

# Methodological Issues in Using Multiple Years of the Medicare Current Beneficiary Survey

Becky A. Briesacher, Jennifer Tjia, Chyke A. Doubeni, Yong Chen, Sowmya R. Rao

University of Massachusetts Medical School, Meyers Primary Care Institute, HealthCore Inc. and Bedford VA Medical Center

## Supplement

doi: <http://dx.doi.org/10.5600/mmrr.002.01.s04>

## Technical Appendix

### Exhibit A1. List of Reviewed Papers Using MCBS

Author	Main study design	Main outcome measure	Main predictor	Statistical methods	MCBS files (CAU/ATC and years)	Weighting approach	Survey design approach	Repeat observations approach
Anderson(Anderson, Norton, & Dow, 2003)	Longitudinal	Medicare home care use (any visits and # of visits)	Billing practice	OLS, Probit	CAU 1992-1997	Weighted estimates, approach not described	Not described	Panel-data estimation
Anderson(Anderson, Norton, & Kenney, 2003)	Longitudinal	Changes in annual Medicare home health care expenditures	Billing practice	OLS, two-part random effects design	CAU 1992-1997	Not described	Not described	Panel-data estimation, bootstrapping SEs
Balsa(Balsa, Cao, & McGuire, 2007)	Longitudinal	1) Enrollment in managed care	Race/ethnicity	Bivariate probit, difference-in-difference	CAU 1996-2000; ATC 2001	Unweighted, sensitivity analyses using average of sampling weights	Not described	Fixed effects model

Author	Main study design	Main outcome measure	Main predictor	Statistical methods	MCBS files (CAU/ATC and years)	Weighting approach	Survey design approach	Repeat observations approach
Briesacher(Briesacher, Stuart, Ren, Doshi, & Wrobel, 2005)	Longitudinal	Changes in total health care expenditures	Gaining drug coverage	Fixed-effects panel estimator and difference in difference	CAU 1995-2000	Unweighted	Did not incorporate survey design	Fixed effects model
Cai(Cai & Lubitz, 2007)	Longitudinal	Changes in disability	Baseline ADLs, IADLs	Multistate life table model, multinomial logistic regression, discrete time hazard model	CAU1992-2003	Full-sample weight	Adjusted for PSU and strata, per Lohr 1999; bootstrapping SEs	Transition models
CDC(Centers for Disease Control and Prevention, 2004)	Repeated cross-sectional	Influenza vaccination rates	Time	Descriptive, no statistics	CAU 1991-2002	Not described	Used statistical software for survey data	No statistical tests over time
Chen(Chen et al., 2001)	Repeated cross-sectional	Cost of non-melanoma skin cancer (NMSC)	Time	Descriptive statistics	CAU 1992-1995	Not described	Not described	No statistical tests over time
Chin(Chin, Zhang, & Rathouz, 2003)	Longitudinal	Changes in health status	Baseline health perceptions	Ordinal logistic regression	CAU 1991-1994	Sampling weights for baseline estimates	STATA, survey estimators	Transition models, robust SE for clustering
Craig(Craig, Kreling, & Mott, 2003)	Repeated cross-sectional	Non-acquisition of prescribed medications	Time	Descriptive statistics	ATC 1996-1999	Not described	Not described	Not described

Author	Main study design	Main outcome measure	Main predictor	Statistical methods	MCBS files (CAU/ATC and years)	Weighting approach	Survey design approach	Repeat observations approach
Crystal(Crystal, Sambamoorthi, Walkup, & Akincigil, 2003)	Pooled	Depression diagnosis, antidepressant and psychotherapy use	Time	Logistic regressions	CAU 1992-1998	Weighted, but not described	SUDAAN	Adjusted for clustering within persons
Doshi(Doshi, Polsky, & Chang, 2007)	Repeated cross-sectional	Prevalence of obesity	Time	Descriptive statistics	CAU 1997-2002	respondent-level weighting	STATA, survey estimators	Not described
Ettner(Ettner, Hermann, & Tang, 1999)	Pooled	Psychiatric service utilization and expenditures	Mental health sector treatment vs. general medicine	Multinomial logit; 2-part models; logit model; 2-stage IV methods	Not described 1991-1993	Not described	Not described	Not described
Fiscella(Fiscella & Holt, 2007)	Pooled	Clinical preventive services	Race	Logistic regression	ATC 1998-2002	Survey weights	SAS survey estimators	Randomly selected 1 observation to eliminate repeat observations
Fiscella(Fiscella, Holt, Meldrum, & Franks, 2006)	Pooled	Self-reported and claims based clinical preventive services	Race	Logistic regression	ATC 1999-2002	Not described, survey weights	SAS survey estimators	Not described
Fisher(Fisher et al., 2003)	Longitudinal	5-year mortality rate, change in functional status and satisfaction	Region-level end of life spending	Cox proportional hazard model	CAU 1991-1996 and 1992-1995	Longitudinal weight, normalized	SUDAAN	Transition models, Robust SEs for clustering

Author	Main study design	Main outcome measure	Main predictor	Statistical methods	MCBS files (CAU/ATC and years)	Weighting approach	Survey design approach	Repeat observations approach
Foote(Foote & Hogan, 2001)	Pooled	Medicare costs by service type and health care costs	Payer type	Annual estimates averaged over 3 years	CAU 1994-1996	Not described	Not described	Not described
Hancox(Hancox et al., 2005)	Pooled	Skin pathology claims/	Time	Descriptive statistics	1992-2000	Not described	Not described	Not described
Hill(Hill, Fillit, Thomas, & Chang, 2006)	Pooled	Degree of functional impairment	Healthcare costs and prevalence of institutionalization	Linear and logistic regressions	CAU 1995-1998	Cross-sectional	Not described	Not described
Holt(Holt, Franks, Meldrum, & Fiscella, 2006)	Pooled	Self-reported mammography and claims mammography	Race	Logistic regression	ATC 1992-2002	Cross-sectional	SAS survey estimators	Not described
Hoover(Hoover, Crystal, Kumar, Sambamoorthi, & Cantor, 2002)	Pooled	Medical Expenditures in Last Year of Life	Time	Linear polynomial model	CAU 1992-1996	Sample weights	SUDAAN, taylor linearization	Adjusted for intraclass correlation, robust covariance
Housman(Housman et al., 2003)	Pooled	Episodes of care for non-melanoma skin cancer	Time	Descriptive statistics	Not described 1992-1995 claims	Not described, weighted estimates	Not described	Not described

Author	Main study design	Main outcome measure	Main predictor	Statistical methods	MCBS files (CAU/ATC and years)	Weighting approach	Survey design approach	Repeat observations approach
Iezzoni(Iezzoni, Davis, Soukup, & O'Day, 2004)	Longitudinal	Changes in physical and sensory functioning	Prior year characteristics	Logistic models	ATC 1996-1997	sampling weights	SUDAAN	Transition model
Karnal-Bahl(Kamal-Bahl, Stuart, & Beers, 2005)	Repeated cross-sectional	Receipt of propoxyphene	Drug coverage and physician characteristics	Descriptive statistics + logistic regression	CAU 1993-1999	Not described, weighted estimates	STATA, survey estimators	Not described
Kemper(Kemper, Weaver, Short, Shea, & Kang, 2008)	Pooled	Not getting help with an ADL	Medicare home care spending, individual income, and other individual characteristics	Linear probability model	Not described 1992,1996,2000	Cross-sectional	STATA, survey estimators	Fixed effect models
Lakdawalla(Lakdawalla et al., 2003)	Longitudinal and repeated cross-sectional	Admission to nursing home, changes in disability and nursing home use	Baseline disability status and marital status	Life-table analysis and logit models	CAU 1992-1996	Normalized weights	Not described.	Robust SEs
Laschober(Laschober, Kitchman, Neuman, & Strabic, 2002)	Repeated cross-sectional	Drug coverage	Beneficiary characteristics	Logistic regression	ATC 1996-1999	Not described	Survey software	Not described

Author	Main study design	Main outcome measure	Main predictor	Statistical methods	MCBS files (CAU/ATC and years)	Weighting approach	Survey design approach	Repeat observations approach
Lubitz(Lubitz, Cai, Kramarow, & Lentzner, 2003)	Longitudinal	Expected health expenditures and life expectancy	Baseline self-rated health Status	Multivariate hazard model	CAU 1992-1998	Cross-sectional weight	Bootstrapping for SEs	Transition models and bootstrapping for SEs
Lundy(Lundy & Craig, 2006)	Repeated cross-sectional	Disease modifying agents for Multiple Sclerosis (MS)	Time	Descriptive statistics	CAU 1995-2002	Not described	Not described	Not described
McCormack(McCormack et al., 2002)	Repeated cross-sectional	Supplemental coverage	Beneficiary characteristics	Multinomial logistic regression	ATC 1992, 1995-1998	Not described	SUDAAN	No statistical tests over time
McKnight(McKnight, 2006)	Repeated cross-sectional	Long term care utilization and health outcomes	Change of Medicare reimburse policy	OLS Probit	CAU 1992-1999	Not described	Not described	Fixed effect models
Mello(Mello, Stearns, Norton, & Ricketts, 2003)	Longitudinal	HMO enrollment	Previous year's Health status, HMO market penetration	Random effects probit model	CAU 1992-1996	Unweighted, longitudinal weights used in sensitivity analyses	Did not incorporate survey design	Panel data estimation, robust Huber-White SEs
Mello(Mello, Stearns, & Norton, 2002)	Longitudinal	Hospital stay, number of inpatient days, physician visits	Prior health plan choice (Medicare FFS or HMO)	2-stage model/discrete factor model	CAU 1993-1995; ATC 1996	probability weighted, longitudinal weights for sensitivity analyses	Not described	Panel data estimation, robust Huber-White corrected SEs

Author	Main study design	Main outcome measure	Main predictor	Statistical methods	MCBS files (CAU/ATC and years)	Weighting approach	Survey design approach	Repeat observations approach
Murray(Murray & Eppig, 2002)	Repeated cross-sectional	Insurance coverage	Time	Descriptive statistics	ATC 1991-1999	Weighted estimates, but not described	Not described	No statistical tests over time
Noyes(Noyes, Liu, & Holloway, 2006)	Pooled	Parkinsonism diagnosis	Neuroleptic use	Logistic model	CAU 1992-2000	survey weights	STATA, survey estimators	Statistical methods to adjust for repeat observations
Noyes(Noyes, Liu, Li, Holloway, & Dick, 2006)	Pooled	Medical utilization and economic burden – adjusted medical expenditures	Time	2-part logit model	Not described, 1992-2000	Survey weights	STATA, survey estimators	Robust Huber-White SEs
Noyes(Noyes, Liu, Holloway, & Dick, 2007)	Pooled	Diagnosis of Parkinson’s disease and medical expenditures	Case identification algorithms	Comparison of means	CAU 1992-2000	Not described	Not described	Robust Huber-White SEs
O’Malley(O’Malley & Forrest, 2006)	Pooled	Immunizations	Patient factors	Hierarchical logistic regression models, for clustering within physician	ATC 2000-2002	Sampling weights	SUDAAN	Hierarchical logistic regressions used to assess clustering
Patel(Patel & Davis, 2006)	Pooled	Drug Expenditures	Time	Descriptive statistics	CAU 1997-2001	Not described, weighted estimates	STATA, survey estimators	Treated each response as independent

Author	Main study design	Main outcome measure	Main predictor	Statistical methods	MCBS files (CAU/ATC and years)	Weighting approach	Survey design approach	Repeat observations approach
Porell(Porell & Miltiades, 2001b)	Longitudinal	1-year change in functional state or death	Baseline HMO vs. FFS supplementary insurance status	Multinomial logit model,	ATC 1991-1996	Normalized weights used in sensitivity analyses	STATA, survey estimators in sensitivity analyses	Transition models, bootstrapping SEs
Porell(Porell & Miltiades, 2001a)	Longitudinal	1-year change in functional state or death	Baseline health insurance status	Multinomial logit model,	ATC 1991-1996	Normalized weights used in sensitivity analyses	STATA, survey estimators in sensitivity analyses	Transition models, bootstrapping SEs
Rice(Rice, Snyder, Kominski, & Pourat, 2002)	Pooled	Medigap switch to Medicare managed care plan	Affordability, need for services, health insurance, sociodemographics, supply of managed care plans	Multivariate logistic regression	CAU 1993-1996	Not described	SUDAAN	Adjusted for multiple observations, sensitivity analysis with Cox models
Riley(G. F. Riley, Lubitz, & Zhang, 2003)	Pooled	Reasons for disability	Health care utilization	Descriptive statistics	CAU 1995-1999	Cross-sectional	SUDAAN	SEs were increased to account for dependence, per Kish 1965
Riley(G. Riley & Zarabozo, 2006)	Repeated cross-sectional	Health and functional status	Time and enrollment (risk contract enrollee or FFS)	Weighted logistic regression	ATC 1991-2004	Cross-sectional	SEs were increased to account for complex sampling, Bye 1994	SEs were increased to account for panel, per Bye 1994



Author	Main study design	Main outcome measure	Main predictor	Statistical methods	MCBS files (CAU/ATC and years)	Weighting approach	Survey design approach	Repeat observations approach
Rosenbach(Rosenbach, Acamache, & Khandker, 1995)	Repeated cross-sectional and pooling	Health care utilization (ER visit, OP visit and physician visit)	Predisposing, enabling and need factors	Weighted Logistic, least square regression	ATC 1991-1993	Not described	SUDAAN	Faced with the choice of controlling for panel data or complex sampling design
Sambamoorthi(Sambamoorthi, Olfson, Walkup, & Crystal, 2003)	Repeated cross-sectional and Pooled	Receipt of newer antidepressants	Time	Sequential binary logistics, probit	CAU 1992-1997	Cross-sectional and longitudinal weights	SUDAAN, STATA, survey estimators	Extra level of clustering for individuals per Liang 1986
Shang(Shang & Goldman, 2007)	Longitudinal	Drug spending	Drug coverage	Instrumental variable analysis -> drug coverage	CAU 1992-2000	Not described	Not described	Fixed effect models, robust SE
Simoni-Wastila(Simoni-Wastila, Zuckerman, Shaffer, Blanchette, & Stuart, 2008)	Pooled	Mental health prescription drug fills for 3 years	Drug Coverage Discontinuity	Zero-inflated negative binomial model	CAU 1997-2001	Unweighted	STATA, survey estimators	Not described.
Stein(Stein, Sloan, & Lee, 2007)	Repeated cross-sectional	Glaucoma medication utilization	NA	Descriptive statistics	CAU 1992-2002	Not described	Not described	Not described
Stuart(Stuart, Simoni-Wastila, & Chauncey, 2005)	Pooled	Drug spending for 3 years	Drug coverage Discontinuity	OLS and generalized linear model	CAU 1998-2000	Not described	Not described	Not described

Author	Main study design	Main outcome measure	Main predictor	Statistical methods	MCBS files (CAU/ATC and years)	Weighting approach	Survey design approach	Repeat observations approach
Stuart(Stuart et al., 2003)	Repeated cross-sectional	Inappropriate drug use (exposure to Beers drug, drug on list of highly severe outcome & on list of less severe outcome)	Demographic characteristics, income, drug coverage, geographic area, health status, chronic conditions.	Logistic regression	CAU 1995 and 1999	Cross-sectional	STATA, survey estimators	Minimal overlap of samples
Tosteson(Tosteson, Gottlieb, Radley, Fisher, & Melton, 2007)	Longitudinal	1 year mortality	Hip fracture; pre-fracture health and functional status	Cox proportional hazard model	CAU 1996-2000	Average of MCBS cross-sectionals	Not described	Transition model
Waidmann(Waidmann & Liu, 2000)	Pooled	Trends in rates of activities of daily living and IADLs	Time	Multinomial logit and least squares regression	ATC 1992-1996	Not described, weighted estimates	STATA, survey estimators	Modeled as correlation within persons
Wei(Wei, Sambamoorthi, Olfson, Walkup, & Crystal, 2005)	Pooled	Psychotherapy	Time	Multilevel logistic regression	Not described, 1992-1999	Used only baseline weights	SUDAAN	Linearization methods to account for intracluster correlations
Yang(Z. Yang, Norton, & Stearns, 2003)	Repeated cross-sectional	Health care expenditures	Longevity	Stratified population means	CAU 1992-1996	Cross-sectional	Adjusted for complex survey design	Bootstrapping SEs

Author	Main study design	Main outcome measure	Main predictor	Statistical methods	MCBS files (CAU/ATC and years)	Weighting approach	Survey design approach	Repeat observations approach
Yang(Zhou Yang, Gilleskie, & Norton, 2004)	Longitudinal	Changes in Health care expenditure and mortality	Drug coverage	2 part expendtr. model (logit and log linear), dynamic multinomial logit model, regressions	CAU 1992-1998	Not described	Not described	Panel models

SOURCE: Data collected by author.

#### Exhibit A2. Weighted analyses for four designs using SAS and SUDAAN Analytic\_file (hypothetical data sorted by year)

Baseid	ADL_cont	Year	Varstrata	Varunit	Sudstrata	Sudunit	Avg_weight*	CS1YRWGT§	Mid-year	
									weight†	L3YRSWGT‡
25	1	2005	41	1	10,485	4588	3,579	3,579	.	.
31	3	2005	28	2	10,355	3073	2,057	2,078	1,982	4,278
47	0	2005	96	2	30,360	98	1,937	1,605	2,096	4,181
75	1	2006	86	1	30,235	324	1,659	1,623	1,623	3,655
97	1	2006	73	1	20,335	314	2,033	1,946	1,946	4,197
16	0	2006	41	1	10,375	3428	3,242	3,464	3,464	7,375
53	0	2007	37	1	10,145	684	2,201	1,965	2,179	4,715
75	2	2007	86	1	30,235	324	1,659	1,574	1,623	3,655
97	1	2007	73	1	20,335	314	2,033	2,071	1,946	4,197

\* Avg\_weight=Sum of cross-sectional weights/ number of years in the survey

People in 05 only=05 cross-sectional weight/1; People in 05 and 06 = (05 cross-sectional weight +06 cross-sectional weight)/2; People in 05, 06 and 07 = (06 cross-sectional weight +06 cross-sectional weight + 07 cross-sectional weight)/3; People in 06 and 07 = (06 cross-sectional weight +07 cross-sectional weight)/2; People in 07 only= 07cross-sectional weight/1

§ CS1YRWGT = cross-sectional weights from each year's data

People in 2005: 05 cross-sectional weights; People in 2006: 06 cross-sectional weights; People in 2007: 07 cross-sectional weights

† Mid-year weight = cross-sectional weights (CS1YRWGT)from the middle year (2006)

‡ L3YRSWGT = longitudinal weights from 2007 file (rix3)

**Pooled: (Exhibit 2a)**

```

/*SAS*/
proc surveymeans data= Analytic_file;
cluster varunit;
strata varstrat;
var ADL_cont;
weight weight;
run;
/*SUDAAN*/
proc descript data= Analytic_file filetype=sas design=wr;
nest sudstrat sudunit;
weight weight;
var adl_cont;
setenv decwidth=4 colwidth=10;
print nsum="SAMSIZE" mean semean/nsumfmt=F9.0 meanfmt=F10.4 semeanfmt=F10.4 style=nchs;
run;

```

**Repeat cross-sectional: (Exhibit 2b)**

Longitudinal1 (individuals with repeated measures of ADL): (Figure 2c)  
 Longitudinal2 (individuals who have been in the MCBS for 3 years): (Figure 2d)

```

/*SAS*/
proc surveyreg data= Analytic_file;
cluster VARUNIT;
strata VARSTRAT;
weight CS1YRWGT;
*replace CS1YRWGT with Mid-year weight for Longitudinal1 design;
*replace CS1YRWGT with l3yrswg for Longitudinal2 design;
class year;
model adl_cont=year/ noint solution vadjust=none;
estimate 'Diff 2005-2006' year 1 -1 0;
estimate 'Diff 2006-2007' year 0 1 -1;
run;
/*SUDAAN*/
proc descript data= Analytic_file filetype=sas design=wr;
nest sudstrat sudunit;
weight CS1YRWGT;
*replace CS1YRWGT with Mid-year weight for Longitudinal1 design;
*replace CS1YRWGT with l3yrswg for Longitudinal2 design;
class new_year;
var adl_cont;
setenv decwidth=4 colwidth=10;
print nsum="SAMSIZE" mean semean t_mean p_mean/nsumfmt=F9.0 meanfmt=F10.4 semeanfmt=F10.4
style=nchs;
run;
proc descript data= Analytic_file filetype=sas design=wr;
nest sudstrat sudunit;

```

```
weight CS1YRWGT;
*replace CS1YRWGT with Mid-year weight for Longitudinal1 design;
*replace CS1YRWGT with l3yrswg for Longitudinal2 design;
class new_year;
var adl_cont;
contrast new_year = (-1 1 0) / name = "2005 versus 2006";
contrast new_year = (0 -1 1) / name = "2006 versus 2007";
setenv decwidth=4 colwidth=10;
print nsum="SAMSIZE" mean semean t_mean p_mean/nsumfmt=F9.0 meanfmt=F10.4 semeanfmt=F10.4
style=nchs;
run;
```

SOURCE: data are hypothetical and are only used for illustration purpose. SAS and SUDAAN codes are written by author.

---

## Medicare & Medicaid Research Review

2012  
Volume 2, Number 1

---

### Mission Statement

*Medicare & Medicaid Research Review* is a peer-reviewed, online journal reporting data and research that informs current and future directions of the Medicare, Medicaid, and Children's Health Insurance programs. The journal seeks to examine and evaluate health care coverage, quality and access to care for beneficiaries, and payment for health services.

<http://www.cms.gov/MMRR/>

---

### *U.S. Department of Health & Human Services*

Kathleen Sebelius,  
Secretary

### *Centers for Medicare & Medicaid Services*

Marilyn Tavenner  
Acting Administrator

### *Center for Strategic Planning*

Anthony D. Rodgers,  
Deputy Administrator and Director

*Editor-in-Chief* David M. Bott, Ph.D.

*Senior Editor* Cynthia Riegler, M.A.

### *Editorial Board*

Gerald S. Adler, M.Phil.  
CMS/Center for Strategic Planning  
Andrew Bindman, M.D.  
University of California, San Francisco  
William J. Buczko, Ph.D.  
CMS/Innovation Center  
Todd Caldis, Ph.D., J.D.  
CMS/Office of the Actuary  
Craig F. Caplan, M.A.  
CMS/ Innovation Center  
Melissa A. Evans, Ph.D.  
CMS/Center for Medicare  
John Hsu, M.D., M.B.A, M.S.C.E.  
Harvard Medical School

Jesse M. Levy, Ph.D.  
CMS/ Innovation Center  
Jennifer Polinski, Sc.D, M.P.H.  
Brigham & Women's Hospital  
Isidor R. Strauss, F.S.A.  
CMS/Office of the Actuary  
Fred G. Thomas, Ph.D., C.P.A.  
CMS/ Innovation Center  
Robert Weech-Maldonado, Ph.D.  
University of Alabama at Birmingham  
Rashid Bashshur, PhD.  
University of Michigan Health System

*Submission Guidelines:* [www.cms.gov/MMRR/InformationforAuthors](http://www.cms.gov/MMRR/InformationforAuthors)

*Contact:* [mmrr-editors@cms.hhs.gov](mailto:mmrr-editors@cms.hhs.gov)

Published by the Centers for Medicare & Medicaid Services

*All material in the Medicare & Medicaid Research Review is in the public domain and may be duplicated without permission.*

*Citation to source is requested.*