples for future studies, and for research purposes.
- Remain open to continuing innovation of sampling strategy, especially to take advantage of alternative data sources (e.g., non-probability online data sources), to adjust selection probability (e.g., burden metrics) and to use model-based prediction (e.g., propensity for nonresponse, magnitude of potential contribution to bias).

PREFACE

The origins of this *FORUM on Coordinating Designs for Multiple Studies*, came from the National Center for Education Statistics (NCES) commission to the National Institute of statistical Sciences (NISS) to *convene an expert panel to explore the opportunities and the expectations for surveys set to be conducted following emergence from the Covid period*. Specifically that panel was asked to place the new needs for information on education post-COVID into the context of contemporary media awareness and capability while addressing the changing modes and roles of education. At the same time, the panel was to take cognizance of the rising rates of refusals and nonresponse at multiple levels from state offices of education to individual participants.

The central concept from this 2020 panel's deliberations and discussions with NCES staff were for NCES/IES to:

> *Implement a recognized functionally coherent and transparent structure to replace/ reorganize the loosely connected collection of separate surveys and assessments.*[1]

In August 2021, NCES again charged NISS with assembling experts to work with NCES staff to formulate a clear objective and a strategy for restructuring multiple surveys into a coordinated process and then to weigh the requirements and the merits of going forward. This charge was broken into two parts, *Connecting the Dots I & II.* On the technical side, Part I, any proposed solution had to simultaneously meet the statistical requirements for each individual survey, facilitate management of sample duplications and enable data leveraging across surveys to improve estimation. With respect to implementation, Part II addressed the feasibility and identification of critical points in the process of implementation of a consolidated/coordinated approach encompassing multiple surveys.

At the beginning of 2022, the positive reports of both parts of *Connecting the Dots I & II*, led to NCES support, through NISS, of preliminary research into the technical issues posed by the heterogeneous requirements, constraints and designs of the individual studies to be coordinated.

At this *FORUM on Coordinating Designs for Multiple Studies*, this preliminary work on several specific strategies was presented with critical commentaries by statistical and survey design experts. A technical working session with experts and NCES staff followed where discussion focused on the research efforts necessary to develop any of these approaches into implementable form.

[1] Report of NCES-NISS Panel on Post COVID Surveys.

NATIONAL INSTITUTE OF STATISTICAL SCIENCES
TECHNICAL EXPERT PANEL

NISS/NCES FORUM AND TECHNICAL WORKSHOP AGENDA

FORUM ON COORDINATING DESIGNS FOR MULTIPLE SURVEYS

AGENDA

Monday, May 9, 2022

PART I:	**10:00 am – Noon**	
10:00 – 10:10 am	Welcome and Introductions	
10:10 – 10:40 am	Setting the Stage for Coordinated Sampling	
	Why? NCES Objectives	Peggy Carr
	Statistics: The Problem Space	Nell Sedransk
10:40 – 11:10 am 11:10 – 11:50 am	Coordinating Sample Design for Multiple NCES Surveys Questions	Trivellore Raghunathan Yajuan Si
	Some Topics for Coordinated Sampling of Schools over Time Questions	Jae-kwang Kim
BREAK	12:30 – 1:20 pm	
PART II	**1:20 – 3:30 pm**	
1:20 – 3:00 pm	Weighing the Comparative Feasibility and Relative Merits	
	Discussants:	Mary Thompson Jerry Reiter Lynne Stokes
Discussion – Open		Jay Breidt
3:00 – 3:30 pm	Specific Design Comments Summary Comparison Implications for Implementation	
3:40 – 4:00 pm	Discussion and Summary Steps and Requirements for Development	

INTRODUCTION

The National Center for Education Statistics (NCES) and other Centers within the Institute of Education Sciences (IES) collect data on a national scale from districts, schools, and individual administrators, teachers, students and parents. The surveys, assessments and other studies vary widely in purpose and scope, but all contribute to the information available about individual schools, districts and states. In addition to creating databases for individual data collections, this compiled information is used in turn for designing future samples.

The origins of this *FORUM on Coordinating Designs for Multiple Studies*, came from the National Center for Education Statistics (NCES) commission in 2020 to the National Institute of Statistical Sciences (NISS) to convene the first in a series of now six expert panels to place the new needs for information on education post-COVID into the context of contemporary media awareness and capability while addressing the changing modes and roles of education. At the same time, the panel was to take cognizance of the rising rates of refusals and nonresponse at multiple levels from state offices of education through to individual participants. The primary concept, as formulated by the first of these panels is for NCES/IES to:

> *Implement a recognized functionally coherent and transparent structure to replace/ reorganize the loosely connected collection of separate surveys and assessments.*

Subsequent panels in 2021 focused on feasibility and defining the entire data collection process from sample design through recruitment to survey implementation and data recording. Separate groups of experts have been assembled to examine the technical issues and the challenges to implementation. Following these panels' deliberations, the next step is more detailed examination of the challenges posed in coordinating multiple surveys with heterogeneous content requirements, partially overlapping populations, and differing constraints as goals.

Several statistical strategies were proposed for investigation. Early in 2022, preliminary research was initiated through NISS with support from NCES. This FORUM presents the context for NCES data collections and outlines the rationale and motivation for considering these strategies. The goal of the FORUM is to engage both statistical experts and NCES staff in discussion of the options and the realistic constraints and to arrive at a list of the next steps to develop promising approaches into implementable form.

The primary presentations follow together with the discussant's formal commentaries and the recommendations reached at the working session.

PART I

APPROACHES TO THE PROBLEM: TECHNICAL ISSUES AND PATHS TO SOLUTIONS

NATIONAL INSTITUTE OF STATISTICAL SCIENCES

COORDINATED SAMPLE DESIGN FOR MULTIPLE NCES SURVEYS

STATISTICS: THE PROBLEM SPACE

Nell Sedransk, PhD
National Institute of Statistical Sciences

ABSTRACT

A sequence of expert panels on design innovation, technical issues and the challenges to implementation led to the research into new design strategies that are the focus of this FORUM. These panels set the context, i.e., motivation, rationale and constraints, that define statistical challenges for coordinated sampling designs. Several specific strategies were proposed for further development and consideration.

Setting the Challenge

- Post-Covid Surveys & Setting Priorities for Federal Data Access: Expand the Context for Education Data
- Context
- Paradigm
- Connecting the Dots, I & II

- School Survey Participation and Burden
- Forum on Coordinated Designs for Multiple Surveys

Two Ideas: Streamlining recruitment for NCES studies
Consolidating and sharing information

Post-Covid Surveys – December 2020 Technical Feasibility

- Implement a recognized functionally coherent and transparent structure to replace/reorganize loosely connected collection of separate surveys and assessments
- Consolidate recruitment with sole recruiter for all surveys and assessments negotiate for program rather than individual survey/assessment.

Setting Priorities . . . To Expand the Context for Education Data – March 2021

- A Integrate data from other internal and external sources with NCES data to inform critical national issues involving education.
- Create a resource set of files of contextual information as a "backbone" to be consistent over time and content to link among multiple surveys, assessments and other data collections.

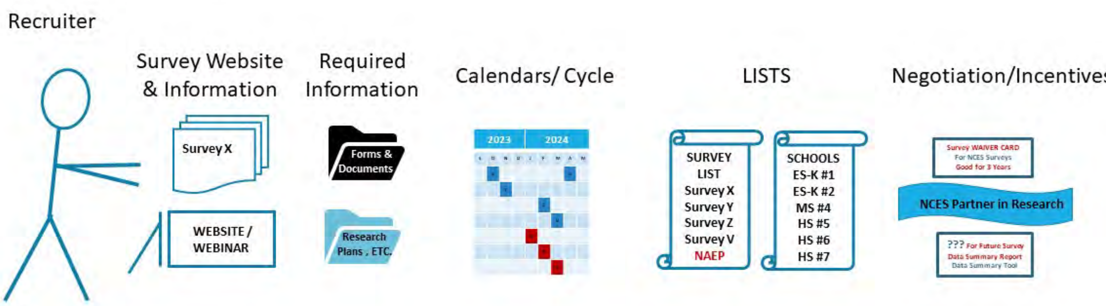
Context from Various Perspectives

- NCES
  - Declining response rates.
  - Non-Response Bias Assessment and Adjustment Review: Further improvement will depend on more informative covariates
- Schools and Districts
  - Over-reliance on specific schools
  - Poorly planned timing relative to school operations
  - Burden – either immediate or accumulated - outweighing (local) benefit
- State and District Administrators
  - Unorganized series of participation requests from IES/NCES
  - Redundant negotiations/solutions to administrative and legal issues
  - Enormous load of total requests for participation (all sources)
- Contractors and Recruiters
  - Lack of recognition or priority for NCES over other proposed studies

Paradigm

- Single Recruiter
- Complete Program of Surveys
  - Including NAEP, PISA, etc.
  - Calendar for upcoming academic year
  - Plans for at least 1 more year
- Survey Information
  - Requirements
  - Burden
- Sampling Frame with history of district and school participation

Recruitment Paradigm



Two Questions: Can it be done (statistically)? What will it take to do it?

- ***Connecting the Dots, I – September 2021 Technical Feasibility***
- ***Connecting the Dots, II – November 2021 Implementation Issues***

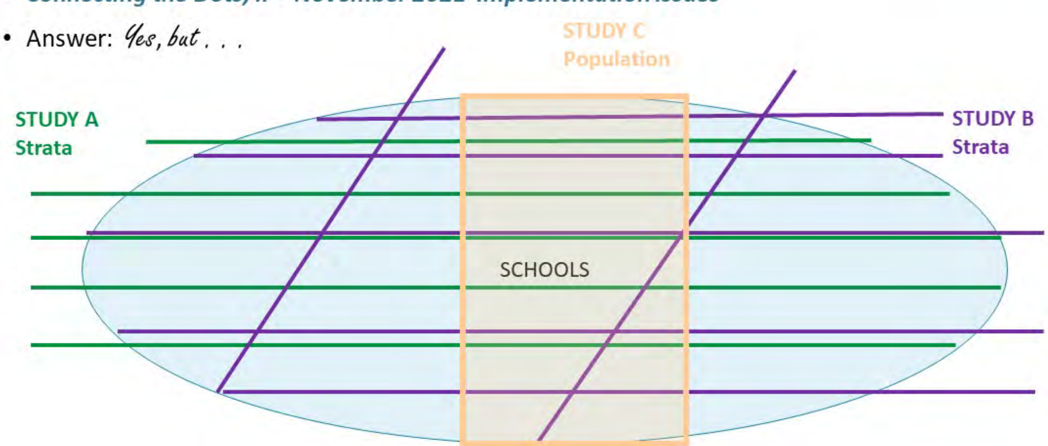
- Answer: *Yes, but . . .*

Constraints on Solution

- Enable data sharing across surveys; eliminate redundancy wherever possible
- Allow separate participation decision for each study
- Avoid overlap among samples or manage overlap to leverage information
- Accommodate heterogeneity of studies
  - Assessments, student surveys, education personnel (no students) surveys, administrative records
  - Cross-section or longitudinal
  - Population specifications – estimates required
- Assure individual study integrity
  - No priority or rank-ordering of studies
  - Individual study requirements and goals met
  - Statistically sound estimates, variance estimates, bias assessment and adjustment
- Facilitate continuing design innovation

Two Questions: Can it be done (statistically)? What will it take to do it?

- ***Connecting the Dots, I – September 2021 Technical Feasibility***
- ***Connecting the Dots, II – November 2021 Implementation Issues***
- Answer: *Yes, but . . .*



Four Approaches

Separate Sequential Sampling

- Structure – minimal embedded common structure across surveys
  - Surveys in sequence each draw samples according to individual plans
  - Overlaps (schools selected more than once) are assigned to one or more surveys

Total Combined Sample

- Structure – complex common structure can be embedded across surveys
  - Total sample based on combined (across surveys) sample configuration estimate
  - Allocation to individual surveys (multiple possibilities)
    - Total sample comprised of large number of “mini replicates” (across or within stratum) to be selected at random as required by individual surveys
    - Random allocation within stratum based on individual survey requirements

Four Approaches

Total Combined Sample – with blocking

- Structure – complex common structure can be embedded across surveys
  - Total sample based on combined (across surveys) sample configuration estimate
  - Allocation to individual surveys (multiple possibilities)
    - Combined fine structure stratification
    - Incomplete block framework for allocation to surveys
    - Random allocation within stratum based on individual survey requirements

Partitioned Population Samples

- Structure – limited common structure across surveys
  - Population partitioned at random into segments
  - Segments assigned to one survey (or more than one if overlap is intended)
  - Surveys draw samples independently from their segments

COORDINATING SAMPLE DESIGNS: STATISTICAL STRATEGIES

Trivellore Raghunathan, PhD, Yajuan Si, PhD, Mike Elliott, PhD, Rod Little, PhD
University of Michigan

ABSTRACT

We consider six exemplar surveys that use the Common Core Data (CCD) as the sampling frame of U.S. public schools. Taking into account the survey stratification and eligibility restriction across surveys, we focus on the joint select probability estimation for each school to be selected for all surveys subject to the response burden.

We have developed four strategies for a coordinated sampling process to potentially reduce response burdens: 1) Independently select schools for each survey, compute the burden for each selected school, and randomly substitute schools from the same stratum. 2) Independently select schools for each survey, compute the burden for each selected school, and reject samples that exceed the burden, 3) Sequentially sample schools based on a random survey order and decrease the selection probability for schools selected in previous surveys, and 4) Use matrix sampling of to assign surveys to schools using a probabilistic mechanism, i.e., create replicates. Strategies 1) and 2) offer flexibility and independence for individual surveys. Strategy 3) involves the ordering of surveys at random and changing the selection probabilities to minimize overlap. Strategy 4) is potentially the most statistically efficient, enhances the utility of variables collected across surveys, and enhances the information for each school in CCD using modeled estimates.

None of these methods will work without the “centralization” of sampling activities. Simulation studies are in demand to compare different strategies for the improvement of data quality and implementation feasibility.

Background

- The National Center for Education Statistics (NCES) annually conducts various surveys and assessments
  - Studies are of multiple types: longitudinal and cross-sectional
  - Students, teachers and administrators as respondents
  - The National Assessment of Education Progress (NAEP) is the largest and the only K-12 survey/assessment with mandated participation.
  - Others rely on voluntary participation from states, school districts, schools, teachers, parents, and children.
- Currently, NCES contractors independently develop the design, conduct surveys/assessments, and contact each office in the approval chain.
- Declining response rates and operational issues call for an integrated approach to multiple surveys

Sampling Frame

- Most surveys use CCD (Common Core Data) as the sampling frame and the primary database on public elementary and secondary education in the U.S.
- Every year the CCD collects information from the universe of state education agencies on all public elementary and secondary schools and education agencies in the U.S. (E.g., 99,763 schools collected in 2020-2021).
- The CCD provides descriptive data about staff and students at the school, school district, and state levels.

Updating the CCD

- Of the administrative information surveys, for public K-12 schools, information (updated annually) for the CCD is mandatory.
- Updating of the CCD occurs when samples drawn for NCES surveys/assessments, including NAEP, lead to corrections and whenever a school or district volunteers new information.
- It is desirable that information be updated on a continuous basis as it becomes available.

Six Exemplar Surveys

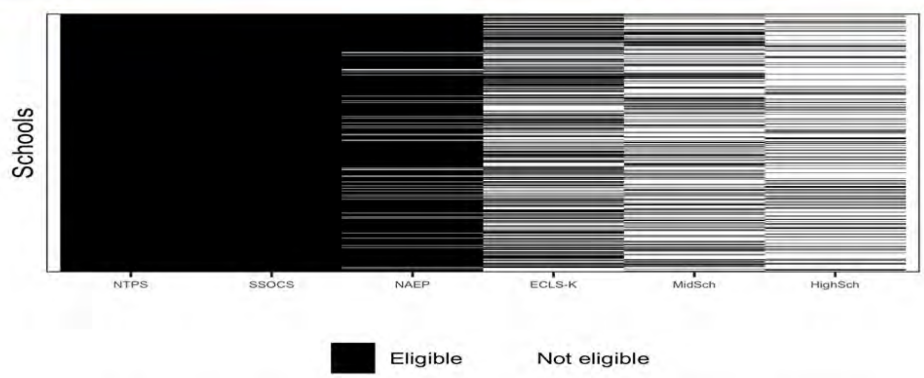
Survey (#sampled public schools)	Grades/Respondents	Time Frame
NAEP (8,300 for Grade 4 Reading and Math)	Grades 4, 8, 12	Every 2 years
Early Childhood Longitudinal Study-Kindergarten (ECLS-K) (1,040)	Grades K-5	2011-16, Future?
Middle Grades Longitudinal Study (MidSch) (3,710)	Grades 6 - 8	2017-18, Future?
High School Longitudinal Study (HighSch) (1,890)	Grades 10 -12	2009, Future?
National Teacher and Principal Survey (NTPS) (10,600)	All public schools	2015-18, Future?
School Survey on Crime and Safety (SSOCS) (4,800)	All Public elementary and secondary schools	2017-18 Future?

Objectives

To develop strategies for a coordinated sampling process

1. To coordinate the process of approaching the schools for participation
2. To potentially reduce burden on decision-makers and respondents and possibly decrease refusal/nonresponse
3. To exploit the variables collected across surveys to leverage data from multiple surveys

School Eligibility for Selected Surveys



#Surveys	2	3	4	5	6
#Schools	2469	9794	64091	19363	3593

Stratification

1. NAEP: urbanization, enrollment of Black, Hispanic, or other race/ethnicity students, state-based achievement scores, and median household income
2. ECLS-K: PSUs (counties) are stratified by Metropolitan Statistical Area (MSA) status, census geographic region, size class (defined using the measure of size), per capita income, and the race/ethnicity of 5-year-old children residing in the PSU (specifically the percent of 5-year-old APIs, the percent of 5-year-old Blacks, and the percent of 5-year-old Hispanics)
3. MidSch: school type (public, Catholic, other private), region (Northeast, Midwest, South, and West), and prevalence of students with disabilities
4. HighSch: none
5. NTPS: none
6. SSOCS: school level, locale, and enrollment size

*Note: factors marked in blue are available in the publicly accessible 2020-21 CCD

Statistical Problem

$s = survey$

$i = school$

$h = stratum$

- Allow stratification differs across surveys
- Some common strata across all surveys
- Individual surveys may differ on oversampling needs

$$I_{ihs} = 1 \text{ if school } i = 1,2,\ldots,N_h, h = 1,2,\ldots,H \text{ is selected for survey } s = 1,2,\ldots,S$$

$$= 0 \text{ otherwise}$$

$$\pi_{ihs} = \text{initial inclusion probability}$$

$$\pi_{ihs} = 0 \text{ (if school } i \text{ in stratum } h \text{ is not eligible for survey } s\text{)}$$

Statistical Problem (Continued)

Burden:

b_{ihs} = *burden if school i in stratum h is selected for survey s*

Total Burden:

$$\sum_{s=1}^{S} b_{ihs} I_{ihs} = b_{ih+}$$

Examples:

$b_{ihs} = 1$

$b_{ihs} = k_{ih} + 1, k_{ih}$ = *number of asks already on the books*

Goal:

$\Pr(I_{ih1}, I_{ih2}, \ldots, I_{ihS} \mid b_{ih+} \le b_h)$

and

Compute the final sampling weight

$w_{ihs} = 1/\Pr(I_{ihs} = 1)$

Four Strategies

1. Independent selections, substitution to reduce burden
2. Independent selections, reject samples exceeding the burden
3. Sequential sampling
4. Matrix sampling

Strategy 1 (Independent Selection, Substitution)

- Independently select schools for each survey
- Compute the burden for each selected school
- Randomly replace/substitute the schools from the same stratum to reduce the burden
- Example
    - Maximum ask is 3
    - A school is selected for four surveys
    - Choose a survey at random, replace a random school from the pool of non-sampled schools from the same stratum
- Extension of Kish and Scott (1971, JASA) idea.
- Substitution may not be possible, if maxed out. Higher burden need to accepted for some schools

Strategy 2 (Independent Selection, Rejection)

- Independently select schools for each survey
- Compute the burden for each selected school
- Reject samples that exceed the burden
- Directly draw from the joint distribution of S dummy variables subject to the burden constraint
- If too many samples are rejected then burden constraint needs to be adjusted

Strategy 3 (Sequential Sampling)

- Choose a random survey order
- Select schools for the first survey
- Select for the second survey, by decreasing the selection probability for those selected in the first survey
- Select for the third survey, decreasing the probability for those selected in the first two surveys
- After S selections, compute the burden
  - Substitute to reduce or
  - Reject the samples
- Random order of surveys in each stratum

Strategy 4 (Matrix Sampling)

- Treat all surveys as "splits" of one large survey
- Use split questionnaire or matrix sampling design to assign splits to the schools using a probabilistic mechanism, i.e., create replicates

Random school	Assignment 1	Assignment 2
1	S1	S6
2	S4	S2
3	S5	S3
4	S2	S6
5	S2	S1
6	S1	S3
7	S2	S3

Incomplete block design with schools as "blocks", "slots" as the number surveys slots open in the school and surveys or splits as the "treatment"

Computing Sampling Weights

- Analytical computation of the ultimate probability that a school is included in a survey
- Monte Carlo Method
  1. Draw a very large number of samples (the joint distribution of S sampling indicators) using any given strategy
  2. Compute the fraction of the samples where the unit was included
  3. The inverse of the fraction is the sampling weight
  4. Adjust the initial selection probability to achieve the desired sampling fraction in each stratum or to control the distribution of the weight
  5. Once settled, choose a random realization from Step 1

Discussion

- None of these methods will work without "centralization" of sampling activities.
- A common core of stratification variables will help in the coordination.
- Strategies 1 and 2 offer flexibility and independence for individual surveys.
- Strategy 3 involves ordering of surveys at random and changing the selection probabilities to minimize overlap.
- Strategy 4 is probably the most statistically efficient, enhances the utility of variables collected across surveys, and enhances the information for each school in CCD using modeled estimates.
- All these strategies are designed to reduce the burden (negative correlation among the sampling indicators). Is there a benefit in going to smaller number of schools for a larger number of surveys? (positive correlation among the sampling indicators)

SOME TOPICS FOR COORDINATED SAMPLING OF SCHOOLS OVER TIME

Jae-kwang Kim, PhD
Iowa State University

ABSTRACT

We consider three topics for coordinated sampling over time. The first topic is the sampling design for negatively coordinated longitudinal surveys. Sampling design for longitudinal surveys is tricky as we wish to balance the trade-off between the response burden and the statistical efficiency. To reduce the response burden, it is desirable to develop a negatively coordinated sampling design with known measures of size. Permanent random numbers play an important role in controlling the coordination of the longitudinal surveys. We propose a modification of the Swiss method to allow for differential coordination for each sampling unit.

The second topic is about imposing balancing conditions in planned missingness. Achieving balancing conditions at the design stage can be achieved by employing the maximum entropy sampling design with the balancing constraints. At the estimation stage, the empirical likelihood method can be used to impose balancing conditions.

The third topic is about adaptive sample size allocation for stratified random sampling. The basic idea is to adopt the Huntington-Hill algorithm to select the sample elements sequentially and update the stratum parameters using the current observations. The updated parameters are used to recompute the priority values for applying the Huntington-Hill algorithm. The stratum with the largest priority values is selected in the adaptive stratified sampling.

- Introduction to NRI (National Resources Inventory) Sampling Design and Estimation
- Sampling Design for Negatively Coordinated Longitudinal Surveys
- Imposing Balancing Condition in the Planned Missingness
- Adaptive Sample Size Allocation for Stratified Sampling
- Concluding Remarks

Introduction to NRI (National Resources Inventory) Sampling Design and Estimation

Introduction to NRI

NRI survey

- Large-scale cross-sectional and longitudinal survey of land use and natural resources conditions and trends on US non-federal lands
- Two-phase sampling
  1. Phase one: 1997 NRI
  2. Phase two: annual NRI
- Multi-mode data collection
  - Photo-interpretation
  - Auxiliary materials
  - Local NRCS
- Multi-purpose survey

Kim (ISU) Sampling designs 05/09/2022 4 / 41

Introduction to NRI

Sampling in time

97	00	01	02	03	04	05	06	07	08	09	10	11	12
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
✓		✓											
✓			✓								✓		
✓				✓									
✓					✓		✓					✓	
✓						✓		✓					✓
✓									✓				
✓										✓			
✓													
✓													

Kim (ISU) Sampling designs 05/09/2022 5 / 41

Generalized least squares (GLS) estimation

- Best estimates: Gold standard for estimation
  - Control totals for weight processing
  - Checks for imputation
  - Basis for replication variance estimation
- Variables for GLS estimation
  - Segment data: small water bodies, small streams, small built-up, large urban and built-up, public roads. (called Control coveruse)
  - Point data: Some coveruse indicators (cultivated cropland, noncultivated cropland, pastureland, rangeland, forestland, barren/other rural land, CRP, and other land in farms), WEQ, and aquahab codes.

Generalized least squares (GLS) estimation

- For example, consider

	00	01	02	03
P0(core)	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$
P1		$\checkmark$		
P2			$\checkmark$	
P3				$\checkmark$
	θ_{00}	θ_{01}	θ_{02}	θ_{03}

where θ_t=population total at time t
- 7 estimators for 4 parameters:

$$\hat{Y} = (\hat{Y}_{0,00}, \hat{Y}_{0,01}, \hat{Y}_{1,01}, \hat{Y}_{0,02}, \hat{Y}_{2,02}, \hat{Y}_{0,03}, \hat{Y}_{3,03})^T$$

where $\hat{Y}_{p,t}$=panel total estimate from panel p at time t $E(\hat{Y}_{p,t}) = \theta_t$

Estimates

The model is

$$\hat{Y} = X\theta + e$$

where X is a matrix linking panel total estimates to the population total θ and E(e)=0.

$$\begin{pmatrix} \hat{Y}_{0,00} \\ \hat{Y}_{0,01} \\ \hat{Y}_{1,01} \\ \hat{Y}_{0,02} \\ \hat{Y}_{2,02} \\ \hat{Y}_{0,03} \\ \hat{Y}_{3,03} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \theta_{00} \\ \theta_{01} \\ \theta_{02} \\ \theta_{03} \end{pmatrix} + \begin{pmatrix} e_{0,00} \\ e_{0,01} \\ e_{1,01} \\ e_{0,02} \\ e_{2,02} \\ e_{0,03} \\ e_{3,03} \end{pmatrix}$$

Estimates

- GLS estimate for θ:

$$\hat{\theta} = (X^T V^{-1} X)^{-1} X^T V^{-1} \hat{Y}$$

- Thus, the GLS estimator combines the information.
- Different choice of V makes different GLS estimator.
- For reference, see Breidt and Fuller (1999).

Sampling Design for Negatively Coordinated Longitudinal Surveys

Sampling design for negatively coordinated longitudinal surveys

Sampling design in longitudinal surveys

- Wish to balance the trade-off between the survey frequency and the response burden.
- Survey frequency can be determined by the following factors:
  1. Response burden
  2. Dynamics (temporal variability)
- Coordination is an important concept for business surveys.
  - Negative coordination: the number of common units between successive samples through time in a population is minimized.
  - Positive coordination: the number of common units is maximized.
- Measure of Size (MOS) is available for each sampling unit. The MOS may change over time. We wish to select the samples using the selection probability proportional to MOS.
- Permanent Random Number (PRN) plays an important role to control the coordination of the longitudinal samples.

Kim (ISU) Sampling designs 05/09/2022 11 / 41

Sampling design for negatively coordinated longitudinal surveys

Poisson method with permanent random numbers (Brewer et al., 1972)

- Let $\pi_k^{(t)}$ be the inclusion probability of unit k in year t.
- Wish to select negatively coordinated Poisson samples with inclusion probabilities $\pi_k^{(t)}$
  - In the first wave, generate $u_k^{(1)}$ from $U(0,1)$ distribution. If $u_k^{(1)} < \pi_k^{(1)}$ then unit k is selected in S_1. The PRNs are modified to

$$u_k^{(2)} = \begin{cases} u_k^{(1)} - \pi_k^{(1)} + 1 & \text{if } k \in S_1 \\ u_k^{(1)} - \pi_k^{(1)} & \text{otherwise.} \end{cases}$$

  - In the second wave, if $u_k^{(2)} < \pi_k^{(2)}$ then unit k is selected in S_2. The PRNs are modified to

$$u_k^{(3)} = \begin{cases} u_k^{(2)} - \pi_k^{(2)} + 1 & \text{if } k \in S_2 \\ u_k^{(2)} - \pi_k^{(2)} & \text{otherwise.} \end{cases}$$

Kim (ISU) Sampling designs 05/09/2022 12 / 41

Check

- Note that

$$\begin{aligned} P(u_k^{(2)} < \pi_k^{(2)}) &= P(u_k^{(2)} < \pi_k^{(2)} \mid k \in S_1) P(k \in S_1) \\ &\quad + P(u_k^{(2)} < \pi_k^{(2)} \mid k \notin S_1) P(k \notin S_1) \\ &= P(u_k^{(1)} - \pi_k^{(1)} + 1 < \pi_k^{(2)}) \pi_k^{(1)} \\ &\quad + P(u_k^{(1)} - \pi_k^{(1)} < \pi_k^{(2)})(1 - \pi_k^{(1)}) \\ &= \left(\pi_k^{(1)} + \pi_k^{(2)} - 1\right) \pi_k^{(1)} + \left(\pi_k^{(1)} + \pi_k^{(2)}\right) \left(1 - \pi_k^{(1)}\right) \\ &= \pi_k^{(2)} \end{aligned}$$

- Similarly, we can show $P(u_k^{(3)} < \pi_k^{(3)}) = \pi_k^{(3)}$.

The Swiss method

- Selection interval concept is used to obtain positive or negative coordination sample.
- For illustration, suppose that we have $T = 2$ years, with longitudinal inclusion probabilities $\pi_k^{(1)}$ and $\pi_k^{(2)}$. Let u_k be the PRN generated from $U(0,1)$ distribution.
- For positive coordination, we select $k \in S_t$ if $u_k \in [0, \pi_k^{(t)}]$.
- For negative coordination, we select samples as follows:
  - For $t = 1$, k is selected if $u_k \in [0, \pi_k^{(1)}]$.
  - For $t = 2$, k is selected if

$$u_k \in [\pi_k^{(1)}, \min\{\pi_k^{(1)} + \pi_k^{(2)}, 1\}] \cup [0, \max\{0, \pi_k^{(1)} + \pi_k^{(2)} - 1\}].$$

- Introduced by Graf and Qualité (2014).

Sampling design for negatively coordinated longitudinal surveys

The Swiss method (cont'd)

For negative coordination,

Case	Selection Interval	$k \in S_1$	$k \in S_2$
$\pi_k^{(1)} + \pi_k^{(2)} \leq 1$	$[0, \pi_k^{(1)}]$	1	0
	$[\pi_k^{(1)}, \pi_k^{(1)} + \pi_k^{(2)}]$	0	1
	$[\pi_k^{(1)} + \pi_k^{(2)}, 1]$	0	0
$\pi_k^{(1)} + \pi_k^{(2)} > 1$	$[0, \pi_k^{(1)} + \pi_k^{(2)} - 1]$	1	1
	$[\pi_k^{(1)} + \pi_k^{(2)} - 1, \pi_k^{(1)}]$	1	0
	$[\pi_k^{(1)}, 1]$	0	1

The location of u_k determines the sample selection status of unit k for each year.

Sampling design for negatively coordinated longitudinal surveys

Proposal

- Idea: Extend the Swiss method to allow for differential coordination for each sampling unit.
- Let $\rho_k \in (0, 1)$ be the sampling coordination parameter that determines overlap across years.
    - $\rho_k = 0$: maximum overlap (no response burden)
    - $\rho_k = 1$: minimum overlap (high response burden)
- Selection interval is a function of ρ_k

Illustration (T=2)

Case	Selection Interval	$k \in S_1$	$k \in S_2$
$\rho_k \pi_k^{(1)} + \pi_k^{(2)} \leq 1$	$[0, \pi_k^{(1)}]$	1	0
	$[\rho_k \pi_k^{(1)}, \rho_k \pi_k^{(1)} + \pi_k^{(2)}]$	0	1
	$[\rho_k \pi_k^{(1)} + \pi_k^{(2)}, 1]$	0	0
$\rho_k \pi_k^{(1)} + \pi_k^{(2)} > 1$	$[0, \rho_k \pi_k^{(1)} + \pi_k^{(2)} - 1]$	1	1
	$[\rho_k \pi_k^{(1)} + \pi_k^{(2)} - 1, \rho_k \pi_k^{(1)}]$	1	0
	$[\rho_k \pi_k^{(1)}, \pi_k^{(1)}]$	1	1
	$[\pi_k^{(1)}, 1]$	0	1

Discussion

- We can allow that the sampling coordination parameter depend on t.
- How to determine the sample coordination parameter ρ_k?
  - If the temporal variability is high (as in urban area), we may reduce ρ_k.
  - If the response burden is high, we may increase ρ_k.
- We have assumed that the survey costs are the same. If the cost function is known, then we can compute the expected total cost by

$$C_T = \sum_{k=1}^{N} C_k \pi_k$$

where $\pi_k = nM_k / (\sum_{i=1}^{N} M_i)$ and M_i is the MOS for unit i. Thus, we can determine the sample size n such that C_T is less than the budget.

Imposing Balancing Condition in the Planned Missingness

Imposing balancing conditions in planned missingness

Planned Missingness

An example of planned missingness structure

Sample	Year 1	2	3	4	5	6
S1	✓		✓		✓	
S2		✓		✓		✓
S3	✓	✓	✓			
S4		✓	✓	✓		
S5			✓	✓	✓	
S6				✓	✓	✓

We have auxiliary variable information (X) available in the sampling frame.

Kim (ISU) Sampling designs 05/09/2022 20 / 41

Imposing balancing conditions in planned missingness

Balancing condition

- Balancing condition: The distribution of X in the sample is close to the one in the population level.
- In cross-sectional survey samples, the balancing condition in the sampling design can be achieved using the balanced sampling procedures such as the cube method (Deville and Tillé, 2004) or the rejective Poisson sampling (Fuller, 2009).
- The rejective Poisson sampling can be viewed as a constrained maximum entropy sampling.
- In the longitudinal survey with planned missingness, the balancing condition is more difficult to achieve.
- We may achieve the balancing condition through optimal weighting.

Kim (ISU) Sampling designs 05/09/2022 21 / 41

Maximum entropy sampling

- Result 1: The sampling design that maximizes entropy

$$I(p) = -\sum_{S \in \mathcal{S}} P(S)\log\{P(S)\}, \quad (1)$$

on $\mathcal{S} = \{S; S \subset U\}$ with fixed inclusion probabilities π_k, $k \in U$, is the Poisson sampling design.

- Poisson sampling is a maximum entropy sampling design, but it is of relatively limited interest because the sample size is random.
- Thus, we may want to impose additional constraint to the above optimization problem: The fixed-sample size constraint.

Maximum entropy design with fixed sample size

- Let $\mathcal{S}_n = \{S; S \subset U, |S| = n\}$ be the set of all possible samples of size n. The maximum entropy sampling can be defined as maximizing the entropy in (1) subject to

$$\sum_{S \in \mathcal{S}_n} I(k \in S)P(S) = \pi_k, \quad k \in U \quad (2)$$

and

$$\sum_{S \in \mathcal{S}_n} P(S) = 1. \quad (3)$$

- The solution to this optimization problem is

$$P(S) = \begin{cases} \dfrac{\exp\left(\sum_{k \in S} \theta_k\right)}{\sum_{S \in \mathcal{S}_n} \exp\left(\sum_{k \in S} \theta_k\right)} & \text{if } |S| = n \\ 0 & \text{otherwise,} \end{cases} \quad (4)$$

where $\theta_1, \cdots, \theta_N$ are determined by (2). Thus, (4) is the sampling design for the maximum entropy sampling (with fixed-sample size n).

Result 2 (Chen et al., 1994)

Any maximum entropy design of parameter θ_k can be written as a conditional design with respect to the fixed size of all Poisson designs of inclusion probabilities:

$$\tilde{\pi}_k = \frac{\exp(\theta_k + C)}{1 + \exp(\theta_k + C)} \tag{5}$$

for all $C \in \mathbb{R}$.

Note: In addition to the fixed sample size constraint, we can also impose the balancing constraints to the maximum entropy design. The theory has not been fully developed in the literature.

Optimal Weighting using Empirical likelihood (Qin et al., 2002)

- Let π_i be the first-order inclusion probability. The log-likelihood can be written as

$$\ell = \ell(p, W) = \sum_{i \in S} (\log \pi_i + \log p_i) + (N - n) \log(1 - W) \tag{6}$$

where $p_i \geq 0$,

$$\sum_{i \in S} p_i = 1, \tag{7}$$

$$\sum_{i \in S} p_i \pi_i = W. \tag{8}$$

- In addition, if $\bar{X}_N = N^{-1} \sum_{i=1}^{N} \mathbf{x}_i$ is available, we can impose

$$\sum_{i \in S} p_i \mathbf{x}_i = \bar{X}_N. \tag{9}$$

as an additional constraint for the optimization problem.

Imposing balancing conditions in planned missingness

Remark 1

- The objective function in (6) is motivated from the observed likelihood function in the missing data literature.
- Under assumption
$$(X_i, Y_i) \overset{iid}{\sim} F$$
and $\pi(x, y) = P(I = 1 \mid x, y)$, the observed likelihood can be written as
$$L(F) = \prod_{i=1}^{N} \{\pi(x_i, y_i) dF(x_i, y_i)\}^{I_i} \times (1 - W)^{N-n}$$
where
$$W = P(I = 1) = \iint \pi(x, y) dF(x, y)$$
is unconditional sampling rate and $n = \sum_{i=1}^{N} I_i$ is the sample size.
- The empirical likelihood is obtained by discretizing F.

Kim (ISU) Sampling designs 05/09/2022 26 / 41

Imposing balancing conditions in planned missingness

Remark 2

- Three constraints, (7), (8) and (9), are used in the empirical likelihood optimization.
- The first constraint is the normalization constraint. If we simply maximize
$$\ell(p) = \sum_{i \in S} \log(p_i)$$
subject to $\sum_{i \in S} p_i = 1$, the solution is $\hat{p}_i = 1/n$.
- The second constraint (8) is used to correct the selection bias. Constraint (8) can be called the internal bias calibration (IBC) condition (Firth and Bennett, 1998).
- The third constraint (9) is used to reduce the variance by incorporating the auxiliary information. Constraint (9) is the balancing condition or calibration condition.

Kim (ISU) Sampling designs 05/09/2022 27 / 41

Imposing balancing conditions in planned missingness

Maximum empirical likelihood estimator

- The final weight $\hat{p}_i$ can be obtained by maximizing (6) subject to the three constraints.
- For estimating $\theta = E(Y)$, the final maximum EL estimator $\hat{\theta}_{EL} = \sum_{i \in S} \hat{p}_i y_i$ is asymptotically equivalent to the following optimal regression estimator of Rao (1994):

$$\hat{\theta}_{opt} = \frac{1}{N} \sum_{i=1}^{N} \left\{ \mathbf{x}_i \beta^* + \frac{I_i}{\pi_i} (y_i - \mathbf{x}_i \beta^*) \right\} \qquad (10)$$

where

$$\beta^* = arg \min_{\beta} V \left\{ \sum_{i \in S} \frac{1}{\pi_i} (y_i - \mathbf{x}_i \beta^*) \right\}.$$

Kim (ISU) Sampling designs 05/09/2022 28 / 41

ADAPTIVE SAMPLE SIZE ALLOCATION FOR STRATIFIED SAMPLING

Adaptive sample size allocation

Basic setup

- Consider stratified sampling (with H strata).
- Sample size allocation: Given $n = \sum_{h=1}^{H} n_h$, how to choose n_h ?
  1. Proportional allocation: choose $n_h \propto N_h$.
  2. Optimal allocation: choose n_h such that

$$\text{minimize } Var\left(\hat{Y}_{HT}\right) \quad \text{subject to } c_0 + \sum_{h=1}^{H} c_h n_h \leq C,$$

  where c_h is the cost of observing an element in stratum h and C is a given total cost.
- We will first focus on proportional allocation.

Kim (ISU) Sampling designs 05/09/2022 30 / 41

Proportional allocation in practice

- Proportional allocation formula:

$$n_h = n \cdot \frac{N_h}{N}$$

  and n_h are integer-valued.
- Two problems
  - Some of n_h can be equal to zero if N_h is very small.
  - Simple rounding to nearest integer does not always work. It may lead to $\sum_{h=1}^{H} n_h \neq n$.
- To guarantee unbiased estimation, we need to impose $n_h \geq 1$ for all $h = 1, \cdots, H$.

Example: apportionment of the House of Representative

One of the fundamental reasons for conducting the decennial census of population is to reapportion the U.S. House of Representatives. Apportionment is the process of dividing the 435 memberships or seats in the U.S. House of Representatives among the 50 states.

- $H = 50$ states in the US
- N_h: Population size for state h, available from the latest Census
- n_h: number of seats in the US House of Representatives. US constitution requires the followings
  1. $n_h \geq 1$
  2. $\sum_{h=1}^{H} n_h = 435$
  3. $N_h/n_h \cong N/n$ as closely as possible for every h.
- Thus, we need to allocate $435 - 50 = 385$ seats proportional to the population size.

Huntington-Hill method

1. Calculate a set of "priority values" for each state, based on the state's apportionment population.
2. Sort those values from largest to smallest.
3. Allocate a seat to a state each time one of its priority values is included in the largest 385 values in the list.

The formula for calculating priority values is as follows:

$$a_{h,s} = \frac{N_h}{\sqrt{s(s+1)}}$$

where s is the number of seats that the state has been allocated so far.

Toy Example ($H = 4, n = 10$)

State	Pop Size	a_1	a_2	a_3	a_4	Seats
A	100,000	70,710	40,825	28,868	22,361	1+4=5
B	50,000	35,355	20,412	14,434	11,180	1+1=2
C	40,000	28,284	16,330	11,547	8,944	1+1=2
D	20,000	14,142	8,165	5,774	4,472	1

Disclaimer: This is an illustrative example. The official rule is more complicated.

Justification

- Using

$$\sum_{h=1}^{H} \frac{N_h^2}{n_h} = \sum_{h=1}^{N} N_h^2 - \sum_{h=1}^{H} \sum_{k=1}^{n_h-1} \left(\frac{1}{k} - \frac{1}{k+1} \right) N_h^2$$
$$= \sum_{h=1}^{N} N_h^2 - \sum_{h=1}^{H} \sum_{k=1}^{n_h-1} \frac{N_h^2}{(k+1)k}$$

we have only to maximize the second term:

$$\sum_{h=1}^{H} \sum_{k=1}^{n_h-1} \frac{N_h^2}{(k+1)k} = \sum_{h=1}^{H} \sum_{k=1}^{n_h-1} a_{h,k}^2.$$

- Because $a_{h,1}, a_{h,2}, \cdots, a_{N_h-1}$ are monotonely decreasing for each h, we have only to select the largest $n - H$ elements of $a_{h,k}$'s.

Optimal allocation: Problem formulation

- We may assume that

$$y_{hi} \stackrel{iid}{\sim} (\mu_h, \sigma_h^2), \tag{11}$$

for $h = 1, \cdots, H$.

- Given the sample size n and known population parameters, we wish to find the sample allocation that minimizes

$$V(\hat{Y}_{HT}) = \sum_{h=1}^{H} \frac{N_h^2 \sigma_h^2}{n_h}$$

subject to the constraint that $\sum_{h=1}^{H} n_h = n$.

- The solution to the above optimization problem is

$$n_h \propto N_h \sigma_h, \tag{12}$$

where σ_h^2 is the population variance of y-values in stratum h.

Research problem formulation

- How to achieve the optimal allocation in (12) when σ_h^2 are unknown ?
- Idea:
  - We can sequentially update the estimate σ_h^2 from the current sample.
  - May apply Huntington-Hill algorithm by trying to achieve $N_h\hat{\sigma}_h/n_h$ as equally as possible.
  - That is, use

$$a_{h,s} = \frac{N_h}{\sqrt{s(s+1)}}\hat{\sigma}_h$$

    as the priority values for applying the Huntington-Hill algorithm.
- The idea of using Huntington-Hill algorithm for sample size allocation in stratified sampling was first investigated by Wright (2012) but not in the context of adaptive sampling design.

CONCLUDING REMARKS

Concluding Remarks

- Three topics considered
  1. Sampling design for negatively coordinated longitudinal surveys
  2. Imposing balancing conditions in planned missingness
  3. Adaptive sample size allocation under stratified sampling
- The ideas are somewhat rough, but can be further developed with more context in developing future sampling designs in NCES.

REFERENCES

SOME TOPICS FOR COORDINATED SAMPLING OF SCHOOLS OVER TIME

Breidt, F.J. and W.A. Fuller (1999), 'Design of Supplemented Panel Surveys with Application to the National Resource Inventory', *Journal of Agricultural, Biological, and Environmental Statistics* **4**, 391–403.

Brewer, K.R.W., L.J. Early and S.F. Joyce (1972), 'Selecting Several Samples from a Single Population', *Australian Journal of Statistics* **3**, 231–239.

Chen, X.H., A.P. Dempster and J.S. Liu (1994), 'Weighted Finite Population Sampling to Maximize Entropy', *Biometrika* **81**, 457–469.

Deville, J.C. and Y. Tille´ (2004), 'Efficient Balanced Sampling: The Cube Method', *Biometrika* **91**, 893–912.

Firth, D. and K.E. Bennett (1998), 'Robust Models in Probability Sampling', *Journal of the Royal Statistical Society: Series B* **60**, 3–21.

Fuller, W.A. (2009), *Sampling Statistic*, Wiley, Hoboken, NJ.

Qin, J., D. Leung and J. Shao (2002), 'Estimation with Survey Data under Non-Ignorable Nonresponse or Informative Sampling', *Journal of the American Statistical Association* **97**, 193–200.

Rao, J.N.K. (1994), 'Estimating Totals and Distribution Functions using Auxiliary Information at the Estimation Stage', *Journal of Official Statistics* **10**, 153–165.

Wright, T. (2012), 'The Equivalence of Neyman Optimum Allocation for Sampling and Equal Proportions for Apportioning the U.S. House of Representatives', *The American Statistician* **66**, 217–224.

OBSERVATIONS FROM THE NASEM PANEL DISCUSSIONS AND FINDINGS

S. Lynne Stokes, PhD
Southern Methodist University

ABSTRACT

In April the National Academic of Science, Engineering and Medicine (NASEM) released a report commissioned by the Institute for Education Science (IES) to provide a vision for the National Center for Education Statistics (NCES) in the rapidly and dramatically changing landscape of education in the United States. The demand for more extensive and better data underlies the need of policymakers, administrators, and educators to make evidence-based decisions on all aspects of education. The growth in data sources is matched by the increased range of urgent issues to address: diversity, inclusion, exclusion, equity, accessibility, technology resources, education economics as well as the traditional evaluations of achievement, attainment, and both short- and long-term education outcomes.

The panel's discussion of this broad charge covered issues of data content and the data acquisition process as part of the larger picture of data needs for decision-making by the agency and its stakeholders and collaborators at the federal, state, and local level, as well as researchers' data needs and expectations. The panel noted in particular that NCES is overburdened and understaffed both across the board and specifically in technical statistical expertise. While NCES needs inhouse capability to provide data analysis to meet stakeholders' needs, they could also play a role by curating and providing tools so that state and district decision-makers can make use of the NCES data conveniently, and for purposes of identifying and comparing their performance with that of other jurisdictions that have similar characteristics to their own. The panel found that some barriers to such comparisons are differences in definitions, inclusion-exclusion criteria, and data adjustment procedures among NCES surveys. Others are the inability for local officials to combine local and NCES data and to access microdata (except their own) conveniently. Finally, some local jurisdictions lack the staff with the time and expertise to analyze the data themselves and could benefit from tools to facilitate comparisons between their own outcomes and others'.

The panel report also recommended that NCES should acquire or supplement current in house expertise to have capability of full technical oversight of contractors. This would also have the benefit of helping the agency retain knowledge about the data and lessons learned about the data collection process. The report made more detailed recommendations about internal NCES/IES structure and staffing, and division of responsibilities with contractors and potential federal agency collaborators. Finally, the report includes a two-year roadmap to strategic planning, goals and implementation milestones on the assumption that NCES must move into the future streamlining, making hard choices and doing more with less.

A number of the recommendations and conclusions of the NASEM panel resonate with findings and recommendations from previous NISS-NCES panels in *Connecting the Dots* series[2]. Three of these themes are: 1) partnership relationships with stakeholders, as data providers and as information users, 2) data integration across internal and external sources, with the ancillary requirement of standardized definition and modern language in addition to technology for efficient sharing, 3) innovation and continuing nimble adaptation to change in all aspects of data definition, acquisition, integration, analysis and publication.

Ideas presented at the *Forum* this morning that relate to these NASEM themes include the objective of designing samples that ensure the ability to create small area models to respond to stakeholders' need for information about "schools/districts like ours." Use of common definitions (at least throughout IES/NCES) of "core" items and common stratum definitions can enable data sharing to allow coordinated designs; where these definitions conform to wider use, integration with other data sources becomes possible. Another NASEM panel concern consistent with those raised in this *Forum* is identifying ways of reducing respondent burden. A separate NISS-NCES panel concurrent with this *Forum* is considering in detail the issues of defining burden and of identifying item/stratum definitions in sufficiently refined terms to characterize similarity of jurisdictions.

Widening data sources by accessing administrative records, online sources and other non-probability sampled information offers potential efficiency as statistical ancillary information to be used in design as well as modeling. Such Innovative design includes responsiveness to stakeholders needs by adjusting for perceived burden to participants or mitigating nonresponse and resulting sampling bias despite the high propensity to decline participation. Whichever strategies are developed for the coordination of designs for multiple surveys, it is clear that they should be expandable to a greater number of surveys, should allow incorporation of ancillary information from new data sources, and should be able to respond to rapidly changing information demands. The words "Innovation", "nimble", and "flexible", that recur throughout the NASEM report also apply here in the context of sampling design requirements.

2 *Post-Covid Surveys,* December 2020; *Setting Priorities for Federal Data Access to Expand the Context for Education Data*, March 2021, *Connecting the Dots, I & II, Integrated Sampling Approach for Multiple Surveys*, August-November 2021 https://www.niss.org/nces-report-library.

PART II

DISCUSSION

DISCUSSANTS' COMMENTS ON COORDINATING NCES SURVEYS

MARY THOMPSON, Distinguished Professor Emerita, University of Waterloo

We have been asked to comment on possible techniques to coordinate the sampling for multiple surveys. The chances of successful coordination of sampling will improve with increased central management of the surveys in other ways, and therefore I will start with a few words about that.

There are some very positive steps to be taken:

- While many of the surveys are carried out by external contractors, there is a possibility of more coordinated management and of bringing the data in house.
- There is a willingness to try to make questioning more consistent across surveys to reduce the numbers of questions being asked of schools participating in multiple surveys.
- There is an intention to investigate extending the use of administrative data to complement survey data.
- There is an intention to try to plan the deployment of survey cycles at least one year ahead.
- There is an intention to have a single contractor lining up participation within each geographic area for which the samples are being coordinated.

The Sampling Frame

There is an excellent frame (Common Core Data or CCD) for the public school system, and the intention is to devote even more resources to keeping it up to date and adding information to it. This is very important because the additional information can be used as auxiliary information to improve estimation and even to help reduce bias due to nonresponse.

Presumably the structure of this frame is two-level in the sense of being a listing of districts and listing of schools within districts. The schools are of various academic levels and types. Each survey will have its own sub-frame, such as schools with a fourth grade.

There are approximately 13,800 public school districts in the United States; on average, they contain 98,000/13,800 = 7.1 schools and 55,200,000/13,800 = 4,000 students. The smallest has one school, and the largest (New York City) has 1800 schools and 1.1 million students.

The NAEP is conducted nation-wide every two years in grades 4, 8 and 12, and is mandatory. The data from it can be leveraged. There are about 8,300 schools in the sample.

The proposed added restricted access data of response history and current "burden" for districts and schools should assist with coordination of sampling.

Priority

We heard that in planning deployment there should be no priority order for NCES surveys. Presumably it means that any survey that is approved for a given year must happen with a design as close as possible to what was approved. Perhaps also, in the joint sampling process, each must in theory have access to the whole relevant subframe.

At the same time, the NAEP and the international surveys are to be considered part of the calendar.

Variety of Designs

The Michigan presentation gave some examples illustrating the variety of NCES surveys:

- One-time cross-sectional surveys
- Repeat cross-sectional surveys at certain time intervals
- Longitudinal surveys

The unit of interest varies among surveys. The unit of interest can be

- The school itself
- The student in a certain grade or grades
- The staff member or administrator

One question to consider is whether repeat cross-sectional surveys should have intentional overlap in their samples so as to be able to measure change more efficiently.

Stratification

The word stratification was used to indicate categories for oversampling as well as geographic divisions.

The stratification differs from survey to survey, according to the survey aims. It may be possible to bring some of the stratification schemes closer together. Primary strata are location-region combined with school type.

Stages of Sampling

The sample may need to represent the relevant school types of various regions, and most simply, a stratified random sample of relevant schools would be taken within each census region or state. However, the ECLS-K survey is one that has PSUs, namely counties. Among the NCES surveys there may be other cases of sampling in stages.

Coordinated Sampling Approaches

A question is the level of geography/governance at which coordination should take place. An approach that works well in some regions or states may not work for all. At the same time, it is worth considering certain technical approaches that might work in some of them.

Burden

We were asked to think about the burden on a school district as well as the burden on a school. In both cases, arranging for surveys takes up staff time and committee time. However, for the school, surveying students also takes up instruction time, and this puts harder limits on the numbers of surveys that are possible.

I would imagine that the nature of the burden on a school district varies substantially with its size. A very large school district will have a large staff and perhaps some staff dedicated to managing requests for surveys. For some, a survey that misses the school district in question could not be regarded as nationally representative, and thus there will be many to be coordinated. A very small school district may hardly ever be asked to participate in an NCES survey. Perhaps the burden is most difficult for one of average size, with a relatively small staff and less frequent requests from NCES.

The rest of this document imagines that the burden for a mid-size district is minimized if it has close to constant numbers of NCES survey requests and administrations per year (mean could be less than one), and if the burden is shared evenly among the mid-size districts in the stratum that are willing.

Comments on the Proposals

Suppose for illustration that we want to make the number of surveys over a two-year period (last year and this year) as even as possible. Suppose that for school k we have a two-year target NCES survey burden of B_k. Let:

- ml_k be the number of NCES surveys administered last year.
- mc_k be the number of NCES surveys already committed to for the current year.
- L_k be the maximum number of NCES surveys possible for this year. Then the target B_k will be between ml_k + mc_k and ml_k + L_k.
- If B_k is the target burden, then this year's number of "slots" Sl_k is B_k – ml_k – mc_k.

We want also to control the burden on the districts, in the sense of trying to spread the burden within a primary stratum evenly over districts (or suitable unions of districts) belonging to the geographic area of the stratum. Let h denote a primary stratum, and let f denote a union of mid-size not necessarily contiguous school districts within h of a suitable number, e.g., containing 50 to 100 schools. We would want to spread the burden evenly over the values of h and f (hf).

In brief:

- The Michigan presentation suggests that for the current year we might list schools and their slots within each hf, and take successive or simultaneous samples from these for the candidate surveys. The four options (Independent selection, substitution; Independent selection, rejection; Sequential sampling, with random survey order; Matrix sampling) all seem worth considering and adapting to the NCES context. The last one seems easiest to gear to a concept of "target burden". It is likely however that B_k is not uniform within an hf in the two-year burden context.
- As the authors point out, with all of the strategies, some kind of planned substitution is likely to be necessary. Computing inclusion probabilities for schools for individual surveys would have to be supplemented with a way of addressing nonresponse bias through information on the frame, administrative data, and perhaps NAEP information.
- Dr. Kim's presentation introduces the idea of permanent random numbers, that is sometimes used in repeated sampling designs to control burden across time by effectively dividing the population into randomly selected subpopulations and rotating the selection of new units among these. This concept seems worth considering here.
- As well, among other useful concepts and techniques, Dr. Kim discusses the concept of entropy of a sampling design.

A Random Number Method

Here is an outline of a strategy, drawing to some extent on the permanent random number idea.

- For each hf, let the number of school-slots be S_hf. Make a list of those school-slots, listing them in random order. Assign to each school-slot a random number from the interval [0, 1].

- Make sure these "random" numbers are spread out within [0,1] by rejecting sets that have too much clumping, i.e., reduce the entropy of each sequence within an hf. Or use stratified random sampling where the strata are small intervals of [0,1].
- For a given survey that needs a certain number of schools, for each hf, take a small subinterval of [0, 1] and select the schools whose "random" numbers fall in the subinterval. (If sampling at the same rate for all hf, this could be the same subinterval for all hf.)
- For a second survey that could or should sometimes be paired with the first one, a subinterval that intersects the first one would be used.
- For a second survey that should not be paired with the first one, a disjoint subinterval would be used.
- With this method, there is a spreading out of the sample within h, using the non-overlapping divisions f, and there is a rotation of schools used as the intervals of "random" numbers take their various possible values.

The exercise of finding exact intended inclusion probabilities for the school-slots (and hence the schools for an individual survey) could require carrying out these steps multiple times in the same manner as in the Monte Carlo method for the options of the Michigan presentation. Again, for final inference nonresponse and the necessity for substitution would have to be accounted for.

JERRY REITER, Professor & Chair, Department of Statistical Science, Duke University

I focus my comments on the four strategies presented by the University of Michigan team. I also describe a hybrid strategy that I believe is worthy of study. First, I describe some general questions to consider for each strategy.

- When do samples need to be selected: at the beginning of the year, or any time before data collection is planned?
    - Allowing samples to be taken at any time has advantages, in that NCES can be responsive to changing conditions, e.g., disruptions from a pandemic or disaster, changes in available funding, and new Congressional mandates for certain surveys.
- Who selects the samples: NCES (or its agent) or the individual contractors?
    - Allowing individual contractors to select samples allows them to specify study designs for their particular survey.
- How simple is the methodology to explain and implement?
    - Ideally, the methodology requires minimal tracking by NCES or the contractors.
- Is the methodology equitable to all surveys?
    - Surveys at different points on the timeline should not have advantage or disadvantage in terms of their ability to select who they want to.
- How effectively does the methodology minimize participant burden?

In my comments, I assume that the sampling unit is a school. Similar considerations apply for other sampling units.

Strategy 1: Independent Selection, Substitution

Pros: This strategy is relatively simple to explain. It requires essentially no theoretical development, as the survey weights can be obtained by simulation. It can be implemented sequentially, that is, samples for any particular survey can be selected at any time as opposed to all at once. It enables contractors to specify survey designs as they best see fit for their surveys. *Cons*: NCES and contractors need to know how many times each school has been contacted, so that they can follow the strategy. This creates extra overhead and sampling complications, and presumably NCES would need to check that the contractors are following the strategy. Surveys at the end of the timeline may not have enough units belonging to small subpopulations left to sample (unless burden is disregarded), so that their contractors may be disadvantaged.

Strategy 2: Independent Selection, Rejection

Pros: This strategy is relatively simple to explain. It requires essentially no theoretical development, as the survey weights can be obtained by simulation. The survey can control participant burden effectively. *Cons*: The sample is taken all at once, making it difficult to adapt to changing conditions. NCES would need to take the sample and likely from a simple design. This is problematic for contractors who have tuned survey designs to their specific populations and objectives; it would be beneficial for these contractors to have control over their designs. It would be challenging to implement this strategy with different stratification schemes per survey. The space of feasible samples may not be large, which could make it challenging to reduce variances from efficient survey design.

Strategy 3: Sequential Sampling

Pros: This strategy can be implemented sequentially, that is, samples for any particular survey can be selected at any time as opposed to all at once. It enables contractors to specify survey designs as they best see fit for their surveys. *Cons*: The survey design is difficult to explain, and the computation of weights is quite complicated—perhaps too complicated in practice with a large number of surveys. NCES and contractors need to know how many times each school has been contacted, so that they can follow the strategy. This creates extra overhead and sampling complications, and presumably NCES would need to check that the contractors are following the strategy. Surveys at the end of the timeline may not have enough units belonging to small subpopulations left to sample (unless burden is disregarded), so that their contractors may be disadvantaged. This strategy does not control burden by design. It relies on randomness to keep burden low; this may fail in practice.

Strategy 4: Matrix Sampling

Pros: This strategy is relatively simple to explain. It requires essentially no theoretical development, as the survey weights can be obtained by probability computations (or simulations if necessary). This strategy controls burden. The strategy is equitable to all surveys in the timeline. *Cons*: The sample is taken all at once, making it difficult to adapt to changing conditions. NCES would need to take the sample and likely from a simple design. This is problematic for contractors who have tuned survey designs to their specific populations and objectives; it would be beneficial for these contractors to have control over their designs. It would be challenging to implement this strategy with different stratification schemes per survey.

Alternative Strategy

An alternative strategy uses the concept of matrix sampling but applies it to create frames rather than samples. That is, NCES randomly puts each school into at most k sampling frames, where k is determined according to burden considerations and sufficient frame size. Each survey gets a frame. Contractors can sample from the frame using whatever design they please. *Pros*: Contractors can make the survey design as they see fit; samples can be selected at any time up to data collection; the method is easy to explain and facilitates simple weight adjustments; the method controls burden; the method is fair to surveys at all points in the timeline. *Cons*: With many surveys, the extra sampling step to create frames could increase standard errors, particularly if burden (k) is kept low.

Simulation Studies

The designs can be evaluated using simulation studies. NCES could use the CCD as a population. They could implement each strategy, attempting to recreate the current designs of the studies (with stratification) as closely as possible. Using a variety of variable types (e.g., binary with common event, binary with rare event, continuous), NCES could repeatedly draw samples and compute estimates of means and sub-domain means. Key questions include the following.

- Mean squared errors: Is each survey able to estimate population quantities with roughly similar relative accuracy as they are able to now? Or, do some surveys get disfavored?

- Convenience: How closely is each survey able to approximate the design that they did previously, as measured by the similarity of design effects?
- Burden: How closely is each survey able to keep burden at the desired level, as measured by distributions of the number of samples each unit is in?

Some factors are not necessarily informed by the simulation results. But, designing the simulations might help NCES think about these issues.

- Simplicity: How easily can each survey adapt to disruptions, e.g., from a pandemic, or to changes in timelines, or to changes in anticipated funding (sample size)?
- Estimation simplicity: How easy is it to estimate means and variances? Can transparent adjustments be made to handle nonresponse, either through weighting or imputation?

S. LYNNE STOKES, Professor and past Chair, Department of Statistics, Southern Methodist University

The presentations of initial theoretical work on coordinating sampling designs raise three issues relevant to the technical aspects of sampling design. Outside the immediate focus of this panel, but nonetheless presenting implicit constraints on sampling design are the concerns expressed by the NASEM panel regarding the practicalities of implementation and about successful engagement of the broad range of stakeholders.

Optimizing Data Definitions

For key data where optimization for data sharing requires some standardization, the case for standardization across surveys is strong.

The potential to reduce burden by seeking out the best source, often state or district rather than school, as respondents for specific items, may facilitate implementation of consistent item questions and response definitions. Another benefit of using a common source across surveys and not re-asking questions is the potential to reduce burden at all levels. In addition, if the best source of key data items is a district or state source, then at least those items may be available for schools/districts in the reserve sample eliminating the burden for schools that are not ultimately sampled or for schools that otherwise decline participation.

At the same time, one concern is the need to be sensitive to information that changes over time, and the inconsistency that can introduce. Even ordinarily stable items can shift under the kinds of dramatic change that covid precautions induced. Consequences for inference are obvious.

For stratum variables and boundaries, commonality of definitions affects sampling designs as well as the potential for data sharing, and analogous arguments for standardization (if it can be implemented) apply. Even apparently obvious variables may currently have varying definitions. For example, the teachers and principals survey defines “school” differently from other surveys; and there is increasing need to incorporate information on non-traditional non-public schools that may not fit current school descriptors.

Probability Sampling and Multiple Selection

It is not clear how calculation of probabilities of selection across surveys will be done with each of the proposed approaches. Starting from the simplest case of school-based cross-sectional surveys, both theoretical development and simulation will be necessary to evaluate each proposed approach with respect to school burden/multiple selection as well as selection bias, precision of estimators, etc. Adding longitudinal studies (from the early childhood study to the high school and beyond) to the mix is a first refinement. Do these simply add burden without altering selection probabilities in subsequent years?

Rules will need to be developed for ensuring that each survey’s coverage is when inclusion of school with unique or rare demographics has to be limited.

Specific question about probabilities of selection arise if the coordinated surveys used different sampling algorithms. For example, some surveys have traditionally stratified the population then ranked schools within stratum by some useful covariate in order to sample systematically – a design to increase precision. Could subsequent surveys continue with systematic sampling? How would probabilities of selection across surveys be calculated?

Integration with Other Surveys

Coordinating all NCES surveys is just one part of controlling overuse of certain schools and of managing burden for districts, schools and individual participants. Within the Department of Education there are other surveys, notably NAEP but also other studies and surveys conducted within IES. What are the consequences for NCES coordinated sampling designs of previous selection (or over-selection) of schools by those surveys?

NAEP frequently selects the same schools within the Trial Urban District Assessment (TUDA) project to obtain a more precise picture of inner-city schools. Would the pre-existing cumulative burden for these schools disallow their inclusion in other NCES surveys?

By adjusting for NAEP years, how would samples in NAEP years be commensurate with samples in non-NAEP years? It seems that this problem should be solvable, but it cannot be ignored.

On the positive side, how could NAEP variables help? Could NAEP information be useful for refining nonresponse adjustment for other studies? Could NAEP information be used in conjunction with aggregate data on nonresponse for other NCES studies to understand the impact and the relative influence of non-responding schools or districts?

Responses to NASEM Panel Observations

The NASEM report[3] recommended that contractors should work collaboratively with NCES/IES staff to build capacity and to incentivize cost-effectiveness.

- This coordination of sampling designs across multiple surveys is exactly in line with this recommendation.

The NASEM report recommended that NCES expand its role in data curation and governance, particularly as it relates to state and local agencies. They are extremely interested in local area data or ways to compare their own performance with others like themselves. The data products currently offered are often not seen as useful for this purpose.

- A product of this coordinated survey design fits nicely with this priority. It could jumpstart the process of standardizing the way data are collected, not just for their own (NCES/IES) surveys, but for state surveys as well. A trusted liaison could begin this dialog with state agencies to arrive at a system that satisfies multiple stakeholders.

The NASEM report recommended that “When designing data collection operations, NCES is encouraged to seek out members of diverse communities to help survey operations experts understand how to approach various populations and provide benefits to survey respondents.” The approach of reciprocity should improve survey response rates as well. Engaging members of the public at this stage of design requires expertise in qualitative methods. NCES can collaborate with other centers in IES or can contract out this work, but permanent NCES staff need to be closely involved to assure that knowledge is retained and to develop a culture of DEIA awareness that can be applied to other areas of NCES’s work.

[3] National Academies of Sciences, Engineering, and Medicine. 2022. *A Vision and Roadmap for Education Statistics*. Washington, DC: The National Academies Press. https://doi.org/10.17226/26392.

- An advantage of the proposal to coordinate studies and work closely with districts and schools and to design deliberately to provide information with local relevance brings the opportunity of engaging diverse communities.

The NASEM report expressed concern about the slowness with which data are produced: i.e., by the time the data come out, they may be out of date.

- Changes by coordinating multiple surveys would be primarily in the design and survey scheduling phases. Possibly, nonresponse adjustments could be accelerated, especially if there was a sample vetted in advance or even as an ongoing process beginning with the first survey in the field.

JAY BREIDT, Senior Fellow, NORC at the University of Chicago, and Professor Emeritus, Department of Statistics, Colorado State University

Coordinating NCES surveys has excellent potential to reduce the overall burden (by consolidating recruiting efforts and paying overhead costs once) as well as distributing burden more equitably. Without coordination, equity does occur (because these are probability sampling designs), but only at random and over the long term. Distribution of burden that is not only equitable but also perceived as equitable could help in recruiting of schools and districts, and a well-coordinated system could clearly demonstrate its equitable distribution.

Survey burden has a useful analogy to taxation (Sunter, 1977). NCES should weigh response load versus response obligation across the levels of districts, schools, and individuals, with response load and burden defined both within and across those levels. The response load is the cost of responding to a survey, including all aspects from the administrative overhead in the approval process down to individuals completing questionnaires. Mandatory data collections like CCD and NAEP are part of the burden. Burden does not scale with size: smaller units will often have disproportionately higher burden.

The response obligation is a determination of the reasonable share of the total burden for an entity (district/school/individual) and can be a function both of size of the entity and the value of the information services returned to that entity by NCES. NCES should be continuously asking the question if burden outweighs local benefits. The coordination system should ensure that the actual burden on a district or school rarely exceeds its fair share. This might be enforced as a hard constraint, or the constraint might be relaxed slightly.

This coordination problem has a long history in surveys, especially establishment surveys, but similar problems arise due to logistical constraints in almost any survey with a field component. As such, there is an extensive body of literature on which to build, dating back at least to Goodman and Kish (1950). Some aspects of experimental design have been explored in this context. For example, see Rao and Nigam (1990) for a critical review of methods that rely on incomplete block designs along with a proposed alternative that uses linear programming.

The Monte Carlo approach to construction of feasible samples and estimation of inclusion probabilities is a useful idea that has appeared in many contexts (Fattorini, 2006). In a NOAA Fisheries application (Papacostas and Foster, 2020), a sample of site-days (fishing access sites crossed with days in the fishing season) is stratified by region, season, type of day (weekday vs. weekend/holiday), and type of fishing (private vs. charter, shore vs. boat). Travel and scheduling constraints on the small crew of field samplers lead to many infeasible samples: for example, too many selected site-days for the available crew or too much geographic dispersion to allow for travel. Instead, NOAA draws many samples, checks the logistical constraints, and throws out the infeasible samples. The final sample is selected from among the feasible samples, and the complete set of feasible samples is used to estimate the inclusion probabilities.

A potential problem with this approach is both very small probabilities and very small estimated probabilities, leading to weight variation that is reflective of logistical constraints but not related to population features. That is, the logistical constraints plus the Monte Carlo error create uninteresting variation in the weights. Design-based estimators with correct inclusion probabilities are valid without any modeling assumptions but may be inefficient when design weights are weakly related to study variables and have wide variation.

One recommendation is to gain computational efficiency and improved inclusion probability estimation by starting with as much design and balance as possible algorithmically (using some combination of detailed stratification, balanced sampling or rejective sampling, etc.), and then select samples that meet logistical constraints only within this well-balanced class of samples. This approach is in contrast to selecting a large set of samples via some simpler design, without detailed stratification and balance, and filtering the larger set of samples.

A second recommendation is to consider weight smoothing (Beaumount, 2008), which fits well within the design-based paradigm. The idea of weight smoothing is to model one variable, the weight, so the procedure works for any survey response variable. Factors that are explanatory of important weight variation (like stratification variables or size variables) could be included in the model. The final weights would smooth out logistical and Monte Carlo variation, leading to estimators with small design bias and reduced variance relative to estimators with original, unsmoothed design weights.

An issue of practical implementation is how survey organizations that might conduct the coordinated samples would play a role in future NCES surveys. Advantages to those organizations would include some simplification of proposal development, as some of the sampling and recruiting burden would be shifted away from them, and a decrease in the need to deal with the coordination problem in the field as their sampled districts and schools refuse cooperation due to burden from other surveys. Potential disadvantages to the survey organizations would be the strong constraints in developing innovative sampling solutions, making it harder for them to achieve efficiencies and demonstrate their competitive advantages; NCES would be directly impacted by these disadvantages. Accordingly, it would be ideal if the coordinated sampling scheme maintained as much flexibility for sampling innovation as possible (e.g., constructing frames within which survey organizations could make selections).

References

- Beaumont, Jean-François. "A new approach to weighting and inference in sample surveys." Biometrika 95.3 (2008): 539-553.
- Fattorini, Lorenzo. "Applying the Horvitz-Thompson criterion in complex designs: a computer-intensive perspective for estimating inclusion probabilities." Biometrika 93.2 (2006): 269-278.
- Goodman, Roe, and Leslie Kish. "Controlled selection—a technique in probability sampling." Journal of the American Statistical Association 45, no. 251 (1950): 350-372.
- Papacostas, Katherine J., and John Foster. "A replication approach to controlled selection for catch sampling intercept surveys." Fisheries Research 229 (2020): 105609.
- Rao, J. N. K., and A. K. Nigam. "Optimal controlled sampling designs." Biometrika 77, no. 4 (1990): 807-814.
- Sunter, A. B. "Response burden, sample rotation, and classification renewal in economic surveys." International Statistical Review/Revue Internationale de Statistique (1977): 209-222.

FINDINGS AND RECOMMENDATIONS

This working session engaged NCES staff, presenters and discussants in open discussion of criteria and specifications for samples, for coverage of subpopulations, for schools with unique or rare attributes, and for estimators.

There were several common threads across all the topics discussed. First is a focus on relationships and the communication opportunities to define and to meet needs of the various aspects of the entire process from survey purpose through data analysis and publication. Expansion, regularization and integration of databases comprise the second thread that establishes the limitations on modeling and data analysis and sets the constraints on data sharing. Third is the pervasive need to think creatively in this dynamic context, anticipating continuing change whether driven by new technologies, new data sources, new opportunities, new demands for information, or new topical content. Practicality is the fourth theme; for sampling design this means preparing with research, studying via simulation, testing with pilot or other cases, and vetting with contractors and administrators/educators to ensure a confident launch of this new approach.

All the proposed approaches to coordinated design of multiple surveys depend on two conditions. The first is a common definition of descriptors, whether these variables are used to partition the population into "equivalent" subpopulations or to define a stratification structure for the various surveys to be coordinated. Without commonality of information to allow compiling across surveys data sharing is moot.

The second component is the use of additional or ancillary information either to provide preliminary information to improve efficient sampling or to provide for adjustments to sample selection probabilities (*e.g.*, based on propensity for nonresponse) or for detailed local description adequate for modeling "schools like ours."

These specifically need *new* and consistent information on additional attributes that can characterize schools and districts adequately to determine the extent of similarity. Acquiring this information will rely on the relationships with potential participants, with stakeholders and on a commitment within NCES to build the needed databases and to align NCES data collections with the goal of a common set of "key variables." Critical elements in this new information are required to illuminate the actual participation decisions of districts and schools and the rationales for those decisions.

Effective modeling of potential design variables like response propensity as well as the adaption of analysis tools for use by individual schools, requires much more specific information about additional relevant school attributes. A variety of sources is likely to be useful. Efficient design strategies can make use of non-probability prior information (e.g., web-scraping, administrative records, schools' own websites) to focus the actual data collection to meet precision requirements for specified estimators.

Another fundamental change will open more innovations for sampling: redefine criteria for survey acceptability by replacing the response rate threshold with stated precision requirements for estimation and bounds on bias. Once upon a time, response rate was a defensible surrogate for survey credibility. Now, as nonresponse continues to increase, it is more important to understand *who* declines participation, *why* they won't participate, and *what* the impact is on precision and on bias assessment and adjustment. Continuing to focus on nonresponse rate by adding easier to recruit schools (already abundantly represented in the respondent group) is inefficient for improving estimation or for reducing nonresponse bias. At best, this apparently easier resolution moves the problem of engaging minority, rare or highly resistant populations from an estimation problem to a bias problem.

Without question, future opportunities will increase and sampling design paradigms will evolve. These may come with strides in technology and data access. They may also come in deeper coordination among surveys than just the sampling design: for example, integration of distinct surveys for at least a subset of schools, addition of selected items for simultaneous incorporation into multiple surveys, or in the use of rapid response brief (even 5 question) surveys to plan or to calibrate larger data collections.

What will be most important will be an openness to further innovation and the opportunities to test multiple ideas, whether as separate pilot or research projects, or as brief interjections into ongoing surveys. The value of coordination will increase as additional surveys are incorporated, so even initial planning will require thinking ahead about expansion. Undoubtedly coping with coordination on a larger scale will present new challenges – openness to innovation will be essential. But always continuing evaluation in a formal, documented process will be critical and should involve the full cast of participants, contractors, NCES staff, what is successful and important to conserve.

Technical points raised by the invited discussants also generated further discussion with clarification about implications for NCES in regard to precision and bias requirements and also with regard to practicality for implementation. Several salient points pertaining to practical considerations and/or technical details not already mentioned follow.

With regard to data definitions and the need for commonality, one practical suggestion is to identify and go to the “best source.” For school attributes, this is often the state, sometimes the district. In addition to the efficiency, a potential side benefit would be obtaining the basic school information prior to approaching the school and without regard to the participation of the school personnel in a survey.

A first step in considering how to enrich the information available for schools in the CCD, for example, is to look at other data collections within the Department of Education, especially *EdFacts*. Another resource for *new* information, although currently this may be primarily anecdotal, are the recruiters from the contractors who conduct the surveys. Their accounts and insights could provide an initial basis for recording information on rationales for response/nonresponse and schools’ and districts’ perceptions of burden and also of potential benefit.

The balance of benefit to burden is regularly raised as the key determinant of decisions to participate or not. Benefit often seems to be construed as return of locally relevant information to the school, i.e., information about “schools like ours.” This information may be about performance, about universal problems faced by schools (e.g., post-covid catch-up), economics or other administrative issues. Dialogue with school and district administrators is the best resource for understanding benefit as well as burden. Modeling to provide this level of information also requires having sufficient detailed information to adequately define “like ours.”

Obviously, gathering these data requires thought at the outset as part of designing the samples. Like most design decisions, the benefits from forethought are even greater in planning for multiple surveys with the potential for incorporating specific school attribute data can be universal or can be deliberately parceled out among contemporaneous surveys.

Successful definition of “like ours” must encompass school or district perception and administrative criteria for similarity. At the same time, equivalence classes can be defined based on past survey outcome and demographic data can be defined (and modeled) statistically. Merged, these can provide the basis for effective modeling and individualized summaries and projections.

A technical note to support the use of common definitions relates to continual updating of school information as individual surveys acquire corrected or updated information. For data to be sharable across surveys, the same data file (same time stamp) needs to be used in calculating probabilities at sample selection and again a common updated data file (single time stamp for post-survey corrections; and these specific files need to be preserved as a synchronized database for future survey designs and for research. Some simplification may come from recruiting at a single point in time and seeking a common source (like state records) for basic information for all the surveys included.

Further design possibilities to think about revolve around finding new ways to use the reserve sample, and new alternatives to individual pilot studies. It is easier at this point to pose questions to prompt thought than to suggest which avenues might be worth pursuing. For example, if the reserve sample is a composite for multiple surveys, could it be constructed differently by targeting likely needs for more precise information by using school matching for non-responders? Another example: Could the role of pilot studies be changed or reduced? For example, could earlier surveys incorporate key items/ information or could different surveys (e.g., pulse surveys) with or without non-probability components (e.g., crowd-sourced or web-scraped) take over at least part of the role of a traditional pilot survey?

The unanimous finding with respect to the technical problems in coordinating multiple surveys is that multiple approaches need to be pursued. As these are developed, semi-hypothetical contexts and sophisticated simulations need to be developed simultaneously to challenge the innovative designs and to illuminate both their strong and their weak points. At this point in time, this design challenge is uncharted territory for IES/NCES. Consequently neither the comparative advantages of various design approaches nor their limitation cannot as yet be known. Constraints intrinsic to this challenge that will affect all approaches currently envisioned are likewise have yet to be identified. It will be important that innovation continue to be facilitated as the numbers and kind of multiple surveys to be incorporated expands.

Comparison of approaches discussed during this technical working session (block design basis, matrix sampling, modified Swiss design with balance, Bayesian design incorporating auxiliary data) remains conjectural without an appropriate test bed using real/realistic data from NCES and IES study designs in current or recent implementation. Thus, development and evaluation of potential sampling strategies will require a sophisticated (semi-synthetic) simulated database to adequately encompass the detail and vagaries encountered in actual NCES study implementation.

Such a simulation plan would include a prototype for the sampling process and construction of a semi-synthetic data set, elaborate enough to mimic the problems of current concern. Such problems include the likelihood of multiple reselections for particular units, uncertain or excessive selection bias, and unacceptable influence of a few selected schools (whether participants or non-responders).

The session culminated in in five recommendations for immediate next steps to be undertaken by NCES and by research statisticians.

Recommendations

- Pursue multiple promising strategies simultaneously, developing each selected methodology sufficiently to permit validation and feasibility testing of implementation and to allow (preliminary) estimation of precision and bounds on bias. Strategies deserving of consideration for development include a strategy derived from block design of experiment from a common sampling frame, also Bayesian design incorporating prior or non-probability information, a modified Swiss method with balance, and matrix sampling that creates separate sampling frames.
- Create both simulation and realistic test files and resources for evaluation of strengths, vulnerabilities and capacity for scaling up with extension to additional IES/Department of Education studies and surveys.
- Initiate consolidation of stratum definitions and items to enable data sharing for sampling design, reduce redundant data requests, and facilitate efficient, balanced sampling designs.
- Expand the detail available for CCD schools; for sampled schools collect data on participation history (e.g., agreement/refusal with rationale at each level of decision-making) as a basis for designing samples for future studies, and for research purposes.
- Remain open to continuing innovation of sampling strategy, especially to take advantage of alternative data sources (e.g., non-probability online data sources), to adjust selection probability (e.g., burden metrics) and to use model-based prediction (e.g., propensity for nonresponse, magnitude of potential contribution to bias).

APPENDICES

APPENDIX A: Biosketches for Expert Panel

Jae-kwang Kim, PhD

Title: Professor, Department of Statistics, Iowa State University

Jae-kwang Kim is Professor in the Statistics Department at Iowa State University (ISU)where he also previously he served Director of the Center for Survey Statistics and Methodology. He also held a dual appointment with KAIST in Korea. In 2015 Professor Kim receive the Gertrude M. Cox Award and he is an elected Fellow of both the American Statistical Association and the Institute for Mathematical Statistics. His doctoral research under Professor Wayne A. Fuller paved the way for a distinguished research career in survey sampling theory and applications with many refereed publications in leading statistics journals. His primary research interest continues to lie in problems survey sampling and statistical analysis with missing data. By introducing missing data (or latent variable) framework, a broad range of applied statistical problems can be reformulated as problems in missing data that then can lead to solution. Two classes of examples are measurement error models and multi-level models that can be treated as special cases of the latent variable modeling. Now, with several of his doctoral students/degree recipients, his work is extending into data integration, an emerging application area of missing data analysis, and into exploration of relevant topics in machine learning.

Yajuan Si, PhD

Title: Research Assistant Professor, Survey Research Center, University of Michigan

Yajuan Si is a Research Assistant Professor in the Survey Research Center at the Institute for Social Research on the University of Michigan-Ann Arbor campus. She received her PhD on Statistical Science in 2012 from Duke University. Dr Si's research lies in cutting-edge methodology development in streams of Bayesian statistics, linking design- and model-based approaches for survey inference, missing data analysis, confidentiality protection involving the creation and analysis of synthetic datasets, and causal inference with observational data. She has established a research agenda on advancing survey inference with Bayesian modeling techniques and adjusting for selection/nonresponse bias in complex data modeling with various types of data (e.g., survey and big data), with a focus on multilevel regression and poststratification and multiple imputation. Yajuan has extensive collaboration experiences with interdisciplinary researchers to improve the application of statistics in many different substantive fields, and she has been providing statistical support to solve sampling design and analysis issues on health and social science surveys.

Mike R. Elliott, PhD

Title: Professor, Department of Biostatistics, University of Michigan

Michael Elliott is Professor of Biostatistics at the University of Michigan School of Public Health and Research Scientist at the Institute for Social Research. He received his Ph.D. in biostatistics in 1999 from the University of Michigan. Prior to joining the University of Michigan in 2005, he held an appointment as an Assistant Professor at the Department of Biostatistics and Epidemiology at the University of Pennsylvania School of Medicine, and prior to that as a Visiting Professor of Biostatistics at the University of Michigan School of Public Health and as a Visiting Research Scientist at the University of Michigan Transportation Research Institute.

Dr. Elliott's statistical research interests focus around the broad topic of "missing data," including the design and analysis of sample surveys, casual and counterfactual inference, and latent variable models. He has worked closely with collaborators in injury research, pediatrics, women's health, and the social determinants of physical and mental health. Dr. Elliott serves as an Associate Editor for the Journal of the American Statistical Association.

Roderick Joseph Little, PhD

Title: Richard D. Remington Distinguished University Professor of Biostatistics; Professor, Department of Statistics; Research Professor, Institute for Social Research, University of Michigan

Dr. Rod Little’s research interests include incomplete data, sample surveys, Bayesian statistics, applied and statistics. A primary research interest is the analysis of data sets with missing values; another interest is the analysis of data collected by complex sampling designs involving stratification and clustering of units. Dr. Little’s inferential philosophy is model-based and Bayesian, which he applies to the development of model-based methods for survey analysis that are robust to misspecification, reasonably efficient, and capable of implementation in applied settings. His applied interests are broad, including mental health, demography, environmental statistics, biology, economics and the social sciences as well as biostatistics.

Trivellore E. Raghunathan, PhD

Title: Professor of Biostatistics; Research Professor, Survey Research Center, Institute for Social Research

Trivellore Raghunathan (Raghu) is Professor of Biostatistics and Director and Research Professor at the Institute for Social Research, also Research Professor in the Joint Program in Survey Methodology at the University of Maryland. He is the Director of Biostatistics Collaborative and Methodology Research Core (BCMRC), a research unit designed to foster collaborative and methodological research with the researchers in other departments in the School of Public Health and other allied schools. He holds additional appointments in the Biostatistics and Measurement Core for the Michigan CTSA, the Center for Research on Ethnicity, Culture and Health (CRECH), and the Center of Social Epidemiology and Population Health (CSEPH). He is an Elected Fellow of the American Statistical Association (ASA) and, among many awards, has received the SPAIG Award from the ASA. His primary research interests are in the analysis of incomplete data, multiple imputation, Bayesian methods, design and analysis of sample surveys, small area estimation, confidentiality and disclosure limitation, longitudinal data analysis and statistical methods for epidemiology. He has developed a SAS based software for imputing the missing values for a complex data set and can be downloaded from www.iveware.org.

APPENDIX B: Biosketches for Discussants

Mary Thompson, PhD

Title: Distinguished Professor Emerita, Department of Statistics and Actuarial Science, University of Waterloo.

Professor Thompson works primarily in survey methodology and sampling theory. Her book Theory of Sample Surveys describes the mathematical and foundational theory in detail; it also contains a systematic approach to using estimating functions in surveys, and a thorough discussion (with examples) of the role of the sampling design when survey data are used for analytic purposes. Estimation for stochastic processes has been another theme of her research. These two themes come together in aspects of inference from complex longitudinal surveys. Issues in the design of longitudinal surveys to support causal inference are central to work on the International Tobacco Control Survey, with which Professor Thompson has been involved since 2002. She studies the application of multilevel models and longitudinal models with time-varying covariates to complex survey data, including the best ways to adapt the estimating functions systems for use with survey weights, and the use of resampling techniques to provide accurate interval estimates. She is also currently collaborating on projects in survival and multistate models and the design of behavioural interventions on random networks. Professor Thompson has been active in the Statistical Society of Canada, serving as president in 2003-04. She is particularly interested in promoting collaboration between statisticians and researchers in the natural, health and social sciences. She was the initial Scientific Director of the Canadian Statistical Sciences Institute (CANSSI).

Jerome P. Reiter, PhD

Title: Professor & Chair, Department of Statistical Science, Duke University

Jerry Reiter is Department Chair and Professor of Statistical Science at Duke University. His primary areas of research include methods for ensuring data privacy, for handling missing and erroneous values, for combining information across sources, and for analyzing complex data in the social sciences and public policy. He is a Fellow of the American Statistical Association and a Fellow of the Institute of Mathematical Statistics. He is the recipient of several teaching and mentoring awards from Duke University, including the Alumni Distinguished Undergraduate Teaching Award, the Outstanding Postdoctoral Mentor Award, and the Masters of Interdisciplinary Data Science Distinguished Faculty Award. He has advised multiple government agencies on creating data products to share with the public, as well as served on multiple panels and committees for the National Academy of Sciences. He received a PhD in statistics from Harvard University in 1999.

S. Lynne Stokes, PhD

Title: Professor, Department of Statistics, Southern Methodist University

Lynne Stokes is Professor in the Department of Statistical Science at Southern Methodist University in Dallas. Her major areas of research are sampling and nonsampling errors in surveys, capture-recapture models, order statistics, and ranking models. She has served on several National Academy of Sciences' Panels considering sampling issues in federal surveys, with the most recent one entitled *A Vision and Roadmap for Education Statistics*. She has served on the Design and Analysis Committee for NAEP for 20 years, and on several NISS panels considering NCES issues. She was previously a mathematical statistician for the Census Bureau, where she worked on measurement errors in surveys. She is a member of Phi Beta Kappa, a Fellow of ASA, and a recipient of the Don Owen Award and Founder's Award from ASA.

F. Jay Breidt, PhD

Title: Senior Fellow, NORC at University of Chicago and Professor Emeritus, Colorado State University

F. Jay Breidt is a Senior Fellow in the Department of Statistics and Data Science at NORC at the University of Chicago, and Professor Emeritus and past Chair of the Department of Statistics at Colorado State University. His expertise and primary research interests are in survey sampling, time series, nonparametric regression, and uncertainty quantification for complex scientific models. Breidt has served as Reviews Editor for the Journal of the American Statistical Association and The American Statistician and has been an associate editor for eight different journals. He has served on numerous advisory panels, including six review committees for the National Academy of Sciences, two terms on the Federal Economic Statistics Advisory Committee, and two terms on the Census Scientific Advisory Committee, which he currently chairs. Breidt has received numerous honors, including being elected as Fellow in both the American Statistical Association and the Institute of Mathematical Statistics.

APPENDIX C: NISS Biosketches

Panel convened by National Institute of Statistical Sciences

***Nell Sedransk**, PhD*

Title: Director, National Institute of Statistical Sciences-DC

Nell Sedransk is the Director of the National Institute of Statistical Sciences. She is an Elected Member of the International Statistical Institute, also Elected Fellow of the American Statistical Association. She is coauthor of three technical books; and her research in both statistical theory and application appears in more than sixty scientific papers in refereed journals. The areas of her technical expertise include: design of complex experiments, Bayesian inference, spatial statistics and topological foundations for statistical theory. She has applied her expertise in statistical design and analysis of complex experiments and observational studies to a wide range of applications from physiology and medicine to engineering and sensors to social science applications in multi-observer scoring to ethical designs for clinical trials.

***Brian Habing**, PhD*

Title: Associate Director for Education Research at NISS, and Associate Professor of Statistics at University of South Carolina

Brian Habing is Associate Director for Education Activities and Research working with the DC Office of the National Institute of Statistical Sciences (NISS) and Associate Professor of Statistics at the University of South Carolina. His research has focused on psychometrics and scale construction, with a particular emphasis on multidimensional item response theory. His research focus includes analysis of education statistics and his interests also extend to statistical education, including work with AP Statistics and the development of new courses at the undergraduate and graduate level.

***Ya Mo**, PhD*

Title: Research Fellow, National Institute of Statistical Sciences; Assistant Professor, Boise State University

Ya Mo is a research fellow at the National Institute of Statistical Sciences and an assistant professor of Curriculum, Instruction, and Foundational Studies at Boise State University. She received a dual major PhD in Measurement and Quantitative Methods and Curriculum, Instruction, and Educational Policy Programs, and an M.S. in Statistics from Michigan State University, as well as an Ed.M. in TESOL from Boston University. She researches quantitative methods, psychometric measures, and survey statistics; she also applies quantitative research methods to study substantive topics in education, especially large-scale assessments.