Medical Expenditure Panel Survey

Household Component
Statistical Estimation Issues

Overview

- **Annual person-level estimates**
  - **Overlapping panels**
- **Estimation variables**
  - **Weights**
  - **Variance**
- **Pooling multiple years of annual data**
- **Longitudinal analysis of MEPS panels**
  - **Two-year period**
- **Family-level estimation**
- **Other miscellaneous issues**

Annual Person-Level Files

MEPS Annual Files

Year Panel	1997	1998	1999	2000	2001
1 (96-97)	Yr. 2				
2 (97-98)	Yr. 1	Yr. 2			
3 (98-99)		Yr. 1	Yr. 2		
4 (99-00)			Yr. 1	Yr. 2	
5 (00-01)				Yr. 1	Yr. 2
6 (01-02)					Yr. 1

MEPS Annual Files

Year	2002	2003	2004
Panel			
6 (01-02)	Yr. 2		
7 (02-03)	Yr. 1	Yr. 2	
8 (03-04)		Yr. 1	Yr. 2
9 (04-05)			Yr. 1

MEPS Annual Person Level Estimation

	1997	1998	1999	2000	2001
File Number	HC-020	HC-028	HC-038	HC-050	HC-060
Persons with weight > 0	32,636	22,953	23,565	23,839	32,122
Weighted Persons: All	271.3 million	273.5 million	276.4 million	278.4 million	284.2 million
INSC1231=1 (in target pop. at end of year)	267.7 million	270.1 million	273.0 million	275.2 million	280.8 million

MEPS Annual Person Level Estimation (continued)

	2002	2003	2004
File Number	HC-070	HC-079	HC-089
Persons with weight > 0	37,418	32,681	32,737
Weighted Persons: All	288.2 million	290.6 million	293.5 million
INSC1231=1 (in target pop. at end of year)	284.6 million	286.8 million	289.7 million

Weights and Variance Estimation Variables

MEPS Sample Design

- **Each panel is sub-sample of household respondents for the previous year's National Health Interview Survey (NHIS)**

  – **NHIS sponsor is National Center for Health Statistics**

- **NHIS sample based on complex stratified multi-stage probability design**

- **Civilian non-institutionalized population**

NHIS Sample Design (1995-2004)

- **U.S. partitioned into 1,995 Primary Sampling Units (Counties or groups of adjacent counties)**

- **PSU's grouped into 237 design strata**
  - **358 PSU's sampled across strata**

- **Second Stage Units (SSU's)**
  - **Clusters of housing units**
  - **Oversample of SSU's with large Black/Hispanic populations**

- **MEPS based on subsample of about 200 PSU's from NHIS**

Oversampling in MEPS

- **Every year: Blacks and Hispanics**
  - Carryover from NHIS

- **1997: Selected subpopulations**
  - Functionally impaired adults
  - Children with activity limitations
  - Adults 18-64 predicted to have high medical expenditures
  - Low income
  - Adults with other impairments

- **2002 and beyond:**
  - Asians
  - Low income
  - Additional oversampling of blacks in 2004

Estimation from Complex Surveys

- **Estimates need to be weighted to reflect sample design and survey nonresponse**
  - Unweighted estimates are biased

- **Use appropriate method to compute standard errors to account for complex design**
  - Assuming simple random sampling usually underestimates sampling error

Development of Person Weights

- **Base Weight (NHIS)**
  - Compensates for oversampling and nonresponse
- **Adjustments for**
  - Household nonresponse (MEPS Round 1)
  - Attrition of persons (Subsequent Rounds)
  - Poststratification (Census Population Estimates)
  - Trimming of extreme weights
- **Final Person Weight**
  - Weight > 0: person selected and in-scope for survey
  - Weight = 0 (about 5% in 2004): person not in-scope for survey but living in household with in-scope person(s)

Distribution of MEPS Sample Person Final Weights

	1997	1998	1999	2000	2001
Average	8,312	11,917	11,730	11,679	8,849
Minimum	299	321	307	454	336
Maximum	68,518	84,587	80,062	78,157	67,537
Variable Name	WTDPER97	WTDPER98	PERWT99F	PERWT00F	PERWT01F

Distribution of Sample Person Final Weights (continued)

	2002	2003	2004
Average	7,702	8,892	8,966
Minimum	367	401	425
Maximum	46,766	60,273	63,728
Variable Name	PERWT02F	PERWT03F	PERWT04F

Types of Basic Point Estimates

- **Means**

- **Proportions**

- **Totals**

- **Differences between subgroups**

Variance Estimation

- **Basic software procedures assume simple random sampling (SRS)**
  - MEPS not SRS
  - Point estimates correct (if weighted)
  - Standard errors usually too small

- **Software to account for complex design using Taylor Series approach**
  - SUDAAN (stand-alone or callable within SAS)
  - STATA (svy commands)
  - SAS 8.2 (survey procedures)
  - SPSS (new complex survey features in 13.0)

Estimation Example: Average Total Expenditures, 2004

- **Weighted mean = $3,284 per capita**
  - Unweighted mean of $2,944 is biased
- **SE based on Taylor Series = 89**
  - SAS: PROC SURVEYMEANS
  - SUDAAN: PROC DESCRIPT
  - Stata: svymean
- **SE assuming SRS = 56 (too low)**
  - SAS: PROC UNIVARIATE *or* MEANS

Computing Standard Errors for MEPS Estimates

- **Document on MEPS website**

- http://www.meps.ahrq.gov/mepsweb/survey_comp/standard_errors.jsp

Example (Point estimates and SEs): SAS V8.2

- **proc surveymeans data=work.h89 mean; stratum varstr; cluster varpsu; weight perwt04f; var totexp04;**

Example (Point estimates and SEs): SUDAAN (SAS-callable)

- **First need to sort file by varstr & varpsu**

- **proc descript data=work.h89 filetype=SAS design=wr;**
  nest varstr varpsu;
  weight perwt04f;
  var totexp04;

Example (Point estimates and SEs): Stata

svyset [pweight=perwt04f], strata(varstr) psu (varpsu)

svymean(totexp04)

Analysis of Subpopulations

- **Analyzing files that contain only a subset of MEPS sample may produce error messages or incorrect standard errors**

- **Each software package has capability to produce subpopulation estimates from entire person-level file**

- **See "Computing Standard Errors for MEPS Estimates"**

  - http://www.meps.ahrq.gov/mepsweb/survey_comp/standard_errors.jsp

Assessing Precision/Reliability of Estimates

- **Sample Sizes**

- **Standard Errors/Confidence Intervals**

- **Relative Standard Errors**
  - standard error of estimate ÷ estimate

Example: Average total expenses per capita, 2004

- **Sample Size = 32,737**
- **Estimate = $3,284**
- **Standard Error = 89**
- **95% Confidence Interval: (3109, 3458)**
- **Relative Standard Error (RSE) or Coefficient of Variation (CV) = 89 ÷ 3284 = .027 = 2.7%**

Types of Basic Point Estimates: Examples

- **Means**
  - Annual per capita expenses in 2004 = $3,284
- **Proportions**
  - Percent with some health expenses in 2004 = 84.7%
  - Two methods to generate estimates:
    - percents obtained from frequency tables
    - means of dichotomous variable
- **Totals**
  - Expenses in 2004 = $963.9 billion
  - Number of persons (weighted) = 293,527,003
- **Differences between subgroups**

Pooling Multiple Years of MEPS Data

Reasons for Pooling

- **Reduce standard error of estimate(s)**

- **Stabilize trend analysis**

- **Enhance ability to analyze small subgroups**

Minimum Sample Sizes

- **CFACT Standards**
  - **Minimum unweighted sample of 100**
  - **Flag estimates with RSE > 30%**
- **Confidence intervals become problematic with small samples and/or highly skewed data**
  - **Consider larger minimum sample sizes for highly skewed variables**
  - **Analysts may be comfortable with smaller minimums for less skewed variables**
  - **ASA Paper: Yu and Machlin (Skewness)**
    http://www.meps.ahrq.gov/mepsweb/data_files/publications/workingpapers/wp_04002.pdf

Example: Annual Sample Sizes (Unpooled)

Year	Total Population	Children 0-5	Asian/PI Children* 0-5
1996	21,571	2,018	58
1997	32,636	3,082	78
1998	22,953	2,114	82
1999	23,565	2,156	93

* Sample sizes do not meet AHRQ minimum requirement (n=100) to publish estimates.
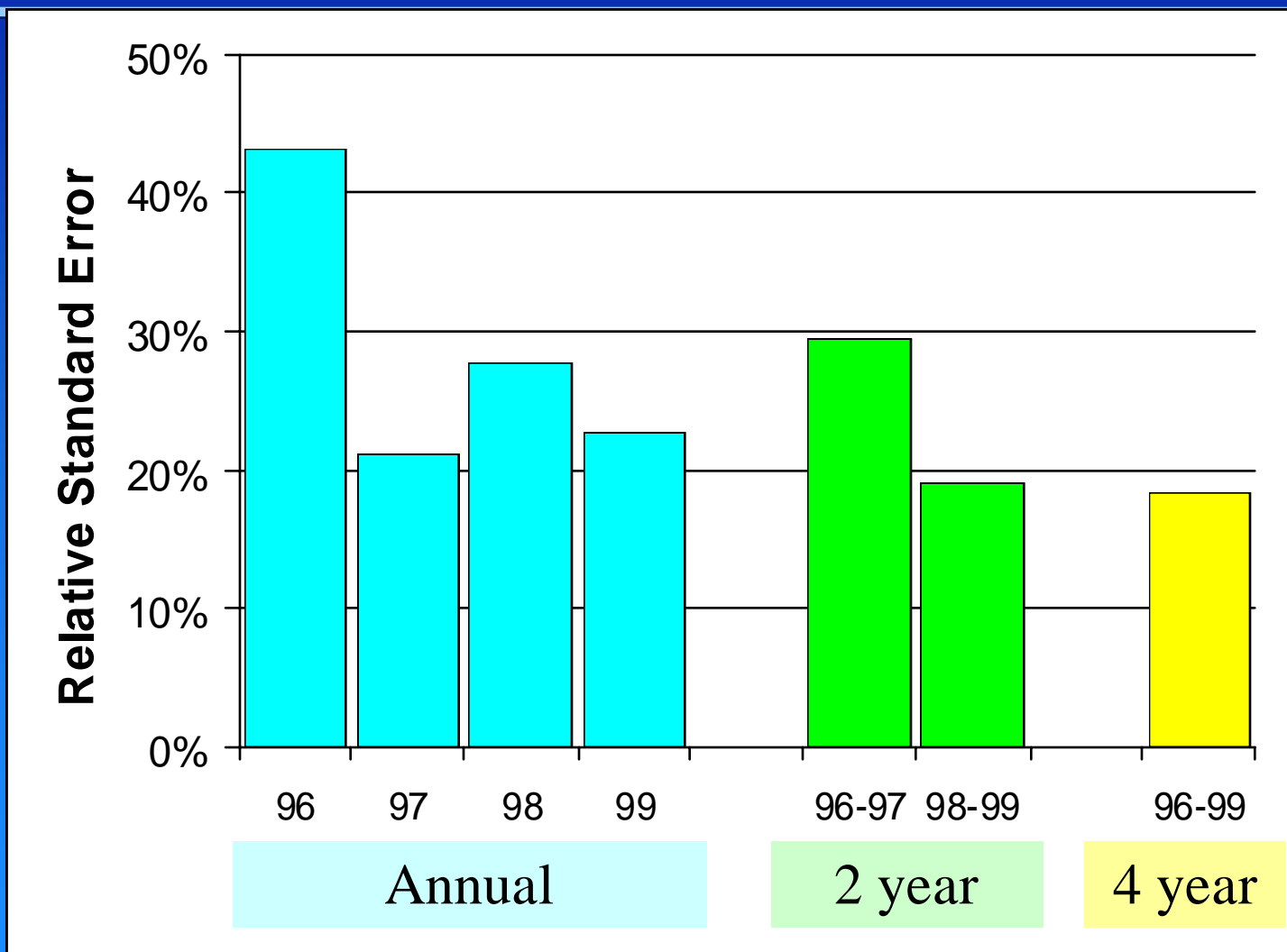
Pooled Sample Sizes

Years	Total Sample	Children 0-5	Asian/PI Children 0-5
1996-1997	54,207	5,100	136
1998-1999	46,518	4,270	175
1996-1999	100,725	9,370	311

Relative Standard Errors for Estimated Mean Expenditures: Asian/PI Children 0-5

Creating a Pooled File for Analysis (1996-2002)

- **Need to work with Pooled Estimation File (HC-036) when 1+ years being pooled include any year from 1996 through 2001**
  - **Stratum and PSU variables obtained from HC-036 for 1996-2004**
  - **Documentation for HC-036 provides instructions on how to properly create pooled analysis file**
- **Stratum (varstr) and PSU (varpsu) variables properly standardized for pooling years from 2002 onward (i.e., do not need HC-036)**

Creating Pooled Files: Summary of Important Steps

- **Rename analytic and weight variables from different years to common names.**
  - Expenditures: TOTEXP99 & TOTEXP00 = TOTEXP
  - Weights: PERWT99F & PERWT00F = POOLWT

- **Divide weight variable by number of years pooled to produce estimates for "an average year" during the period.**
  - Keep original weight value if estimating total for period

- **Concatenate annual files**

- **Merge variance estimation variables from HC-036 onto file (only if 1+ years prior to 2002)**
  - Strata variable: STRA9604
  - PSU variable: PSU9604

Estimates from Pooled Files

- **Produce estimates in analogous fashion as for individual years**
- **Estimates interpreted as "average annual" for pooled period**
- **Example: Pooled 1996-99 data**
  - **The average annual total health care expenditures for Asian/Pacific Islander children under 6 years of age during the period from 1996-1999 was $525 (SE=97).**

Pooling Annual Data: Lack of Independence Across Years

- **Legitimate to pool data for persons in consecutive years**
  - Each yr. constitutes nationally representative sample
  - Pooling produces average annual estimates
  - Stratum & PSU variables sufficient to account for lack of independence between years

- **Lack of independence actually begins with first stage of sample selection**
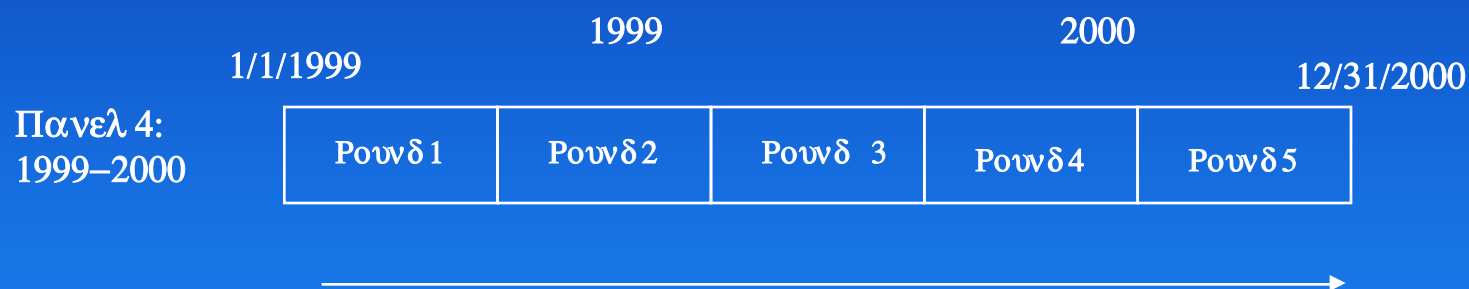 - Same PSUs are used to select each MEPS panel

- **See HC-036 documentation**

Longitudinal Analysis of MEPS Panels

MEPS Longitudinal Analysis: Panel 4: 1999-2000

1999

2000

1/1/1999

12/31/2000

Πανελ 4:
1999–2000

Ρουνδ 1	Ρουνδ 2	Ρουνδ 3	Ρουνδ 4	Ρουνδ 5

MEPS Longitudinal Analysis

- **National estimates of person-level changes over two-year period**
  - two-year period is relatively short
- **Examine characteristics associated with changes**
  - mainly round 1 data

Variables that may change between years or rounds

- **Insurance coverage**
  - Monthly indicators (24 measures)
  - Annual summary (2 measures per person)
- **Health status**
  - Each round (5 measures)
- **Having a usual source of care**
  - Rounds 2 & 4 (2 measures)
- **Use and expenditures**
  - Annual (2 measures per person)

MEPS Longitudinal Weight Files Currently Available (Oct 2007)

MEPS Panel	Years Covered	PUF Number
1	1996-97	HC-023
2	1997-98	HC-035
3	1998-99	HC-048
4	1999-00	HC-058
5	2000-01	HC-065
6	2001-02	HC-071
7	2002-03	HC-080
8	2003-04	HC-086

Creating Longitudinal Files (Panel 4) : Summary of Important Steps

- **Select Panel 4 records from annual files**
  - **1999 (PUF HC-038)**
  - **2000 (PUF HC-050)**

- **Obtain MEPS Longitudinal File (HC-058)**
  - **Contains weight and variance estimation variables**
  - **Contains variable indicating whether complete data are available for 1 or both years of panel**

- **Link using DUPERSID**

Longitudinal Weight

- **Variable Name: LONGWTP#**

- **Produces estimates for persons in civilian noninstitutionalized population in two consecutive years when applied to persons participating in both years of a given panel (YRINDP# = 1)**

Examples: Longitudinal Estimates

- **Of those without insurance at any time in 1999, estimated 76.9% (SE=1.6) also uninsured throughout 2000**

- **Estimated 8.2% (SE=0.4) of the population had no insurance throughout 1999-2000**

- **Of those with no expenses in 1999, estimated 47.6% (SE=1.3) had some expenses in 2000**

- **Of top 5% of spenders in 1996, 30% retain this position in 1997.**

Family-Level Estimation

Family-Level Estimation

- **Need to create families from person-level files (see documentation)**
- **Two family type options:**
  - MEPS: includes unmarried couples/foster children
  - CPS: unmarried couples not family unit
- **Two time frame options:**
  - December 31 (MEPS, CPS)
  - Any time during year (MEPS only)

MEPS Annual Files: Family Sample Sizes, 2004

	MEPS Full yr.	MEPS Dec 31	CPS Dec 31
Unweighted	13,018	12,913	13,349
Weighted	123.0 million	121.8 million	125.8 million
Family Weight Variable Name	FAMWT04F	FAMWT04F	FAMWT04C

Family-Level Estimation Example: Average Expenses per MEPS Family, 2004

- **Based on MEPS families in scope at any time during year**

- **Average number of persons per family is about 2.4.**

Family size	Estimate	SE
All	$7,674	187
1	$5,337	270
2	$9,670	370
3	$7,435	286
4	$8,815	702
5+	$8,265	405

Other Miscellaneous Estimation Issues

Special Supplements

- **Self Administered Questionnaire (SAQ)**
  - **Use SAQ weight**
- **Parent Administered Questionnaire (PAQ)**
  - **2000 only**
  - **Use PAQ weight**
- **Diabetes Care Survey (DCS)**
  - **Use DCS weight**
- **Variables on person-level files**
  - **Consult documentation for appropriate weight**