

6TH INTERNATIONAL CONFERENCE ON

HEALTH POLICY RESEARCH

Methodological Issues in
Health Services and
Outcomes Research

October 28–30, 2005

BOSTON
MASSACHUSETTS

Sponsored by



nchs National Center for
Health Statistics... Monitoring
the Nation's Health



MISSION



The Health Policy Statistics Section (HPSS) of the American Statistical Association was founded in 1994. Our primary objective is to improve the quality, effectiveness, and efficiency of health care in the United States and abroad through the systematic use of quantitative statistical methods. HPSS fosters the development of statistical methods specifically designed for health care and policy decisionmaking by organizing and sponsoring conferences and workshops on methodological issues relevant to health care and policy research. It also cooperates with government agencies to develop ways to analyze their data, offering assistance to legislators and their staff members with their decisionmaking process.

Specific Activities of the HPSS Include:

- Organizing and sponsoring invited, contributed, and luncheon sessions at the Joint Statistical Meetings to foster statistical methodological research relevant to health care and policy research.
- Organizing the biennial International Conference on Health Policy Research (ICPHR).
- Sponsoring student paper contests with various research organizations and societies, encouraging students to develop interest in statistical methods for health services and policy research.
- Publishing proceedings volumes from the JSM and ICHPR meetings to facilitate communication among government, academia, and private sectors.
- Sponsoring Continuing Education courses and workshops to encourage and facilitate ASA members' involvement in health care and policy research, both methodologically and substantively.
- Cooperating with government agencies—such as the Agency for Health Care Research and Quality (AHRQ), the National Center for Health Statistics (NCHS), and the Centers for Medicare & Medicaid Services—to establish ways to utilize and disseminate their data for the study of health policy issues.
- Offering assistance to legislators and their staff members regarding health statistics and their relationship to decisionmaking.
- The HPSS maintains a list server (hpss-news) to facilitate communication among colleagues. To join, go to <http://www.amstat-online.org/sections/hpss/news.htm> and follow the instructions.

For more information, visit our web site at
www.amstat-online.org/sections/hpss/hpssindex.htm.

SCIENTIFIC PLANNING COMMITTEE

Co-Chairs

Arlene Ash, Boston University

Tom Belin, University of California, Los Angeles

COMMITTEE MEMBERS

Gerald Adler, Centers for Medicare & Medicaid Services

Anirban Basu, University of Chicago

Joe Cappelleri, Pfizer Inc.

Steve Cohen, Agency for Healthcare Research and Quality

Bonnie Ghosh-Dastidar, RAND Corporation

Frank E. Harrell, Jr., Vanderbilt University

Alberto Holly, University de Lausanne

Mary Beth Landrum, Harvard Medical School

Stacy Lindborg, Eli Lilly and Company

Lisa Lix, University of Manitoba

Jim Lubitz, National Center for Health Statistics

Alistair James O'Malley, Harvard Medical School

Christopher H. Schmid, Tufts-New England Medical Center

Julianne Soucek, Houston VA Medical Center

Doug Staiger, Dartmouth College

Thérèse A. Stukel, Institute for Clinical Evaluative Sciences

Recai Yucel, University of Massachusetts-Amherst

ADVISORY COMMITTEE

Naihuan Duan, University of California, Los Angeles

Kim McGuigan, Pfizer Inc.

Sharon-Lise Normand, Harvard Medical School

Carolyn Rutter, Group Health Cooperative

Alan Zaslavsky, Harvard Medical School

CONFERENCE PLANNING ASSISTANT

Ann Farrell, Boston University

ASA REPRESENTATIVES

Margarita E. Navas

Elaine L. Powell

TABLE OF CONTENTS

1	Committee Members
2	Conference at a Glance
3	Hotel Floor Plans
4	Friday, October 28, 2005
6	Saturday, October 29, 2005
9	Sunday, October 30, 2005
11	Workshops, Luncheon, and Reception
13	Overall Conference Agenda
14	Proceedings Publication Instructions
15	Abstracts

The Program Committee and the Health Policy Statistics Section thank the following organizations for their generous support of the ICHPR 2005:

*Agency for Healthcare Research and Quality, an Agency of the Department of Health and Human Services
National Center for Health Statistics*

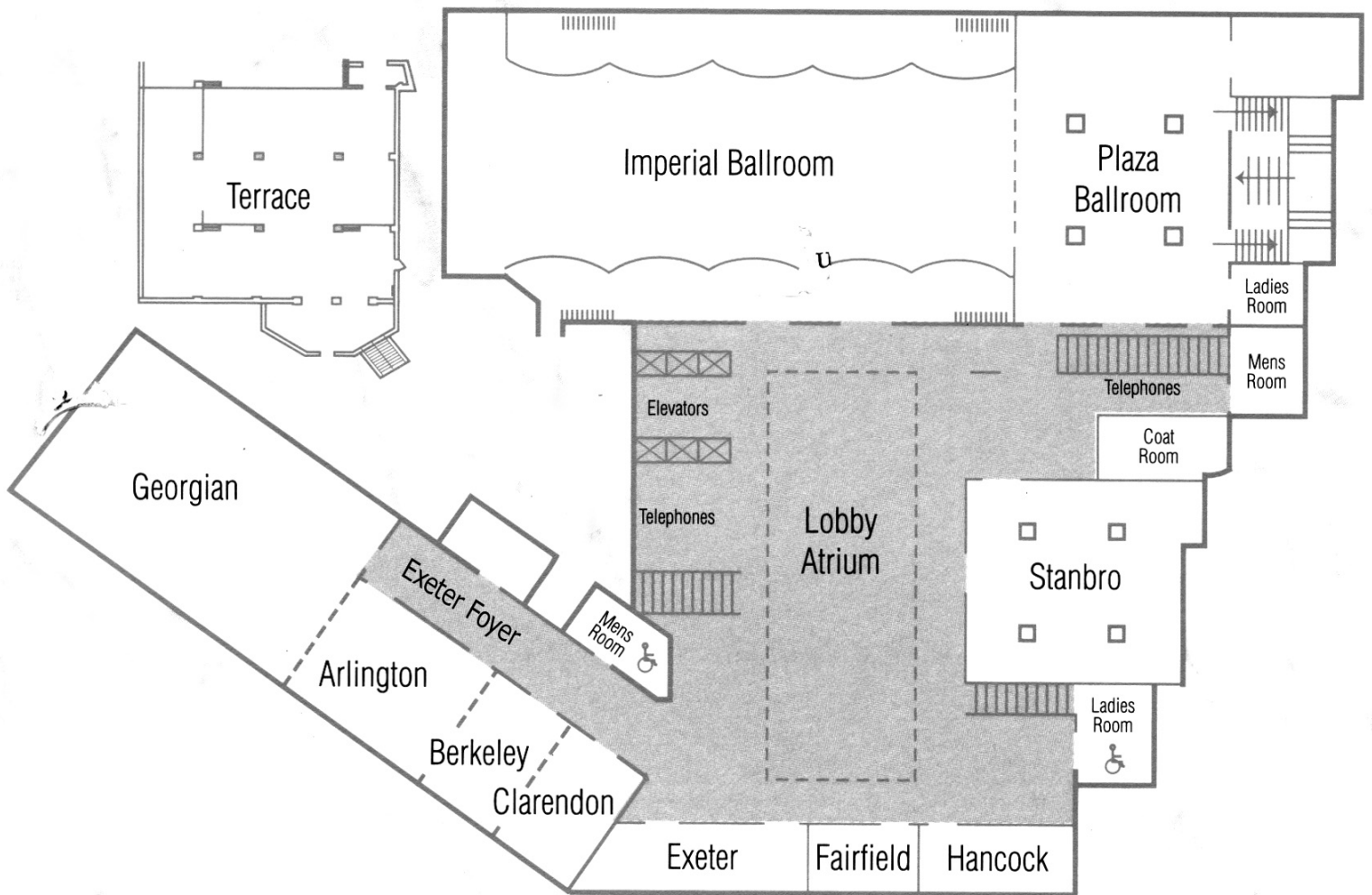
Conference at a GLANCE

Friday October 28	Arlington	Berkeley	Clarendon	Fairfield	Stanbro	Georgian	Exeter Foyer	Exeter
7:00 a.m. 8:30 a.m.	Breakfast on own							
8:30 a.m. 10:15 a.m.	C-1	C-2	W-1	Empty	Empty	Empty	Registration	Office/Speaker Work Room
10:15 a.m. 10:30 a.m.	Coffee Break - Refresh meeting rooms						Coffee Break	
10:30 a.m. 12:15 p.m.	C-3	W-3	W-2	Empty		I-1	Registration	
12:30 p.m. 2:00 p.m.	Refresh meeting rooms				Luncheon	Empty		
2:15 p.m. 4:00 p.m.					Empty	Plenary Session		
4:00 p.m. 4:15 p.m.	Coffee Break					Empty	Coffee Break	
4:15 p.m. 6:00 p.m.	I-2	C-4	W-4	W-5			Registration	
6:30 p.m. 8:00 p.m.	Offsite reception at Mary Baker Eddy Library							

Saturday October 29	Arlington	Berkeley	Clarendon	Stanbro	Exeter Foyer	Exeter
7:00 a.m. 8:30 a.m.	Refresh meeting rooms				Continental Breakfast	Office/ Speaker Work Room
8:30 a.m. 10:15 a.m.	I-3	W-6	W-7	C-5		
10:15 a.m. 10:30 a.m.	Coffee Break - Refresh meeting rooms				Coffee Break	
10:30 a.m. 12:15 p.m.	I-4	W-9	W-8	C-6	Registration	
12:30 p.m. 2:00 p.m.	Lunch on own - Refresh meeting rooms					
2:15 p.m. 4:00 p.m.	I-5	C-7	W-10	I-6		
4:00 p.m. 4:15 p.m.	Coffee Break - Refresh meeting rooms				Coffee Break	
4:15 p.m. 6:00 p.m.	I-7	C-8	W-11	I-8	Registration	

Sunday October 30	Arlington	Berkeley	Clarendon	Exeter Foyer	Exeter
7:00 a.m. 8:30 a.m.	Refresh meeting rooms			Continental Breakfast	Office/ Speaker Work Room
8:30 a.m. 10:15 a.m.	I-9	C-9	W-12	Registration	
10:15 a.m. 10:30 a.m.	Coffee Break - Refresh meeting rooms			Coffee Break	
10:30 a.m. 12:15 p.m.	I-10	C-10	W-13	Empty	

MEZZANINE



INTERNATIONAL CONFERENCE ON HEALTH POLICY RESEARCH

October 28–30, 2005 • Boston Park Plaza Hotel • Massachusetts

FRIDAY, OCTOBER 28, 2005

BREAKFAST (On own)

7:00 a.m.–8:30 a.m.

REGISTRATION

7:00 a.m.–6:00 p.m.

Exeter Foyer, Mezzanine Level

CONCURRENT SESSIONS

8:30 a.m.–10:15 a.m.

W-1 (Workshop–Fee event)

Clarendon, Mezzanine Level

Statistical Graphics for Exploring Data, Presenting Information, and Understanding Statistical Models (Part I)

Frank E. Harrell, Jr., Vanderbilt University

C-1 (Contributed)

Arlington, Mezzanine Level

Propensity Score Methods and Analysis of Noncompliance

Chair: Steven Scott, University of Southern California

8:35 a.m. Assessing Physician Performance: Critical First Step in Pay-for-performance—Sherrie Kaplan*, UCI School of Medicine

8:50 a.m. *Not Doing It Until I Do?* Determining the Efficacy of Virginity Pledges—Janet Rosenbaum*, Harvard University

9:05 a.m. Propensity Score Methodology Combined with Modified ANCOVA: An Example—Bradley E. Huitema*, Western Michigan University; Joseph McKean, Western Michigan University

9:20 a.m. Too Much Ado about Propensity Score Matching?—Onur Baser*, Thomson-Medstat

9:35 a.m. Noncompliance Bias Correction Based on Covariates in Randomized Experiments—Yves Atchade*, University of Ottawa; Leonard Wantchekon, New York University

9:50 a.m. Principal Stratification Approach to Dealing with Treatment Noncompliance and Subsequent Nonresponse—Robert Ware*, The University of Queensland

10:05 a.m. Floor Discussion

* = PRESENTING AUTHOR

C-2 (Contributed)

Berkeley, Mezzanine Level

Imputation Methods

Chair: Kara Bambuer, Harvard Medical School

8:35 a.m. *The Impact of Expanded Medicaid Eligibility for Pregnant Women on Health Care Utilization and Outcome*—John Engberg*, RAND Corporation; Donna Farley, RAND Corporation; Lisa Shugarman, RAND Corporation

8:50 a.m. *Robustness of a Multivariate Normal Approximation for Imputation of Incomplete Binary Data*—Tom Belin*, University of California, Los Angeles; Coen Bernaards, Genentech, Inc.; Joseph Schafer, Pennsylvania State University

9:05 a.m. Comparison of Hot Deck and Multiple Imputation Methods Using HCSDB Data—Donsig Jang*, Mathematica Policy Research, Inc.; Amang Sukasih, Mathematica Policy Research, Inc.

9:20 a.m. Semiparametric Approach for Multiple Imputations of Unobserved Values in Longitudinal Studies—Yulei He*, Harvard Medical School; Trivellore Raghunathan, University of Michigan

9:35 a.m. Multiple Imputation for Correcting Verification Bias in Estimating Sensitivity and Specificity—Ofer Harel*, University of Connecticut; Andrew Zhou, University of Washington

9:50 a.m. Floor Discussion

BREAK

10:15 a.m.–10:30 a.m.

Exeter Foyer, Mezzanine Level

CONCURRENT SESSIONS

10:30 a.m.–12:15 p.m.

W-2 (Workshop–Fee event)

Clarendon, Mezzanine Level

Statistical Graphics for Exploring Data, Presenting Information, and Understanding Statistical Models (Part II)

Frank E. Harrell, Jr., Vanderbilt University

W-3 (Workshop–Fee event)

Berkeley, Mezzanine Level

Strategies for Using Propensity Scores Well

Thomas E. Love, Case Western Reserve University

I-1 (Invited)

Georgian, Mezzanine Level

Combining Estimates/Information Using

Multiple Data Sources

Organizers: Recai Yucel, University of Massachusetts-Amherst, and Tom Belin, University of California, Los Angeles

Chair: Recai Yucel, University of Massachusetts-Amherst

10:35 a.m. Obtaining Cancer Risk Factor Prevalence Estimates in Small Areas—Michael Elliott*, University of Pennsylvania

11:00 a.m. New Modeling Strategies for Combining Data from Multiple Surveys To Obtain Small Domain Estimates of Obesity—Dawei Xie*, University of Pennsylvania

11:25 a.m. Maximum Likelihood Estimation of Multiple Source Predictor Regression Models—Nicholas J. Horton*, Smith College; G. M. Fitzmaurice, Harvard University; T. L. Lash, Boston University; N. M. Laird, Harvard University

11:50 a.m. Discussant: Donald B. Rubin, Harvard University

12:10 p.m. Floor Discussion

C-3 (Contributed)

Arlington, Mezzanine Level

Insights for Health Studies from Surveys and Psychometric Research

Chair: Norma Terrin, Tufts-New England Medical Center

10:35 a.m. Survey Conditioning in Self-reported Mental Health Service Use: Results from a Randomized Trial—Naihua Duan*, University of California, Los Angeles

10:50 a.m. Self-rated Health among Foreign- and Native-born Individuals: A Test of Comparability—Elena Erosheva*, University of Washington; Daniel Takeuchi, University of Washington; Emily Walton, University of Washington

11:05 a.m. An Experiment To Explain What Influences Clinical Decisions—Carol Link*, New England Research Institutes; Lisa Marceau, New England Research Institutes; John McKinley, New England Research Institutes; Amy O'Donnell, New England Research Institutes

11:20 a.m. Latent Class Structure of IQ in Preschool Children with Autism—Elizabeth Koehler*, University of Washington, Robert Abbott; Ted Beauchaine; Geraldine Dawson; Catherine Lord; Jeffrey Munson; Sally Rogers; Marian Sigman; Andrew Zhou, University of Washington

11:35 a.m. Evaluating the Sample Invariance Property of the Standard Error of Measurement—Joseph Cappelleri*, Pfizer Inc.; Andrew Bashmakin, Pfizer Inc.; Samiran Ghosh, University of Connecticut; William Lenderking, Pfizer Inc.

11:50 a.m. Floor Discussion

LUNCHEON (Fee event)

12:30 p.m.–2:00 p.m.

Stanbro, Mezzanine Level

Chair: Thomas E. Love, Case Western Reserve University

Your Money or Your Life—David Cutler, Otto Eckstein Professor of Applied Economics and Dean for the Social Sciences, Harvard University

PLENARY SESSION

2:15 p.m.–4:00 p.m.

Georgian, Mezzanine Level

Assessing Pharmaceutical Safety and Efficacy in the Wake of COX-2 and HRT

Organizers: Frank Harrell, Vanderbilt University, and Thérèse A. Stukel, Institute for Clinical Evaluative Sciences

Chair: Thérèse A. Stukel, Institute for Clinical Evaluative Sciences

Panelists: Muhammad Mamdani*, Institute for Clinical Evaluative Sciences; Alan Breier*, Eli Lilly and Company; Robert O'Neill*, U.S. Food and Drug Administration; Frank Harrell*, Vanderbilt University

3:45 p.m. Floor Discussion

BREAK

4:00 p.m.–4:15 p.m.

Exeter Foyer, Mezzanine Level

CONCURRENT SESSIONS

4:15 p.m.–6:00 p.m.

W-4 (Workshop–Fee event)

Clarendon, Mezzanine Level

Modern Metaanalysis

Christopher H. Schmid, Tufts-New England Medical Center

W-5 (Workshop–Fee event)

Fairfield, Mezzanine Level

Privacy, Confidentiality, and Data Security Training for Health Services Research

Alan M. Zaslavsky, Harvard University

I-2 (Invited)

Arlington, Mezzanine Level

Methods of Risk Adjustment for Skewed Outcome Data

Organizer: Julianne Soucek, Houston VA Medical Center

Chair: Julianne Soucek, Houston VA Medical Center

4:20 p.m. Addressing Skewness and Kurtosis in Risk Adjustment—Alberto Holly*, University of Lausanne; Yevhen Pentsak, University of Lausanne

4:45 p.m. Using Diagnosis-based Risk Adjustment and Self-reported Health Status To Predict Mortality—Kenneth Pietz*, VA Medical Center and Baylor College of Medicine; Laura A. Petersen, VA Medical Center and Baylor College of Medicine

5:10 p.m. Risk Adjustment with Flexible Link and Variance Function Models—Anirban Basu*, University of Chicago; Bhakti Arondekar, GlaxoSmithKline; Paul Rathouz, University of Chicago

5:35 p.m. Discussant: Andrew Zhou, University of Washington

5:50 p.m. Floor Discussion

C-4 (Contributed)

Berkeley, Mezzanine Level

Competing Risks and Advanced Survival Modeling in Health Services Research

Chair: Paula Diehr, University of Washington

4:20 p.m. The Influence of Spousal Morbidity and Mortality on Proband Mental Health Diagnosis and Treatment—Kara Bambauer*, Harvard Medical School; Nicholas Christakis, Harvard Medical School

4:35 p.m. Differences in Regression-based Decomposition Depending on Functional Form—Jinn-Ing Liou*, University of Wisconsin-Madison; Michael Finch, Finch & King, Inc.; Jennifer Frylak, Ingenix; Nilay Shih, University of Wisconsin-Madison; Maureen Smith, University of Wisconsin-Madison

4:50 p.m. Floor Discussion

OFFSITE EVENT (Fee event)

6:30 p.m.– 8:00 p.m.

Mary Baker Eddy Library

SATURDAY, OCTOBER 29, 2005

CONTINENTAL BREAKFAST

7:00 a.m.–8:30 a.m.

Exeter Foyer, Mezzanine Level

REGISTRATION

7:30 a.m.–5:30 p.m.

Exeter Foyer, Mezzanine Level

CONCURRENT SESSIONS

8:30 a.m.–10:15 a.m.

W-6 (Workshop—No fee)

Berkeley, Mezzanine Level

Research Opportunities Using AHRQ Databases

Karen Beauregard, Agency for Healthcare Research and Quality

W-7 (Workshop—Fee event)

Clarendon, Mezzanine Level

Advances in Latent Variable Modeling (Part I)

Bengt Muthen, University of California, Los Angeles

I-3 (Invited)

Arlington, Mezzanine Level

Advanced Methods for Estimating Health Disparities

Organizers: Anirban Basu, University of Chicago, and Douglas Staiger, Dartmouth College

Chair: Douglas Staiger, Dartmouth College

8:35 a.m. Racial Disparities in Self-rated Health at Older Ages: The Contribution of Neighborhood-level Factors—Kathleen Cagney*, University of Chicago; Christopher Browning, The Ohio State University; Ming Wen, The Ohio State University

9:00 a.m. Valuation of Arthritis Health States across Ethnic Groups and between Patients and Community Members—Julianne Soucek*, Baylor College of Medicine; Margaret Byrne, University of Pittsburgh; Adam Kelly, Baylor College of Medicine; Marsha Richardson, Baylor College of Medicine; Chong Pak, Baylor College of Medicine; Harlan Nelson, Baylor College of Medicine; Maria Suarez-Almazor, Baylor College of Medicine; Michael E. DeBakey, VA Medical Center

9:25 a.m. Implementing the IOM Definition of Disparities: An Application to Mental Health Care—Benjamin L. Cook*, Harvard University; Thomas G. McGuire, Harvard University; Margarita Alegria, Harvard University; Kenneth B. Wells, University of California, Los Angeles; Alan Zaslavsky, Harvard University

9:50 a.m. Discussant: Amitabh Chandra, Harvard University

10:10 a.m. Floor Discussion

C-5 (Contributed)

Stanbro, Mezzanine Level

Cost, Risk, and Allocation of Health Care Resources

Chair: Michael Stoto, RAND Corporation

8:35 a.m. Causes and Consequences of Regional Variations in Health Care Resources in Ontario—Thérèse A. Stukel*, Institute for Clinical Evaluative Sciences

8:50 a.m. Adjusting SARS-affected Data for Canadian Inpatient Case Mix Indicators—Sheril Perry*, Canadian Institute for Health Information; Qian Yang, Canadian Institute for Health Information; Douglas Yeo, Canadian Institute for Health Information

9:05 a.m. Why Are We Still Using Charlson To Measure Comorbidity?—Jeanne Speckman*, Boston University Medical Center; Arlene Ash, Boston University School of Medicine; Jennifer Fonda, Boston University Medical Center; Amresh Hanchate, Boston University School of Medicine; Nancy McCall, Research Triangle Institute; Thomas Williams, TRICARE Management Activity, HPA&E

9:20 a.m. Propensity Score Modeling of Antibiotics from Inpatient Data—Michael O'Connell*, Insightful

9:35 a.m. Analysis Methods for Volume-outcome Studies—Katherine Panageas*, Memorial Sloan-Kettering Cancer Center

9:50 a.m. A New, Nonparametric Method for Predicting Health Care Costs with Heteroscedasticity in Risk-adjustment Models—Andrew Zhou*, University of Washington; Hauzhen Lin

10:05 a.m. Floor Discussion

BREAK

10:15 a.m.–10:30 a.m.

Exeter Foyer, Mezzanine Level

CONCURRENT SESSIONS

10:30 a.m.–12:15 p.m.

W-8 (Workshop-Fee event)

Clarendon, Mezzanine Level

Advances in Latent Variable Modeling (Part II)

Bengt Muthen, University of California, Los Angeles

W-9 (Workshop-no fee)

Berkeley, Mezzanine Level

Research Opportunities Using Data from the CDC National Center for Health Statistics

Jim Lubitz and Robert Weinzimer, CDC National Center for Health Statistics

I-4 (Invited)

Arlington, Mezzanine Level

Predicting High-cost Users of Medical Care and the Persistence of High Expenditures over Time

Organizer: Steve Cohen, Agency for Healthcare Research and Quality

Chair: Joseph Cappelleri, Pfizer Inc.

10:35 a.m. Using the SF-12 To Predict Health Care Expenditures—John Fleishman*, Agency for Healthcare Research and Quality; Joel Cohen, Agency for Healthcare Research and Quality; Mark Kosinski, Quality Metric

11:00 a.m. An Evaluation of the Performance of Prediction Models To Identify High-expenditure Cases—Steven B. Cohen*, Agency for Healthcare Research and Quality; Trena Ezzati-Rice, Agency for Healthcare Research and Quality; William Yu, Agency for Healthcare Research and Quality

11:25 a.m. The Impact of Diagnosis Accuracy on Predictive Power of Cost Prediction Models Using the MEPS—Arlene Ash*, Boston University; Joel Cohen, Agency for Healthcare Research and Quality; John Fleishman, Agency for Healthcare Research and Quality

11:50 a.m. Discussant: Michael Davern, University of Minnesota

12:10 p.m. Floor Discussion

C-6 (Contributed)

Stanbro, Mezzanine Level

Metaanalysis Methods

Chair: Donsig Jang, Mathematica Policy Research, Inc.

10:35 a.m. Bayesian Metaanalysis of the Dose-response Relationship of Alcohol Consumption and Health Outcomes—Michael Stoto*, RAND Corporation; Graham Colditz, Brigham and Women's Hospital; Sharon-Lise Normand, Harvard Medical School

10:50 a.m. Fixed and Random Effects Sequential Models for Monitoring Quality of Care—Karl Heiner*, SUNY New Paltz; Bruce Agins, New York State Department of Health

11:05 a.m. Separating the Effects of Publication Bias and Heterogeneity in Systematic Reviews—Norma Terrin*, Tufts-New England Medical Center; Michael Dowd, Tufts-New England Medical Center; Christopher Schmid, Tufts-New England Medical Center

11:20 a.m. Hospital Ownership and Financial Performance: An Integrative Research Review—Karen Eggleston*, Tufts University; Yu-Chu Shan, Naval Postgraduate School

11:35 a.m. Quality Review of Recently Published Metaanalyses in a Five-year Period—Kelly H. Zou*, Harvard Medical School/Brigham and Women's Hospital; Jui Bhagwat, Brigham and Women's Hospital; Jacqueline Campbell, Brigham and Women's Hospital; Daniel Goldberg-Zimring, Brigham and Women's Hospital; Ferenc Joesz, Brigham and Women's Hospital; Lucila Ohno-Machado, Brigham and Women's Hospital

11:50 a.m. The State of the Art in Metaanalyses of Diagnostic Tests—Christopher Schmid*, Tufts-New England Medical Center; Mei Chung, Tufts-New England Medical Center; Joseph Lau, Tufts-New England Medical Center; Athina Tatsioni, Tufts-New England Medical Center

12:05 p.m. Floor Discussion

LUNCH (on own)
12:30 p.m.–2:00 p.m.

CONCURRENT SESSIONS
2:15 p.m.–4:00 p.m.

W-10 (Workshop-Fee event)

Clarendon, Mezzanine Level

Issues When Using Hierarchical Models To Estimate Provider Performance

Michael Shwartz and Arlene Ash, Boston University

I-5 (Invited)

Arlington, Mezzanine Level

Imputation in High-dimensional Complex Surveys

Organizers: Tom Belin, University of California, Los Angeles, and Recai Yucel, University of Massachusetts-Amherst

Chair: Tom Belin, University of California, Los Angeles

2:20 p.m. Multiple Imputation Using Chained Hierarchical Models—Recai Yucel*, University of Massachusetts-Amherst

2:50 p.m. Multiple Imputation by Ordered Monotone Blocks: The Case of the Anthrax Vaccine Clinical Trial—M. Baccini, University of Florence; S. R. Cook, Columbia University; C. Frangakis, Johns Hopkins University; F. Li, Johns Hopkins University; Fabrizia Mealli*, University of Florence; D. B. Rubin, Harvard University

3:20 p.m. Multiple Imputation of Missing Income Data in the National Health Interview Survey—Nathaniel Schenker*, National Center for Health Statistics

3:50 p.m. Floor Discussion

I-6 (Invited)

Stambro, Mezzanine Level

Statistical Issues in the Hospital CAHPS (HCAHPS) Survey

Organizer: James O'Malley, Harvard Medical School

Chair: Roopa Seshadri, Northwestern University

2:20 p.m. Overview of HCAHPS, Design of the HCAHPS Instrument, and Political Issues—Paul Cleary*, Harvard Medical School

2:45 p.m. Issues Concerning Sample Size Calculation and Reporting—Marc Elliott*, RAND Corporation

3:10 p.m. Hierarchical Factor Analysis for Survey Data with Structured Nonresponse—James O'Malley*, Harvard Medical School

3:35 p.m. Discussant: Ron Hays, RAND Corporation

3:50 p.m. Floor Discussion

C-7 (Contributed)

Berkeley, Mezzanine Level

Advanced Methods for Assessing Intervention Effects

Chair: Carol Link, New England Research Institutes

2:20 p.m. Estimating Drug Effects in Claims Data Using the Prescribing Physician as an Instrumental Variable—M. Alan Brookhart*, Brigham and Women's Hospital; Sebastian Schneeweiss, Brigham and Women's Hospital

2:35 p.m. Prescription Drug Insurance and Its Effect on Prescription Drug Utilization and Health of the Elderly—Nasreen Khan*, University of Illinois at Chicago; Robert Kaestner, University of Illinois at Chicago; Swu Jane Lin, University of Illinois at Chicago

2:50 p.m. Evaluation of a Nursing Home Informatics Tool To Reduce Adverse Outcomes: Methodological Issues—Richard Gardiner*, New York Association of Homes and Services for the Aging

3:05 p.m. Professional Caregiver Insurance Risk: Implications of Health Care Provider Insurance Risk Assumption—Thomas Cox*, Seton Hall University College of Nursing; Colene Byrne, New York Association of Homes and Services for the Aging

3:20 p.m. Using CHAID for Instrument Development and Practice Guidelines—James Bost*, University of Pittsburgh

3:35 p.m. A Case Example of Data Mining and Causal Analysis—Surprising? True? Useful?—Andrew Brunskill*, University of Washington

3:50 p.m. Floor Discussion

BREAK

4:00 p.m.–4:15 p.m.

Exeter Foyer, Mezzanine Level

CONCURRENT SESSIONS

4:15 p.m.–6:00 p.m.

W-11 (Workshop-Fee event)

Clarendon, Mezzanine Level

Risk Adjustment and Predictive Modeling

Randall P. Ellis, Boston University

I-7 (Invited)

Arlington, Mezzanine Level

Causal Inference with Longitudinal Data

Organizers: Mary Beth Landrum, Harvard Medical School, and Arlene Ash, Boston University

Chair: Arlene Ash, Boston University

4:20 p.m. The Analysis of Sequential Treatments: Practical Issues—Babette Brumback*, University of Florida

4:50 p.m. Random Effects Logistic Models for Analyzing Efficacy of a Longitudinal Randomized Treatment with non-Adherence—Dylan Small*, University of Pennsylvania

5:20 p.m. Inferring Causal Effects in Clustered Longitudinal Data: The Effect of Publicly Reporting Outcomes in Cardiac Surgery—Mary Beth Landrum*, Harvard Medical School; Robert S. Huckman, Harvard University; David M. Cutler, Harvard University and NBER

5:35 p.m. Discussant: James O'Malley, Harvard Medical School

5:50 p.m. Floor Discussion

I-8 (Invited)

Stanbro, Mezzanine Level

Population Needs-based Funding Models

Organizers: Lisa Lix, University of Manitoba, and Thérèse A. Stukel, Institute for Clinical Evaluative Sciences

Chair: Lisa Lix, University of Manitoba

4:20 p.m. Health Care Funding Models Based on Population Needs: The UK Experience—Peter C. Smith*, University of York

4:45 p.m. National Population Needs-based Funding: The New Zealand Experience—Peter Crampton*, University of Otago

5:10 p.m. Allocating Health Care Resources According to Need: An Approach to Developing Needs-based Formulas Using Linked Health Survey and Administrative Data—Jeremiah Hurley*, McMaster University

5:35 p.m. Discussant: Thérèse A. Stukel, Institute for Clinical Evaluative Sciences

5:55 p.m. Floor Discussion

C-8 (Contributed)

Berkeley, Mezzanine Level

Multilevel Models and Bayesian Methods

Chair: Naihua Duan, University of California, Los Angeles

4:20 p.m. Hierarchical Modeling of Inpatient Rehabilitation Facility Responses to Prospective Payment—Susan Paddock*, RAND Corporation

4:35 p.m. Extending the Capture-recapture Methodology To Estimate Subpopulation Sizes—Ulysses Diva*, University of Connecticut; Dipak Dey, University of Connecticut

4:50 p.m. GEE and Summary Measures Analysis in Medication Use over Time in Asthmatic Pregnant Women—Tebeb Gebretsadik*, Vanderbilt University; Patrick Arbogast, Vanderbilt University; Tina Hartert, Vanderbilt University; Ayumi Shintani, Vanderbilt University

5:05 p.m. Methods for Profiling the Value of Hospital Care following Acute Myocardial Infarction—Justin Timbie*, Harvard Medical School; Sharon-Lise Normand, Harvard Medical School

5:20 p.m. Model Selection versus Information Selection for Performance Evaluation: Where Best To Invest?—Kevin L. Sloan*, U.S. Veterans Health Administration/University of Washington; James F. Burgess, Jr., U.S. Veterans Health Administration; Xiao-Hua (Andrew) Zhou, University of Washington; Chuan Zhou, Vanderbilt University; Paul Fishman, Group Health Cooperative; Li Wang, U.S. Veterans Health Administration

5:35 p.m. Hidden Markov Models for Longitudinal Comparisons—Steven Scott*, University of Southern California; Gareth James, University of Southern California; Catherine Sugar, University of Southern California

5:50 p.m. Floor Discussion

SUNDAY, OCTOBER 30, 2005

CONTINENTAL BREAKFAST

7:00 a.m.–8:30 a.m.

Exeter Foyer, Mezzanine Level

REGISTRATION

8:00 a.m.–10:30 a.m.

Exeter Foyer, Mezzanine Level

CONCURRENT SESSIONS

8:30 a.m.–10:15 a.m.

W-12 (Workshop-Fee event)

Clarendon, Mezzanine Level

Bayesian Hierarchical Modeling with Applications to Provider Profiling (Part I)

David Draper, University of California, Santa Cruz

I-9 (Invited)

Arlington, Mezzanine Level

Methods in Longitudinal Data Analysis

Organizer: Jim Lubitz, National Center for Health Statistics

Chair: Jim Lubitz, National Center for Health Statistics

8:35 a.m. Validation of Life Table Approaches to Estimating Population Health Status—Liming Cai*, National Center for Health Statistics

9:00 a.m. New Findings on non-Response from the Medicare Current Beneficiary Survey (MCBS)—John Kautter*, RTI

9:25 a.m. Handling Incomplete Data in Longitudinal Clinical Trials—Geert Molenberghs*, Limburgs Universitair Centrum

9:50 a.m. Discussant: Joe Schafer, Pennsylvania State University

10:10 a.m. Floor Discussion

C-9 (Contributed)

Berkeley, Mezzanine Level

Health Care Cost and Payment Systems

Chair: John Engberg, RAND Corporation

8:35 a.m. Specification of Regression Models in the Development of Inpatient Case Mix Grouping Methodology—Qian Yang*, Canadian Institute for Health Information; Jeff Hatcher, Canadian Institute for Health Information

8:50 a.m. Estimating Incremental Cost-effectiveness Ratios from Cluster-randomized Intervention Trials—Mohammad Chaudhary*, Johns Hopkins Bloomberg School of Public Health; Mohamad Shoukri, King Faisal Specialist Hospital and Research Center

9:05 a.m. Estimating the Effects of Drug co-Payments on Statin Adherence Using Cross-sectional Time Series Data—Teresa Gibson*, Medstat; Kirsten Axelson, Pfizer Global Pharmaceuticals; Tami Mark, Medstat; Kimberly McGuigan, Pfizer Global Pharmaceuticals; Shaohung Wang, Medstat

9:20 a.m. Effect of Health Interventions on Longevity, Morbidity, Years of Healthy Life, and Costs—Paula Diehr*, University of Washington; Liming Cai, U.S. Centers for Disease Control and Prevention; Ann Derleth, University of Washington; Anne Newman, University of Pittsburgh

9:35 a.m. Competing Risk Analysis Applied to Health Economic Evaluations—George Carides*, Merck Research Laboratories; Shannan Allen, Merck Research Laboratories

9:50 a.m. Health Care Costs following Treatment Initiation for Alzheimer's Disease (AD) in Managed Care—Ann Harada*, Prescription Solutions; Ann Vanderplas, Prescription Solutions

10:05 a.m. Floor Discussion

BREAK

10:15 a.m.–10:30 a.m.

Exeter Foyer, Mezzanine Level

CONCURRENT SESSIONS

10:30 a.m.–12:15 p.m.

W-13 (Workshop-Fee event)

Clarendon, Mezzanine Level

Bayesian Hierarchical Modeling with Applications to Provider Profiling (Part II)

David Draper, University of California, Santa Cruz

I-10 (Invited)

Arlington, Mezzanine Level

Selection Bias in Observational Studies

Organizers: Anirban Basu, University of Chicago, and Douglas Staiger, Dartmouth College

Chair: Anirban Basu, University of Chicago

8:35 a.m. More Ado about Two: Endogenous Switching, Sample Selection, Endogenous Treatment Effects, and the Modified Two-part Model—Joseph V. Terza*, University of Florida

9:00 a.m. Two Approaches to Adjusting for Selection Bias in Cross-sectional Data: Instrumental Variables and Propensity Score—Matthew L. Maciejewski*, University of Washington; Song Wang; Xiao-Hua (Andrew) Zhou, University of Washington

9:25 a.m. Causal Estimation Using Quasi-experimental Designs—Sharon-Lise T. Normand*, Harvard Medical School; Richard G. Frank, Harvard Medical School; Thomas G. McGuire, Harvard Medical School

9:50 a.m. Discussant: Douglas Staiger, Dartmouth College

10:10 a.m. Floor Discussion

C-10 (Contributed)

Berkeley, Mezzanine Level

Health Disparities and Access to Care

Chair: Yu-Chu Shen, Naval Postgraduate School

10:35 a.m. Access to High-cost Medicines in Australia: Evaluating Health Outcomes Using National Claims Data—Christine Lu*, University of New South Wales; Ric Day, University of New South Wales; Ken Williams, University of New South Wales

10:50 a.m. Racial Disparities in Primary Care and Health Care Utilization at the End of Life—Andrea Kronman*, Boston University Medical Center; Arlene Ash, Boston University School of Medicine; Karen Freund, Boston University Medical Center

11:05 a.m. Health Insurance Availability and Racial Disparities in Total Knee Arthroplasty (TKA)—Amresh Hanchate*, Boston University School of Medicine; Arlene Ash, Boston University School of Medicine

11:20 a.m. Health Infrastructure and Rural Immunization in India—Arnab Mukherji*, Pardee-RAND Graduate School; Ashlesha Dator, RAND Corporation; Necaj Sood, RAND Corporation

11:35 a.m. Medicare Beneficiaries' Access to Physician Services in Local Markets in 2003—Anne Ciemnecki*, Mathematica Policy Research, Inc.; Tim Lake, Mathematica Policy Research, Inc.

11:50 a.m. Floor Discussion

Conference Adjourns—12:15 p.m.

WORKSHOPS

Fee Events, \$60 each; Students—\$30 each

Workshops on statistical methods and introductions to important national research databases will be offered in conjunction with the conference.

Workshops are available at additional cost and require advance registration.

Workshops #6 and #9 are free, but require advance registration.

FRIDAY, OCTOBER 28, 2005

(WK1) 8:30 a.m.-10:15 a.m.—Statistical Graphics for Exploring Data, Presenting Information, and Understanding Statistical Models (Part I); Frank E. Harrell, Jr., Vanderbilt University

Graphical methods are increasingly used for exploratory data analysis. Useful graphical tools in this setting include scatterplot matrices, nonparametric smoothers, and tree diagrams. We will use graphical horror stories from the scientific and lay press to illustrate that most graphics used in papers, presentations, and the popular media today—such as bar charts and pie charts—communicate quantitative information poorly. Then, we will discuss elements of graphical perception and good graph construction, many from the writings of Bill Cleveland, with practical suggestions for choosing the best chart or graph type, making good and clear graphics, formatting, and simultaneously presenting multiple variables. Nonstatisticians do not grasp easily complex outcome or risk adjustment models. We will discuss and show examples of effect charts and nomograms, graphics that help physicians and other consumers of statistical analyses understand statistical models, and use them to obtain predictions for individual subjects.

(WK2) 10:30 a.m.-12:15 p.m.—Statistical Graphics for Exploring Data, Presenting Information, and Understanding Statistical Models (Part II); Frank E. Harrell, Jr., Vanderbilt University

In Part II, we will interactively demonstrate how to make effective statistical graphics in the freely available R environment for data analysis and graphics (www.r-project.org) and share some graphical marvels (especially from Edward Tufte and Howard Wainer).

(WK3) 10:30 a.m.-12:15 p.m.—Strategies for Using Propensity Scores Well; Thomas E. Love, Case Western Reserve University

This intermediate-level workshop describes and demonstrates effective strategies for using propensity score analysis in causal modeling. Attendees should be familiar with basic risk adjustment, logistic regression, and the use of the propensity score to deal with selection bias in observational studies. We will begin with a brief review of propensity score methods, and then discuss strategies for estimating the propensity score effectively, assessing and displaying covariate balance, choosing analytic techniques, and communicating results to a nonstatistical audience. The instructor will use examples from health policy and health services research to motivate and illustrate ideas. Time permitting, additional (advanced)

CONFERENCE LUNCHEON

\$20 for registrants (part of the luncheon cost is included in the registration fee)

\$45 for guests (actual cost)

Friday, October 28

12:30 p.m.—2:00 p.m.

Boston Park Plaza Hotel & Towers

The luncheon speaker will be David Cutler, Otto Eckstein Professor of Applied Economics and Dean for the Social Sciences at Harvard University. His recent book, "Your Money or Your Life: Strong Medicine for America's Health Care System" (Oxford Press), was praised as "a clear and concise guide to how one should think about the costs and benefits of health care, the value of medical advances, and options for reforming the health care system."

MENU

Spring greens and julienne vegetables, honey mustard dressing, Boston baked scrod with lemon thyme butter sauce, and New York-style cheesecake

Vegetarian Option—

Vegetable Napoleon: layered roasted vegetables, including portabella mushroom, eggplant, zucchini, yellow squash, red pepper, and marinara sauce, served with steamed rice

OFFSITE EVENT

**Reception at the
Mary Baker Eddy Library
Friday, October 28
6:30 p.m.—8:00 p.m.**

The Mary Baker Eddy Library for the Betterment of Humanity, located a short distance from the Boston Park Plaza Hotel & Towers, features a variety of exhibits on the history and power of ideas. Mary Baker Eddy was an influential, 19th-century American author, teacher, and religious leader noted for her groundbreaking ideas about spirituality and health, which she named Christian Science. She also founded the Church of Christ, Scientist, in Boston, Massachusetts, in 1879. The library has exhibits that review motivating factors, such as the medical use of blood-letting and other now-discarded techniques, and the library's Mapparium offers a widely acclaimed visual display of the spread of ideas across the globe. A fee of \$20 for registrants and guests is required for this event. More details can be found at www.marybakereddylibrary.org.

topics will be discussed. Attendees will receive detailed handouts and access to software developed at the Center for Health Care Research and Policy.

(WK4) 4:15 p.m.–6:00 p.m.—Modern Metaanalysis; Christopher H. Schmid, Tufts-New England Medical Center

While the fixed versus random effects debate still crops up, most statisticians have moved beyond this and now are concerned with methods for exploring heterogeneity, whether in efficacy trials or diagnostic test studies. Topics such as metaregression, indirect comparisons, baseline rate regression, and summary ROC curves have been discussed in both the statistical and clinical literature. Most analysts now recognize that random effects models are necessary to describe most sets of studies, as these usually display considerable heterogeneity. Many of the models proposed recently have included a Bayesian component because many of them are most easily formulated as hierarchical structures that can be fit most easily with Markov chain Monte Carlo simulation. We will describe uses of hierarchical models in a variety of practical applications, many with aspects of missing data; contrast Bayesian and nonBayesian approaches and discuss the choice of appropriate prior distributions; and use examples from the literature to demonstrate methods and software choices.

(WK5) 4:15 p.m.–6:00 p.m.— Privacy, Confidentiality, and Data Security Training for Health Services Research; Alan M. Zaslavsky, Harvard University

Privacy, confidentiality, and data security (PCDS) are of broad concern, especially in health care. Strong PCDS regulations are part of the Health Insurance Portability and Accountability Act of 1996 (HIPAA). The complexity of these regulations cannot be overstated: The final rule on HIPAA's privacy provisions alone comprises more than 1,500 pages of text. Professionals who work with identifiable data are seeking guidance in understanding PCDS issues, recommendations for best practices for PCDS compliance, and tools and methods to increase compliance in their work. This workshop will present overviews of general principles of privacy and confidentiality; requirements of relevant regulations as they affect health services research; technical methods for assessing and limiting unauthorized disclosure, especially for microdata files; and administrative procedures and training that can help to reduce errors affecting PCDS.

SATURDAY, OCTOBER 29

(WK6) 8:30 a.m.–10:15 a.m.—No fee Research Opportunities Using AHRQ Databases; Karen Beauregard, Agency for Healthcare Research and Quality

The Medical Expenditure Panel Survey (MEPS) is a vital national data source designed to continually provide health services researchers, policymakers, health care administrators, businesses, and others with timely comprehensive information about health care use and costs in the United States. The objective of this workshop is to provide data users with an understanding of the unique analytic capabilities

of the MEPS. To meet this objective, participants will be provided with an orientation to the MEPS and MEPS data files. To develop a working knowledge of linking techniques, participants will be walked through exercises and provided with worksheets that illustrate SAS programming techniques using MEPS public use files. The complex survey design aspects of MEPS also will be covered.

(WK7) 8:30 a.m.–10:15 a.m.—Advances in Latent Variable Modeling (Part I); Bengt Muthen, University of California, Los Angeles

This sequence of two workshops gives an overview of recent developments in statistical analysis with latent variables that are of particular relevance to health policy researchers. We will show how the idea of latent variables captures a variety of statistical concepts, including random effects, sources of variation in hierarchical data, frailties, missing data, finite mixtures, latent classes, and clusters. The workshop discusses the integration of such statistical latent variable modeling with the traditional latent variable modeling of psychometrics with its focus on measurement error and hypothetical constructs measured by multiple, fallible indicators as seen in item response theory, factor analysis, and structural equation modeling. The integration leads to a general latent variable framework introduced in the Mplus computer program, facilitating applications such as factor models, growth curve models, multilevel models, latent class models, latent transition models, loglinear modeling, complier-average causal effect estimation in randomized trials, growth mixture modeling with latent trajectory classes, nonignorable missing data models, finite mixture models, discrete-time survival models, and combinations of such models. Various outcome types (such as continuous, censored, count, zero-inflated, semicontinuous, and categorical) of relevance to health research are handled. We will provide numerous examples of health analyses, with Part I emphasizing cross-sectional analyses and Part II emphasizing longitudinal analyses.

(WK8) 10:30 a.m.–12:15 p.m.—(see description above) Advances in Latent Variable Modeling (Part II); Bengt Muthen, University of California, Los Angeles

(WK9) 10:30 a.m.–12:15 p.m.—No fee Research Opportunities Using Data from the CDC National Center for Health Statistics; Jim Lubitz and Robert Weinzimer, CDC National Center for Health Statistics CDC

NCHS is the nation's principal health statistics agency, providing data to identify and address health issues. We will focus on three of the major CDC NCHS data-collection programs: the National Health Interview Survey, the National Health and Nutrition Examination Survey, and the CDC National Health Care Survey. We will describe data collection methods, analytic considerations, data findings, methods of data access, and present examples of research using data from each of these surveys to analyze trends in health and factors affecting health and health outcomes. Target Audience: Health researchers and policymakers who use data for research and health policy decisionmaking.

(WK10) 2:15 p.m.–4:00 p.m.—Issues When Using Hierarchical Models To Estimate Provider Performance; Michael Schwartz and Arlene Ash, Boston University

When, and in what sense, is a hierarchical modeling (HM) framework, in which “shrinkage” estimates of individual provider performance are used, “better for comparing providers” than raw (or traditionally risk-adjusted) mean performance measures? We describe reasons for preferring shrinkage estimators and discuss the difficulty of evaluating the actual results of HM versus other methods in real situations where we do not know the “true” (underlying) means. We also discuss these issues in the context of real data used for profiling, demonstrating that the assumptions that justify HM, especially exchangeability, should not be taken for granted. No previous knowledge of HM is required.

(WK11) 4:15 p.m.–6:00 p.m.—Risk Adjustment and Predictive Modeling; Randall P. Ellis, Boston University

We will provide an overview of the development and use of models that predict person-level spending and health care resource use. Risk adjustment models are used in the United States and internationally for “health-based payment” to health plans, geographic areas, and provider groups where economic incentives may matter greatly. Predictive models that worry less about incentives are increasingly being used to predict person level resource use for many purposes, including case-mix severity controls, identifying patients for case management, provider profiling, and forecasting. We will focus on diagnosis based models, with some comparisons to other predictive frameworks. Attendees will be introduced to a range of uses of such models, the incentive problem, statistical issues, and implementation challenges. Specifically, we will cover how risk adjustment and predictive models differ, comparing alternative risk adjustment model approaches; how the Diagnostic Cost Group (DCG) system works; and issues in implementing risk adjustment and predictive modeling internationally.

SUNDAY, OCTOBER 30

(WK12) 8:30 a.m.–10:15 a.m.—Bayesian Hierarchical Modeling with Applications to Provider Profiling (Part I); David Draper, University of California, Santa Cruz

Datasets with a nested or hierarchical character (e.g., patients within hospitals) abound in health policy research. Often, the units at most or all levels of the hierarchy have either been drawn randomly or we find it useful to think of them as “like” having been drawn randomly, giving rise to a desire to fit random-effects and mixed models. Bayesian fitting of such models can have distinct technical advantages over likelihood-based methods, particularly when the outcome variable is noncontinuous (e.g., binary or count data). In this short course, I will begin with a quick overview of Bayesian inference in general; this will be followed by two detailed case studies in the use of Bayesian hierarchical modeling in provider profiling (one will involve random effects logistic regression in assessing the appropriateness of hospital mortality rates; the other will be based on random effects Poisson regression to examine evidence on whether RN versus non-RN nurse staffing has an effect on patient falls. No previous exposure to Bayesian inference or random effects modeling will be assumed; all of the ideas will be developed in a self-contained fashion. Extensive details on the fitting of the models in WinBUGS and MLwiN will be provided.

(WK13) 10:30 a.m.–12:15 p.m.—(see description above) Bayesian Hierarchical Modeling with Applications to Provider Profiling (Part II); David Draper, University of California, Santa Cruz

OVERALL CONFERENCE AGENDA:

FRIDAY, OCTOBER 28

7:00 a.m.–8:30 a.m.

Breakfast on own

8:30 a.m.–10:15 a.m.

Concurrent Sessions and Workshops

10:15 a.m.–10:30 a.m.

Break

10:30 a.m.–12:15 p.m.

Concurrent Sessions and Workshops

12:30 p.m.–2:00 p.m.

Conference Luncheon (*optional, fee event*)

2:15 p.m.–4:00 p.m.

Plenary Sessions

4:00 p.m.–4:15 p.m.

Break

4:15 p.m.–6:00 p.m.

Concurrent Sessions and Workshops

6:30 p.m.–8:00 p.m.

Offsite Event (*optional, fee event*)

SATURDAY, OCTOBER 29

7:00 a.m.–8:30 a.m.

Continental Breakfast

8:30 a.m.–10:15 a.m.

Concurrent Sessions and Workshops

10:15 a.m.–10:30 a.m.

Break

10:30 a.m.–12:15 p.m.

Concurrent Sessions and Workshops

12:30 p.m.–2:00 p.m.

Lunch on own

2:15 p.m.–4:00 p.m.

Concurrent Sessions and Workshops

4:00 p.m.–4:15 p.m.

Break

4:15 p.m.–6:00 p.m.

Concurrent Sessions and Workshops

SUNDAY, OCTOBER 30

7:00 a.m.–8:30 a.m.

Continental Breakfast

8:30 a.m.–10:15 a.m.

Concurrent Sessions and Workshops

10:15 a.m.–10:30 a.m.

Break

10:30 a.m.–12:15 p.m.

Concurrent Sessions and Workshops

Post Conference Proceedings Publication Instructions

Copyright notice: Authors retain copyright on individual papers published in the Proceedings. Eligible presenters are not prohibited from publishing in the Proceedings as well as other publications; authors are advised to consult journal exclusivity agreements for restrictions.

You are authorized to publish if you presented at the:

2006 Joint Statistical Meetings (JSM), August 6–10, 2006

International Conference on Health Policy Research (ICHPR), Oct. 28–30, 2005 (sponsor: Section on Health Policy Statistics)

If your paper was withdrawn or not orally presented, it is not eligible.

Paper Format and Page Limits - The Proceedings will be published in electronic format only. Papers must be double-column, single-spaced, and fully justified (figures may span both columns, if necessary). 10-point Times New Roman is the preferred font. Titles must appear at the top of the first page; span both columns; and be centered, bold, and in 12-point type. Author name and affiliation should also span both columns; be centered directly below the title; and appear in regular, 10-point type. Color graphics are acceptable.

Paper/Poster page limits:

Invited papers: 12 pages

Contributed papers: 8 pages

Discussant/panelist papers, posters: 3 pages

Do not put page numbers on your pages; these will be generated automatically.

Instructions for Electronic Submission

Your paper will be converted to PDF; you MUST proofread it on your computer screen and approve it to finalize your submission.

You will need:

Java Script-enabled

Cookies-enabled

Browser 4.06 or newer (either Microsoft Internet Explorer or Netscape Navigator)

10 to 30 minutes to upload your paper and finalize your submission

A password unrelated to your personal passwords used for banking, etc. (passwords are not secure)

Mailing address

Email address

Paper presentation type (invited, contributed, etc.)

Sponsoring Section

Credit card number for secure online payment of page charges (You are requested to pay a page charge of \$15 per page to help cover the cost of production. In return, you will receive an automatic copy of the JSM 2005 Proceedings CD-ROM when it is released.)

Your paper saved with one of the following file extensions:

.DAR DVI to PDF compiler (archive file)

.DGZ DVI to PDF compiler (archive file)

.DOC Microsoft Word

.DVI DVI to PDF compiler

.DVZ DVI to PDF compiler (archive file)

.EPS Adobe Acrobat Distiller

.GZ TeX to PDF compiler (archive file)

.HTM Microsoft Word

.HTML Microsoft Word

.PDF Adobe Acrobat

.PRN Adobe Acrobat Distiller

.PS Adobe Acrobat Distiller

.RTF Supported by many word processors

.TAR TeX to PDF compiler (archive file)

.TAZ TeX to PDF compiler (archive file)

.TEX TeX to PDF compiler

.TGZ TeX to PDF compiler (archive file)

.TXT Supported by many word processors and text editors

.WK1 Microsoft Excel

.WK3 Microsoft Excel

.WK4 Microsoft Excel

.WPD WordPerfect

.XLS Microsoft Excel

Please note that TeX documents, especially those including large or complex graphics (equations), may require more processing time. If a message is displayed indicating your paper has not been completely converted, exit the site and check back in about an hour. You may resume the submission process where you left off.

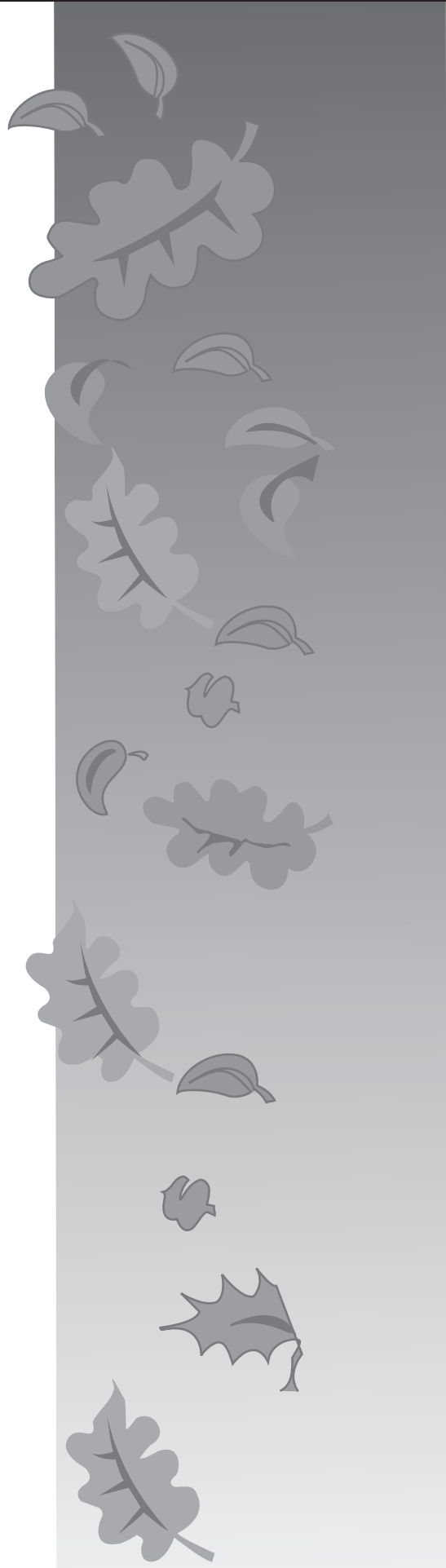
Referencing your Proceedings Paper

Following is a sample of the standard format for referencing your paper (page numbers are assigned by the vendor; users may search by author, title, or Section):

Freund, R. J. (2005), "An Example of Prediction with Regression: a Comparison of Methods," 2006 Proceedings of the American Statistical Association, Statistical Computing Section [CD-ROM], Alexandria, VA: American Statistical Association: pp-pp.

Deadline - The web site will close to new submissions on Friday, October 20, 2006, at 11:59 p.m. EDT. Submit early. You can expect delays due to heavy volume around the deadline.

The submission site will remain open for revisions until Tuesday, November 14, 2006, at 11:59 p.m. EST.



ABSTRACTS

C-1 (Contributed): Propensity Score Methods and Analysis of Noncompliance

Reliably Assessing Physician Performance: The Critical First Step in Pay-for-Performance

Sherrie Kaplan, UCI School of Medicine*

Pay-for-performance programs require accurate and reliable physician-level performance assessment. Using diabetes care as a model, I create a fair and reliable physician-level quality performance score. I abstracted 11 diabetes quality measures from the medical records of a sample of 210 physicians and their 7,574 patients in the Diabetes Provider Recognition Program (DPRP). I tested each measure for the magnitude of the physician effect (or thumbprint), using generalized estimation equations. The inflation factor (the square of the ratio of the estimated model coefficient standard errors from models with and without adjustment for clustering) measured physician effect. I combined five process and four outcome measures into an aggregate, case-mix adjusted, physician-level diabetes quality score with high reliability (Cronbach's $\alpha = .80$). Case-mix variables accounted for only 3% of the variation in the physician-level score. The aggregate case-mix-adjusted score could discriminate physicians scoring in the highest from those scoring in the lowest quartiles of the quality-of-care distribution. Thus, with careful attention to measurement principles in aggregating well-tested quality measures that maximize the physician effect, I could develop a fair and reliable physician-level quality-of-care score for supporting programs to improve physician performance.

Not Doing It Until I Do? Determining the Efficacy of Virginitiy Pledges

Janet Rosenbaum, Harvard University*

This study explores the efficacy of virginitiy pledges in delaying sexual debut for sexually inexperienced adolescents in the National Longitudinal Study of Adolescent Health (Add Health). Subjects were virgin respondents without wave 1 pledge who reported their attitudes toward sexuality and birth control at wave 1 ($n=3,443$). Nearest-neighbor matching within propensity score calipers was used to match wave 2 virginitiy pledgers ($n=291$) with non-pledgers based on wave 1 attitudes, demographics, and religiosity. Treatment effects due to treatment assignment and compliance were calculated. Seventeen percent of virginitiy pledgers were found to be compliant with their pledge and not recant at wave 3 their earlier report of having taken a pledge. Similar proportions of virginitiy pledgers and non-pledgers reported having had premarital sex (54% and 61%, $p=0.16$) and tested positive for chlamydia (2.7% and 2.9%, $p=0.89$). Five years after taking a virginitiy pledge, most virginitiy pledgers failed to report having pledged. Virginitiy pledges do not affect the incidence of self-reported premarital sex or assay-determined chlamydia.

Propensity Score Methodology Combined with Modified ANCOVA: An Example

Bradley Huitema, Western Michigan University; Joseph McKean, Western Michigan University*

Propensity score methods were applied to a large sample program evaluation of the effects of a program designed to improve the performance of at-risk students in a university setting. Data on approximately 30,000 students were available as a reservoir from which matches for treated subjects were selected using propensity scores. The final analysis was based on a modified analysis of

covariance procedure. A comparison of the results of the final analysis with results based on more conventional analyses is presented. It was concluded that insufficient evidence was available to conclude that the treatment program improved performance.

Too Much Ado about Propensity Score Matching?

Onur Baser, Thomson-Medstat*

We evaluate several matching techniques and provide a suggested guideline for selecting the best technique. The following approach is proposed to check for balance: 1) two sample t -statistic between the mean of the treatment group for each explanatory variable with the mean of these variables in the control group; 2) the mean difference as a percentage of the average standard deviations; 3) percent reduction bias in means of explanatory variables after matching and initially; 4) compare treatment and control density estimates for the explanatory variables; 5) the propensity scores of control units with that of the treated units. MarketScan data were used to provide empirical examples. I examined 2 to 1 matching, nearest neighborhood matching with replacement, MM matching, MM with calipers, stratification method, kernel matching, and radius matching. MM with calipers where calipers is selected as a quarter of standard deviation of estimated propensity score provided the best results and was thus the optimal approach. The suggested joint consideration offers an approach to assess the robustness of the estimates.

Noncompliance Bias Correction Based on Covariates in Randomized Experiments

Yves Atchade, University of Ottawa; Leonard Wantchekon, New York University*

We propose a new method for consistent estimation of causal effects in randomized experiments when compliance to assignments is only partial. I follow the potential outcome approach, but in contrast to Imbens and Rubin (1997), I require no prior classification of the compliance behavior. When noncompliance is not ignorable, I show that adjusting for arbitrary covariates actually can increase the estimation bias and I propose a new method for selecting relevant covariates to reduce noncompliance bias. The method uses prior information about the experiment. Next, I investigate cases when the overlap assumption does not hold and, on the basis of covariates, some units are excluded from the experiment or, equivalently, never comply with their assignments. In that context, I show consistent estimates of the causal effect of the treatment are possible if there exists a global representation of the conditional expectation of the outcome given the covariates. I illustrate the methodology with several examples, such as the access to influenza vaccine experiment and the PROGRESA experiment.

Principal Stratification Approach to Dealing with Treatment Noncompliance and Subsequent Nonresponse

Robert Ware, The University of Queensland*

In a trial comparing experimental and standard treatments, the treatment a participant receives may not be the treatment he or she was assigned (i.e., there is imperfect compliance). The standard way of analyzing such trials is by the "intention-to-treat" principle, which can yield biased estimators for causal effects of treatments.

A treatment effect of interest in the presence of noncompliance is the complier-average causal effect, which is the treatment effect for participants who would comply with treatment protocol regardless of treatment assigned. I present a model, based on the principal stratification framework, that estimates the CACE in the presence of both noncompliance and missing data. When estimating model parameters, I assume each participant has some latent compliance state that describes their behavior under all possible treatment assignments. I illustrate methods using data from a trial investigating the effectiveness of daily sunscreen application in reducing the incidence of basal-cell carcinoma in an adult population.

C-2 (Contributed): Imputation Methods

The Impact of Expanded Medicaid Eligibility for Pregnant Women on Health Care Utilization and Outcome

*John Engberg**, RAND Corporation; *Donna Farley*, RAND Corporation; *Lisa Shugarman*, RAND Corporation

Pregnant women in Arkansas between 134% and 200% of the Federal Poverty Line are newly eligible for Medicaid coverage. I compare trends in prenatal care utilization and birth weights between periods before and after expanded Medicaid eligibility. Separate trends are estimated for always eligible, newly eligible, and never eligible segments of the population. Both prenatal care and birth weight information are from birth certificates. Eligibility status is imputed for all pregnant women based on age, race, and marital status using probabilities estimated from the 2000 Decennial Census Public Use Micro Sample (PUMS) of Arkansas women with children under 6 years old. Information about county-level poverty rates from Census Summary Files also is incorporated. I find that the expanded Medicaid program led to increased access to prenatal care for newly eligible women but did not reduce the probability of giving birth to a baby with a low birth weight.

Robustness of a Multivariate Normal Approximation for Imputation of Incomplete Binary Data

*Tom Belin**, University of California, Los Angeles; *Coen Bernaards*, Genentech, Inc.; *Joseph Schafer*, Pennsylvania State University

Multiple imputation has become easier to perform with the advent of several software packages that provide imputation under a multivariate normal model, but imputation of missing binary data remains an important practical problem. Here, I explore three alternative methods for converting a multivariate normal imputed value into a binary imputed value: simple rounding of the imputed value to the nearer of 0 or 1, a Bernoulli draw based on a “coin flip” where an imputed value between 0 and 1 is treated as the probability of drawing a 1, and an adaptive rounding scheme where the cutoff value for determining whether to round to 0 or 1 is based on a normal approximation to the binomial distribution, making use of the marginal proportions of 0s and 1s on the variable. I perform simulation studies on a dataset of 206,802 respondents to the California Healthy Kids Survey, where the fully observed data on 198,262 individuals

defines the target population and the incomplete cases are used to impose realistic patterns of missing data. For a range of estimands, I found satisfactory bias and coverage properties for all the procedures, suggesting that approaches such as these that are based on statistical approximations are far preferable to either avoiding the use of certain variables or relying on complete-case analysis. However, a flaw with the coin-flipping approach appeared in inferences for odds ratios, as independent coin-flipping on the respective binary variables fails to preserve associations between variables. Considering both the occurrence and the extent of deficits in coverage, I found that adaptive rounding provided the best performance overall.

Comparison of Hot Deck and Multiple Imputation Methods Using HCSDB Data

*Donsig Jang**, Mathematica Policy Research, Inc.; *Amang Sukasih*, Mathematica Policy Research, Inc.

This paper presents the results from an exploratory study on imputation of missing data for the Health Care Survey of DoD Beneficiaries (HCSDB). Using data from the 2003 HCSDB, I investigated the extent of item nonresponse in several variables of interest. I also studied and implemented two methods of data imputation for these variables: Sequential Hot-deck—a method of single imputation—and Sequential Regression Multivariate Imputation (SRMI)—a method of multiple imputation. I made comparisons between the statistics prior to and after imputation and between statistics computed based on the hot-deck imputation and the multiple imputation. The domain for comparison was catchment areas, where quarterly estimates of scores or proportions and rating were produced for each of these catchment areas. In this study, I focused on only catchment areas with a sample size of 20 or larger, as estimates with a sample size of less than 20 are considered unstable.

Semiparametric Approach for Multiple Imputations of Unobserved Values in Longitudinal Studies

*Yulei He**, Harvard Medical School; *Trivellore Raghunathan*, University of Michigan

Unbalanced data, where not all individuals are observed at the same time points, is a common feature in many longitudinal studies. A case study motivating this research involves studying the impact of wealth of parents during the critical developmental age on their children’s health development. Missing data occur in wealth measure and family income. The multiple imputation approach provides a framework for handling missing data in such instances. This paper discusses a Bayesian semiparametric approach using spline models to create imputations. The spline models allow for modeling for the nonlinear trends of the longitudinal data. Gibbs sampling is used to obtain the draws from the posterior predictive distribution of the missing observations conditional on the observed values. I illustrate the proposed method by applying it to the dataset of the Panel Study of Income Dynamics. Results from simulation studies evaluating the repeated sampling properties of the inferences obtained using this approach are presented. The properties of the multiple imputation inference under uncongeniality also are discussed.

Multiple Imputation for Correcting Verification Bias in Estimating Sensitivity and Specificity

Ofer Harel*, *University of Connecticut*; Andrew Zhou, *University of Washington*

Sensitivity and specificity are used widely to describe a diagnostic test. When all subjects have both test results and true status, the estimation of the sensitivity and specificity is built on two binomial distributions. This estimation is not a trivial task. In the case in which all subjects are screened using a common test and a subset of these subjects are tested using a golden standard test, there is a risk for verification bias. When not all subjects have been verified, special methods of estimation need to be used. There are several methods to estimate the sensitivity, specificity, and standard error in this kind of situation. The standard methods [were] developed under special cases of the verification choices. Approaching this problem from a missing data perspective allows us to use the Multiple Imputation (MI) technique to impute the data. We adopt MI framework and develop MI procedures using the most common “complete data” methods. I compare the procedures against themselves and the standard (incomplete) methods. I illustrate our procedure using a biomedical data example.

I-1 (Invited): Combining Estimates/Information Using Multiple Data Sources

Obtaining Cancer Risk Factor Prevalence Estimates in Small Areas

Michael Elliott*, *University of Pennsylvania*

Cancer surveillance research requires accurate estimates of risk factors at the small-area level. Unfortunately, no one population-based survey provides ideal prevalence estimates of such risk factors. One strategy is to combine information from multiple surveys, using the complementary strengths of one survey to compensate for the weakness of the other. The National Health Interview Survey (NHIS) is a nationally representative, face-to-face survey with a high response rate; however, it cannot produce state or substate estimates of risk factor prevalence because of small sample sizes and/or concerns about privacy. The Behavioral Risk Factors Surveillance System (BRFSS) is a state-level telephone survey that excludes non-telephone households and has a lower response rate, but does provide reasonable sample sizes in all states and many counties. We develop a model-based approach that considers the joint distributions of the risk factor of interest, associated covariates, and the small-area identifier in a missing data framework. The sample designs are incorporated into the model to allow robust estimates of the risk factor prevalence.

New Modeling Strategies for Combining Data from Multiple Surveys To Obtain Small-domain Estimates of Obesity

Dawei Xie*, *University of Pennsylvania*

Cancer surveillance research requires accurate estimates of risk factors, such as prevalence of obesity for small domains. The small domain is defined by the cross classification of year, region, age categories, gender, race, education, and occupation. Unfortunately, no one population-based survey provides ideal prevalence estimates of such risk factors. One strategy is

to combine information from multiple surveys, using the complementary strengths of one survey to compensate for the weakness of the other. The National Health Interview Survey (NHIS) is a nationally representative, face-to-face survey with a high response rate; however, it cannot produce reliable small-domain estimates of risk factor prevalence because of small sample sizes. The Behavioral Risk Factors Surveillance System (BRFSS) is a state-level telephone survey that excludes non-telephone households and has a lower response rate, but does provide reasonable sample sizes. In our previous work, we obtained estimates on small-area (U.S. county) level via a multivariate linear mixed model after transforming the mean prevalence in each small area by an arcsin square root transformation. However, this transformation might not perform well when sample size is small. Therefore, we are going to adopt a logit-normal model in this new research. Results will be compared between the two model strategies. The sample designs are incorporated into the model.

Maximum Likelihood Estimation of Multiple Source Predictor Regression Models

Nicholas J. Horton*, *Smith College*; G. M. Fitzmaurice, *Harvard University*; T. L. Lash, *Boston University*; N. M. Laird, *Harvard University*

Multiple-source data commonly arise in health services research when parallel reports are solicited from informants and medical databases. The medical comorbidity status of a patient, for example, might be ascertained by interviewing the patient, her doctor(s), and by reviewing her medical records. We review novel regression models for analyzing multiple source risk factors as special cases of generalized linear models, albeit with correlated outcomes. These models allow testing for source differences in the relationships between risk factors and outcome, estimating source-specific effects when necessary (testing if the effects of other risk factors on the outcome differ by source) and incorporating subjects with incomplete observations. We consider quasi-likelihood methods and develop maximum likelihood regression methods for analyzing incomplete multiple source predictor data. These methods are applied to an analysis of use of tamoxifen among a cohort of breast cancer survivors in which multiple source reports of comorbidity were collected.

C-3 (Contributed): Insights for Health Studies from Surveys and Psychometric Research

Survey Conditioning in Self-reported Mental Health Service Use: Results from a Randomized Trial

Naihua Duan*, *University of California, Los Angeles*

To test the effect of survey conditioning (whether observed survey responses are affected by previous experience in the same survey or similar surveys) in a survey instrument used to assess psychiatric disorders and mental health service use, participants in the National Latino and Asian American Study are randomly assigned to an interleaved instrument that places service use questions after detailed questions on disorders or an ensemble instrument that screens for service use near the beginning of the survey, which is hypothesized to be less susceptible to survey

conditioning. In-person, computer-assisted interviews are conducted in respondent's preferred language. Self-reported mental health service use measures are compared between recipients of the two instruments. Survey conditioning is found, as higher service use rates are reported with the ensemble instrument than with the interleaved instrument for all service use measures; ORs range from 1.41 to 3.11, all p-values < 0.001. Results are similar across ethnic groups and insensitive to model specification. An ensemble instrument is therefore recommended when it is feasible for measures susceptible to survey conditioning.

Self-rated Health among Foreign- and Native-born Individuals: A Test of Comparability

Elena Erosheva, University of Washington; Daniel Takeuchi, University of Washington; Emily Walton, University of Washington*

Self-rated health is an indicator of general health and a robust predictor of morbidity, mortality, subsequent disability, and health care utilization. The five-category health status scale is used in a wide range of surveys across many countries. Some research suggests self-rated health response categories may be biased for certain social groups. Using data on Asian Americans from the National Latino and Asian American Study, I test whether immigrants are less likely to report the extreme ends of the scale than their native-born counterparts. Because individuals may differ on a number of dimensions, I use propensity score matching to derive groups that share similar demographic and health characteristics. Each native-born person is matched to a foreign-born of the same ethnicity by nearest available Mahalanobis metric within a caliper defined by the propensity score. Propensity score framework allows us to make descriptive comparisons of self-rated health responses by nativity status, controlling for background characteristics. Results indicate that for Asian Americans, nativity is not associated with higher likelihood of reporting the extreme ends of the health status scale.

An Experiment To Explain What Influences Clinical Decisions

Carol Link, New England Research Institutes; Lisa Marceau, New England Research Institutes; John McKinely, New England Research Institutes; Amy O'Donnell, New England Research Institutes*

Numerous studies have looked at disparities in clinical decisionmaking. However, these studies often are beset with the problems of causality, the confounding of many factors, and generalizability. This presentation describes a factorial experiment that addresses these problems. Two medical conditions—coronary heart disease (CHD) and depression—were portrayed on videotape by professional actors. The “patient’s” gender, age, race, and social class were systematically varied. The physician subjects were UK- or U.S.-trained primary care providers. The physician subjects were randomly sampled from eight health authorities in the UK and Massachusetts. They were asked to manage the “patient” in the context of their current practice. Experiments solve the problem of causality by manipulating experimental factors. A balanced factorial experiment solves

the problem of confounding by allowing the estimation of unconfounded effects. The random selection of physicians allows for the generalization of the results to the health authorities and sampled state and to the countries as a whole if the sampled areas are representative.

Latent Class Structure of IQ in Preschool Children with Autism

Elizabeth Koehler, University of Washington; Robert Abbott; Ted Beauchaine; Geraldine Dawson; Catherine Lord; Jeffery Munson; Sally Rogers; Marian Sigman; Andrew Zhou, University of Washington*

Currently, autism is viewed as a spectrum disorder, meaning that strikingly different severity levels are being classified as the same disorder. However, past studies and clinical experience frequently hint at more than one type. The intent of this study is to quantitatively investigate the possibility of more than one distinct IQ class existing in autism. This study uses a combination of latent class analysis and taxometric methods to examine the Mullen IQ scores of 347 children diagnosed with autism. I would like to determine whether it is reasonable to think the IQ of autistic children is being comprised of more than one distinct class and have some statistical certainty about the existence and composition of the classes. Commonly, latent class analysis is used for both, but I believe that by including the taxometric approach, I can safeguard against the weaknesses in the latent class method. Preliminary results support the existence of at least two classes. While this is only a first step in exploring a multiple class model, it should be considered strongly amongst the growing body of evidence for such a model.

Evaluating the Sample Invariance Property of the Standard Error of Measurement

Joseph Cappelleri, Pfizer Inc.; Andrew Bashmakin, Pfizer Inc.; Samiran Ghosh, University of Connecticut; William Lender King, Pfizer Inc.*

Under classical test theory (CTT), it often is claimed that the standard error of measurement (SEM) of a measurement scale is sample invariant [SEM = standard deviation* $\sqrt{1 - \text{scale reliability}}$]. To evaluate this claim, I conducted Monte Carlo simulations on real and simulated data. The simulations involved random samples of varying size in which the standard deviation and Cronbach's alpha (α , a measure of scale reliability) were computed for 300,000 replications. The invariance property was tested based on a theoretical regression model [$SD^2 = SEM^2 * 1 / (1 - \alpha)$] in which the regression coefficient SEM2 and the coefficient of determination (r^2) were estimated. For the real data on the 10-item Schwartz Outcome Scale administered to 145 patients with obsessive-compulsive disorder, the estimated SEM2 was 12.19 (close to the true value of 12.41) and the estimated r^2 of 0.99 was virtually perfect (and close to the theoretical value of 1). Results were confirmed with simulated data. Thus, the claim that the SEM of a measurement scale is sample invariant is supported.

Plenary Session: Assessing Pharmaceutical Safety and Efficacy in the Wake of COX-2 and HRT

Muhammad Mamdani*, Institute for Clinical Evaluative Sciences; Alan Breier*, Eli Lilly and Company; Robert O'Neill*, U.S. Food and Drug Administration; Frank Harrell*, Vanderbilt University

This plenary session will address recent controversies about benefits and risks of drug therapies that have important international health policy implications. The talks will address a range of key issues regarding improved regulation of drugs, with a focus on Cox-2 inhibitors and hormone replacement therapy. The panelists will discuss improved study designs, such as head-to-head drug comparisons, the role of observational studies in informing drug safety and efficacy, the role of post-marketing surveillance in longterm evaluations, and the opportunities for health policymakers and the public to influence the dynamic interactions between science and policy. Speakers are intimately involved in reviewing and researching drug safety issues from a variety of perspectives. The session will be lively, informative, and possibly controversial.

I-2 (Invited): Methods of Risk Adjustment for Skewed Outcome Data

Addressing Skewness and Kurtosis in Risk Adjustment

Alberto Holly*, University of Lausanne; Yevhen Pentsak, University of Lausanne

In this presentation, we consider a linear model of the form $\mathbf{y} = \mathbf{x}\beta^0 + \mathbf{u}^0$, and our primary interest is to estimate the vector of regression coefficients β^0 without transforming the data (for example, by taking the logarithm of \mathbf{y}). To this end, we assume the conditional distribution of \mathbf{y} given \mathbf{x} belongs to four-parameter families of distribution, these parameters being related to the first four conditional moments of \mathbf{y} given \mathbf{x} . Although OLS procedures yield consistent estimator of β^0 , it should be less efficient than the Maximum Likelihood Estimator (MLE) of β^0 , which explicitly takes into account the additional information on the third and fourth moments. Similarly, the distribution of the OLS t -ratio should be affected by the departure from normality of the distribution of \mathbf{u}^0 conditional on \mathbf{x} , as the variance of the asymptotic distribution of the variance of OLS estimator also depends on these moments. In order to evaluate the order of magnitude of these expected effects, we consider in detail two particular cases: the Pearson's type IV and the generalized gamma distribution.

We also address the following research questions:

- What are the properties of the OLS estimator of the variance?
- What are the properties of the OLS t -ratio?
- What are the properties of the optimal predictor when the true distribution is either a Pearson's type IV or a generalized gamma distribution?

We apply these ideas to merged hospital records and insurance data and a health-based risk adjustment model in Switzerland.

Using Diagnosis-based Risk Adjustment and Self-reported Health Status To Predict Mortality

Kenneth Pietz*, VA Medical Center and Baylor College of Medicine; Laura A. Petersen, VA Medical Center and Baylor College of Medicine

Both diagnosis-based risk adjustment variables and self-reported health status have been found to predict mortality. This presentation compares the ability of two diagnosis-based risk adjustment systems and health self-report obtained at approximately the same time to predict long-term mortality and attempts to determine whether health self-report data contains health information not contained in diagnosis-based risk adjustment systems. We tested the ability of Diagnostic Cost Groups (DCGs), Adjusted Clinical Groups (ACGs), and SF-36V (SF-36 for veterans) Physical Component Score (PCS) and Mental Component Score (MCS) to predict one-year and five-year mortality. The additional predictive value of adding PCS and MCS to ACGs and DCGs also was evaluated. Logistic regression models were compared using Akaike's information criterion and the c -statistic. The outcome was all-cause mortality. The diagnosis-based risk adjustment variables showed slightly better performance than the health self-report variables in predicting mortality. Health self-reports may add health risk information in addition to age, gender, and diagnosis for predicting mortality.

Risk Adjustment with Flexible Link and Variance Function Models

Anirban Basu*, University of Chicago; Bhakti Arondekar, GlaxoSmithKline; Paul Rathouz, University of Chicago

Traditional models, such as ordinary least squares (OLS) regression, and transformation models, such log-OLS regression, have been shown to be problematic. Researchers have suggested the use of generalized linear models (GLM) to overcome these problems. However, specifications of a link function and/or variance function in GLM seldom are driven by theory and often are difficult to ascertain using available diagnostic tests. Recently, we proposed an extension to the estimating equations in generalized linear models to estimate parameters in the link function and variance structure simultaneously with regression coefficients. Rather than focusing on the regression coefficients, the purpose of these models is consistent estimation of the mean of the outcome as a function of a set of covariates, and various functionals of the mean function. Here, we illustrate the biases that may arise in using alternative estimators to model expenditure data. We conclude that careful selection of the estimator is important for modeling cost data. The EEE estimator seems to perform better than alternative estimators studied.

C-4 (Contributed): Competing Risks and Advanced Survival Modeling in Health Services Research

The Influence of Spousal Morbidity and Mortality on Proband Mental Health Diagnosis and Treatment

*Kara Bambauer**, Harvard Medical School; *Nicholas Christakis*, Harvard Medical School

Spouses have a higher risk of mental illness during care giving and widowhood, yet this phenomenon is poorly understood. This presentation will evaluate whether probands (i.e., the people being studied) with ill spouses have a higher risk of developing mental health or substance abuse (MHSA) disorders than probands who have healthy spouses. Used are Medicare claims from 1993–2001 for 949,408 married couples and Cox models to determine the effect of spouse hospitalization and death on proband MHSA diagnosis, controlling for all demographic and clinical characteristics. Husbands whose wives become hospitalized are 1.8 times more likely to have an MHSA diagnosis than husbands with healthy wives. Wife death leads husbands to be 0.9 times less likely to have an MHSA diagnosis. Wives whose husbands become hospitalized are 2.1 times more likely to have an MHSA diagnosis than wives whose husbands are healthy. Husband death leads wives to be 1.1 more likely to have an MHSA diagnosis. Extensions to fixed effects Cox models will be discussed.

Differences in Regression-based Decomposition Depending on Functional Form

*Jinn-Ing Liou**, University of Wisconsin-Madison; *Michael Finch*, Finch & King, Inc.; *Jennifer Frylak*, Ingenix; *Nilay Shih*, University of Wisconsin-Madison; *Maureen Smith*, University of Wisconsin-Madison

Regression-based decomposition has played a significant role in explaining sources of differences between two groups (e.g., blacks versus whites). This technique has been expanded to nonlinear functional forms, but empirical differences between these forms have not been examined. Here, black-white differences in the probability (logistic) and cumulative hazards (Cox) of rehospitalization for the year after stroke are decomposed. Administrative data are obtained in 11 regions for 39,068 Medicare beneficiaries aged 65 years and older who were discharged during 1998–2000 with a diagnosis of acute ischemic stroke. To account for non-proportional hazards, the main outcome measures are rehospitalization within the first 30 days after discharge and within the subsequent 11 months for 30-days survivors. Standard errors are bootstrapped using 1,000 replications. Although logistic and Cox models give similar results for the overall average predicted probability and cumulative hazards of rehospitalization, the results of the decomposition differ significantly. Similar conclusions for overall differences between groups do not imply similar conclusions for decomposition results.

I-3 (Invited): Advanced Methods for Estimating Health Disparities

Racial Disparities in Self-rated Health at Older Ages: The Contribution of Neighborhood-level Factors

*Kathleen Cagney**, University of Chicago; *Christopher Browning*, The Ohio State University; *Ming Wen*, The Ohio State University

Racial differences in self-rated health at older ages are well-documented. African Americans consistently report poorer health, even when education, income, and other health status indicators are controlled. The extent to which neighborhood-level characteristics mediate this association remains largely unexplored. We ask whether neighborhood social and economic resources help to explain the self-reported health differential between African Americans and whites. Using the 1990 Decennial Census, the 1994–95 Project on Human Development in Chicago Neighborhoods Community Survey, and selected years of the 1991–2000 Metropolitan Chicago Information Center Metro Survey, we examine the impact of neighborhood structure and social organization on self-rated health for a sample of Chicago residents 55+ (N=636). We use multilevel modeling techniques to examine both individual and neighborhood-level covariates. Findings indicate that affluence, a neighborhood structural resource, contributes positively to self-rated health and attenuates the association between race and self-rated health. When the level of affluence in a community is low, residential stability is negatively related to health. Collective efficacy, a measure of neighborhood social resources, is not associated with health for this population.

Valuation of Arthritis Health States across Ethnic Groups and between Patients and Community Members

*Julianne Soucek**, Baylor College of Medicine; *Margaret Byrne*, University of Pittsburgh; *Adam Kelly*, Baylor College of Medicine; *Marsha Richardson*, Baylor College of Medicine; *Chong Pak*, Baylor College of Medicine; *Harlan Nelson*, Baylor College of Medicine; *Maria Suarez-Almazor*, Baylor College of Medicine; *Michael E. DeBaakey*, VA Medical Center

We surveyed 193 community members identified by random digit dialing: 64 white (W), 65 African American (AA), and 64 Hispanic (H). The patient sample included 198 individuals diagnosed with osteoarthritis (OA) and drawn sequentially from a health-provider institution clinic list, 66 per ethnic group. Participants were interviewed face to face and asked to rate two scenarios describing patients with arthritis (mild and severe) using visual analog scale (VAS), standard gamble (SG), and time trade-off (TTO). Differences were adjusted for cohort, age, age-squared, gender, and education. The difference between the utility scores for mild OA and severe OA was significantly smaller for AA than W by the VAS, TTO, and SG methods. The difference between mild and severe states was smaller for H than W by the SG method. For the severe OA state, the odds that AA had scores > 0.80 relative to W was 2.22 using the TTO method. Preferences for the mild OA state were not different among ethnic groups. Using the SG method, the odds that the scores were > 0.80 in the public cohort vs. the patient cohort were greater than 1 for severe OA and for mild OA. The public gave the severe OA state a higher preference score than patients did using the VAS method. Education and age had significant, independent effects on utility scores. Age increased the SG utility scores, and the difference between severe and mild health states was less by VAS for older individuals. Education ameliorated the effects of other variables on TTO and SG scores.

Implementing the IOM Definition of Disparities: An Application to Mental Health Care

Benjamin L. Cook*, Harvard University; Thomas G. McGuire, Harvard University; Margarita Alegria, Harvard University; Kenneth B. Wells, Harvard University; Alan Zaslavsky, Harvard University

In a recent report, the Institute of Medicine (IOM) defined a health service disparity between population groups to be the difference in treatment or access not justified by the differences in health status or preferences of the groups. This paper proposes an implementation of this definition and applies it to disparities in outpatient mental health care. The Health Care for Communities (HCC) survey re-interviewed 9,585 respondents from the Community Tracking Study in 1997–98, over sampling individuals with psychological distress, alcohol abuse, drug abuse, or mental health treatment. The HCC is designed to make national estimates. We modeled expenditures using a Generalized Linear Model (GLM) with quasilielihoods and a probit model. We adjusted for group differences in health status by transforming the entire distribution of health status for minority populations to approximate the white distribution. We compared disparities according to the IOM definition to other methods commonly used to assess health services disparities. Our method, based on the IOM definition, finds significant service disparities between whites and both blacks and Latinos. Estimated disparities from this method exceed those for competing approaches, due to the inclusion of effects of mediating factors (such as income) in the IOM approach. A rigorous definition of disparities is needed to monitor progress against disparities and to compare their magnitude across studies. With such a definition, disparities can be estimated by adjusting for group differences in models for expenditures and access to mental health services.

C-5 (Contributed): Cost, Risk, and Allocation of Health Care Resources

Causes and Consequences of Regional Variations in Health Care Resources in Ontario

Thérèse A. Stukel*, Institute for Clinical Evaluative Sciences
The health and policy implications of regional variations in health care resources across regions raise issues of equity and efficiency. I estimate health care resources for 14 Ontario health service regions and examine relationships among health care resource supply, utilization, and outcome. Ontario has fewer health care resources per capita than the United States. Variations in regional resources are lower for hospital beds and cardiac catheterization labs but similar for specialist supply and diagnostic test equipment. Patients admitted with AMI, hip fracture, or GI bleed residing in areas with more resources were similarly ill but received more health care services, such as readmissions, physician visits, and invasive procedures. No differences in long-term mortality were found across regions. Ontario health care resources that are centrally managed, such as hospital beds, demonstrate lower variations in supply across regions. Increased intensity of health care does not appear to reduce mortality for similarly ill patients. This emphasizes the importance of managing health system capacity and of matching health system resources to need for long-term health system sustainability.

Adjusting SARS-affected Data for Canadian Inpatient Case Mix Indicators

Sheril Perry*, Canadian Institute for Health Information;
Qian Yang, Canadian Institute for Health Information;
Douglas Yeo, Canadian Institute for Health Information

The 2003 Severe Acute Respiratory Syndrome (SARS) outbreak in the greater Toronto area of Canada resulted in a WHO travel advisory and a suspension of all non-essential hospital services within Ontario by the Ontario Ministry of Health. These actions led to a decrease in hospital admissions and modifications to discharge practices. Inpatient length of stay (LOS) and cost values in many hospitals were affected greatly. The Canadian Institute for Health Information (CIHI) uses hospital inpatient discharge and case-cost data to calculate two annual health resource indicators associated with its inpatient grouping methodology: Expected Length of Stay (ELOS) and Resource Intensity Weights (RIW). For this purpose, the effect of SARS on the inpatient data was assessed and mitigated. The 2003–2004 fiscal year data were partitioned by SARS phases and compared to previous-year levels. Both LOS and cost were modeled using regression techniques to identify hospitals and SARS phases significantly affected. As a result, hospital-specific data identified as being statistically different during the SARS phases were removed from the national database for ELOS and RIW calculation.

Why Are We Still Using Charlson To Measure Comorbidity?

Jeanne Speckman*, Boston University Medical Center;
Arlene Ash, Boston University School of Medicine; Jennifer Fonda, Boston University Medical Center; Amresh Hanchate, Boston University School of Medicine; Nancy McCall, Research Triangle Institute; Thomas Williams, TRICARE Management Activity, HPA&E

Health expenditures are strongly influenced by overall illness burden, thus understanding factors that influence health care utilization requires risk adjustment. “Charlson” is an older and common risk adjustment method. I compare it to four others: Adjusted Clinical Groups, Chronic Disease and Disability Payment System, Clinical Risk Groups, and Diagnostic Cost Groups to predict total cost. 2.3 million continuously enrolled TRICARE Prime beneficiaries older than 65 years of age in FY2001–2002 were split into estimation and validation subsamples of 1.8 and 0.5 million. Each model used FY2001 diagnoses to predict FY2002 expenditures. I assessed concordance between predicted and actual expenditures. While there was some differential in model performance, the newer methods all did well and vastly outperformed Charlson on overall and subgroup-specific measures of predictive accuracy. This was a time of substantial cost shifts in this population, nevertheless the four newer models all performed well—and better than Charlson. When accurate risk adjustment is key, researchers may need more complex measures of total morbidity burden than Charlson.

Propensity Score Modeling of Antibiotics from Inpatient Data

Michael O'Connell*, *Insightful*

Retrospective observational studies using medical claims examine the effect of alternate treatments but have a number of recognized limitations compared with randomized prospective clinical trials, most notably with respect to balance in known and unknown prognostic factors that may cause a preference for one treatment over another. Various methods have been used to address such differences, including case-control matching on known factors such as age, gender, or significant comorbidities. This paper describes a method used to compare length of stay for patients receiving Zyxon and Vancomycin as treatments for a variety of infection conditions in 120,465 patients identified in medical claims data collected from more than 500 hospitals representing more than 12.7 million patients. The analysis includes logistic regression and screening of additional variables to generate a propensity score model for treatment and a variety of matching techniques for selecting controls with the closest propensity score for each case. Hypothesis testing using simple two-sample comparisons and survival curve comparisons, combined with graphical analysis before and after the matching, assess covariate balance and treatment effects on length of stay and cost.

Analysis Methods for Volume-outcome Studies

Katherine Panageas*, *Memorial Sloan-Kettering Cancer Center*

Numerous studies have appeared in the literature in recent years linking hospital and surgeon procedure volume with patient outcome. Evaluation of an association between hospital or surgeon procedure volume and outcome involves complex statistical issues that arise from the fact that the unit of observation is the patient, but these studies include multiple patients per hospital or surgeon as well as multiple hospitals or surgeons. Hence, patient outcomes tend to be correlated or clustered within hospitals or within surgeons. In the presence of clustering, it is well-known that standard statistical methods are not valid. Furthermore, the volume-outcome setting is unique in that volume reflects both the primary factor under study and the cluster size—a fact that may well invalidate assumptions inherent in methods that correct for clustering. Through a simulation study, the statistical validity of available statistical techniques is evaluated critically. I compare properties of generalized estimating equations (GEE), random effects models, and the weighted GEE approach to account for informative clustering in the context of volume-outcome studies.

A New, Nonparametric Method for Predicting Health Care Costs with Heteroscedasticity in Risk-adjustment Model

Andrew Zhou*, *University of Washington*; Hauzhen Lin

I present a new, non-parametric heteroscedastic transformation regression model that allows me to predict the expected value of an outcome of a patient with given covariates when the distribution of the outcome is highly skewed with a heteroscedastic variance. In the new model, I allow both the transformation function and the error distribution function to be unknown. I show that estimators for regression parameters, the expected value of the original outcome, and the transformation function converge to their true values at the rate $n^{-1/2}$, the convergence rate that one can expect only for a parametric model. In a simulation study, I demonstrate that the proposed nonparametric method is robust with little loss of efficiency. Finally, I apply the new model to a study on health care costs.

I-4 (Invited): Predicting High-cost Users of Medical Care and the Persistence of High Expenditures over Time

Using the SF-12 To Predict Health Care Expenditures

John Fleishman*, *Agency for Healthcare Research and Quality*; Joel Cohen, *Agency for Healthcare Research and Quality*; Mark Kosinski, *Quality Metric*

Risk adjustment models often use demographic or diagnostic information to predict health care expenditures. Some research, based on samples from restricted populations, suggests that patient-reported health status measures can enhance prediction of expenditures and improve risk adjustment. This project examines relationships between measures of physical and mental health status, based on the SF-12 fielded in the MEPS self-administered questionnaire in 2000, and expenditures for health care in 2001. We examine the extent to which the SF-12 physical and mental health scores improve prediction over and above demographic characteristics and self-reported chronic conditions. We also examine whether the SF-12 adds to predictions based on prior expenditures. The results can point to potential enhancements in risk adjustment models.

An Evaluation of the Performance of Prediction Models To Identify High-expenditure Cases

Steven B. Cohen*, *Agency for Healthcare Research and Quality*; Trena Ezzati-Rice, *Agency for Healthcare Research and Quality*; William Yu, *Agency for Healthcare Research and Quality*

In order to satisfy analytic objectives for nationally representative population based surveys, the adopted sample designs often include over sampling techniques to ensure sufficient sample sizes are achieved for specific policy-relevant subgroups. This strategy is attractive in terms of both cost efficiency and precision, with respect to meeting underlying survey design requirements. For population subgroups defined by characteristics that are more static in nature, such as race/ethnicity, gender, age interval, and chronic conditions of long durations, ensuring sufficient sample size through the implementation of an over sampling strategy is a more straight forward operation. Alternatively, achieving sample size targets for population subgroups that are more dynamic in nature, such as the poor or near poor, individuals with high levels of medical expenditures, and the uninsured, is a more difficult enterprise. In this paper, the performance of alternative prediction models to identify future high expenditure cases is evaluated.

The Impact of Diagnosis Accuracy on Predictive Power of Cost Prediction Models Using the MEPS

Arlene Ash*, *Boston University*; Joel Cohen, *Agency for Healthcare Research and Quality*; John Fleishman, *Agency for Healthcare Research and Quality*

A number of models have been developed that use claims data for purposes of risk adjustment and expenditure prediction. One widely used algorithm, developed by DxCG, Inc., uses ICD-9 codes, typically obtained from large claims databases, to generate expenditure predictions. Applying this algorithm to household survey data

presents some challenges. One issue is whether using three-digit ICD-9 codes versus five-digit ICD-9 codes has a major impact on the expenditure predictions. A second issue revolves around insurance status: Different prediction models have been developed for data from different claims databases (private payer, Medicare). Household survey respondents, however, may have multiple sources of insurance in the course of a year, and procedures for dealing with such individuals need to be developed. This presentation will discuss these issues in the context of using DxCG models to predict expenditures for respondents in the Medical Expenditure Panel Survey, a nationally representative household survey.

C-6 (Contributed): Metaanalysis Methods

Bayesian Metaanalysis of the Dose-response Relationship of Alcohol Consumption and Health Outcomes

Michael Stoto*, RAND Corporation; Graham Colditz, Brigham and Women's Hospital; Sharon-Lise Normand, Harvard Medical School

To analyze the association between alcohol consumption and a variety of health conditions, I developed a Bayesian approach to meta-regression. The primary analytical variable in our analysis is the logRR for the health outcome under study at a particular exposure level within a particular study. Within the study, I posit that the logRR has a linear or quadratic relationship to exposure, namely $Y_i = \beta_i + u_i$ and $u_i \sim \text{Normal}(0, \sigma_i)$. To accommodate inter-study variation in the parameters of this relationship, I assume they follow the model $\beta_i = \mu + d_i$ where $d_i \sim \text{Normal}(0, \tau^2)$ and $\mu \sim \text{Normal}(0, V_\mu)$ and $\tau^2 \sim \text{Inverse-Gamma}(a, b)$. This approach provides a unified framework for utilizing all the available data, permitting us to account for several key sources of variation: sampling error due to variability in the risk of the health condition within each study, systematic variation across subgroups defined by study levels of alcohol intake and other study-level covariates, and random inter-study variation due to differences in conducting each study. I illustrate this approach to analyze approximately 60 epidemiological studies of breast cancer and alcohol consumption.

Fixed and Random Effects Sequential Models for Monitoring Quality of Care

Karl Heiner*, SUNY New Paltz; Bruce Agins, New York State Department of Health

Fixed and random effects models are available for monitoring performance as measured by quality-of-care indicators. When measures are gathered sequentially, the power prior used in meta-analysis is available for incorporating the historical information into the prior. How this approach is helpful for increasing precision when scoring and ranking care providers is demonstrated. Examples are drawn from quality-of-care data gathered routinely from the medical records of individuals with HIV in New York State.

Separating the Effects of Publication Bias and Heterogeneity in Systematic Reviews

Norma Terrin*, Tufts-New England Medical Center; Michael Dowd, Tufts-New England Medical Center; Christopher Schmid, Tufts-New England Medical Center

Publication bias and related biases that favor the inclusion of positive results in systematic reviews can lead to overly optimistic estimates of treatment benefit. Heterogeneity occurs when results vary from study to study because of differences in study protocol, study quality, and patient characteristics. Heterogeneity can create the appearance of publication bias even when bias is absent. We apply selection models with and without covariates to 206 published meta-analyses of binary outcome studies to see whether adjusting for covariates altered findings of publication bias. Results show that 26 of the 206 meta-analyses had statistically significant ($p < 0.05$) publication bias when analyzed with selection models with no covariates. When year of publication was included as a covariate, five of the 26 meta-analyses (19%) became non-significant for publication bias. Inclusion of quality measures in the selection model also changed a finding of publication bias in some cases.

Hospital Ownership and Financial Performance: An Integrative Research Review

Karen Eggleston*, Tufts University; Yu-Chu Shan, Naval Postgraduate School

The large empirical literature on hospital ownership and performance gives frustratingly unclear evidence, inviting selective reference to studies that support an analyst's views. This review applies quantitative meta-analysis techniques to analyze the empirical literature between 1990 and 2004 on U.S. hospital ownership and financial performance. I use meta-analysis and metaregression methods with input from a survey of 13 top empirical researchers in the field, paying special attention to how studies account for market interaction and selection bias. Our results suggest the contradictory findings derive largely from differences in authors' underlying theoretical frameworks, empirical model specifications, and assumptions about the functional form of the dependent variables. Weaker methods and functional forms tend to predict larger differences in financial performance between private not-for-profits and for-profits. Studies that control for a wider range of confounding factors—including at the patient, hospital, and market levels or using panel data estimation techniques—find less difference in performance between for-profit and private nonprofit hospitals.

Quality Review of Recently Published Metaanalyses in a Five-year Period

Kelly Zhou*, Brigham and Women's Hospital; Jui Bhagwat, Brigham and Women's Hospital; Jacqueline Campbell, Brigham and Women's Hospital; Daniel Goldberg-Zimring, Brigham and Women's Hospital; Ferenc Joesz, Brigham and Women's Hospital; Lucila Ohno-Macado, Brigham and Women's Hospital

Evidence-based medicine (EBM) is useful to systematically and efficiently summarize existing clinical trials without incurring the high cost of conducting new prospective studies. Quality meta-analyses are essential and recommended. In this evaluation, I focused on EBM for diagnostic imaging studies in the latest five-year period. I performed a comprehensive review to evaluate the quality and validity

of recent meta-analyses based on the EBM quality guidelines, such as the Standards for Reporting of Diagnostic Accuracy (STARD). I focused on articles in Radiology and the articles published by highest frequency in Pubmed by anatomical region and imaging modality. I summarized in the following areas: topic, image modality, sample size, extraction method, number of extractors, test of homogeneity and analytic models, outcome measures, software programs, affiliation of first author, and whether the articles followed the STARD principles. Finally, I evaluated the meta-analyses with respect to whether they were able to provide conclusive results, which would represent a valuable alternative to prospective studies.

The State of the Art in Metaanalyses of Diagnostic Tests

*Christopher Schmid**, Tufts-New England Medical Center; *Mei Chung*, Tufts-New England Medical Center; *Joseph Lau*, Tufts-New England Medical Center; *Athina Tatsioni*, Tufts-New England Medical Center

We report on a systematic review of meta-analyses of diagnostic test performance published through 2003, assessing their quantity, quality as measured by adherence to the Standards for Reporting of Diagnostic Accuracy (STARD), and methods of analysis. Our study discovered 251 meta-analyses: six published before 1990, 25 between 1990–1994, 94 between 1995–1999, and 126 since 2000. Of these, 50% involved imaging tests, 27% biomarkers tests, 17% clinical examinations, 8% histological or cytological tests, and 4% electrophysiological evaluations. The two major search strategies were Medline (90%) and bibliographies (74%); only 16% used unpublished sources. The majority reported languages searched, an increasing number of which included non-English literature. Quality of reporting has improved with more studies using appropriate methods of enrollment and blinded assessment. Eight hundred seventy eight random effects and summary receiving operating characteristic (SROC) curve methods are supplementing or replacing older methods based on fixed effect summaries of sensitivity and specificity in more than 50% of meta-analyses. Examination of between-study heterogeneity with meta-regression is still underutilized, however.

I-5 (Invited): Imputation in High-dimensional Complex Surveys

Multiple Imputation Using Chained Hierarchical Models

*Recai Yucel**, University of Massachusetts, Amherst

Multiple imputation is an increasingly popular method for handling missing data due to item nonresponse in surveys. When using multiple imputation, it is beneficial to reflect the sample design in the imputation model. If the sample design involves clustering, one way to represent the cluster effects is via random effects in the imputation model. Although this idea has been developed in detail for imputing continuous variables, it is less well-developed for imputing categorical variables and mixtures of categorical and continuous variables. In this paper, we describe two approaches to producing multiple imputations for such variables. The first approach extends the general location model proposed by Olkin and

Tate (1961) to include random effects. Imputations under this approach are drawn from the joint predictive distribution of the missing values, and thus follow the fully model-based paradigm for multiple imputation. This approach is problematic in highly multivariate problems, however, due to the number of parameters in the imputation model. For such situations, we propose an extension of the methods given by Raghunathan et al. (2001), in which we produce imputations by fitting chained hierarchical models and by drawing missing values variable-by-variable from the chained models. We illustrate and compare these techniques using simulated data.

Multiple Imputation by Ordered Monotone Blocks: The Case of the Anthrax Vaccine Clinical Trial

*Fabrizia Mealli**, University of Florence; *M. Baccini*, University of Florence; *S. R. Cook*, Columbia University; *C. Frangakis*, Johns Hopkins University; *F. Li*, Johns Hopkins University; *D. B. Rubin*, Harvard University

Multiple imputation generally involves specifying a joint distribution for all variables in a dataset; the data model often is supplemented by a prior distribution for the parameters' vector governing the distribution of the variables in the Bayesian setting. The Anthrax Vaccine Trial data created new challenges for multiple imputation because of the large number and different types of variables in the dataset and the limited number of units within each treatment arm: In order to ensure no data from one treatment arm contaminates imputed data from another arm, imputations must be done independently across treatment arms. In addition, data models for multiple imputation often are based on the multivariate normal or general location model, neither of which is appropriate for the Anthrax Vaccine dataset. An intuitive method for handling such complex datasets with missing values is to specify for each variable with missing values a univariate conditional distribution given all other variables. Such univariate distributions take the form of regression models (e.g., linear regression, logistic regression), which are straightforward to work with and can reflect different data types accurately. Software such as MICE and IVEWare impute missing data this way. Imputation based on univariate conditional distributions is valid for monotone missing data, if for each variable, the univariate distribution involved is conditional only on those other variables that are more observed than the variable being imputed. However, when missing data are not monotone, univariate imputation strategies have the theoretical drawback that the collection of fully conditional distributions may not correspond to any joint distribution for all the variables. The multiple imputation proposed here, and implemented for the Anthrax Vaccine Trial as the motivating case, aims to capitalize on the simplicity of univariate conditional modeling while minimizing incompatibility. Different types of univariate models also are used, depending on the variables being continuous, semi-continuous, binary, ordinal, or categorical.

Multiple Imputation of Missing Income Data in the National Health Interview Survey

*Nathaniel Schenker**, National Center for Health Statistics

The National Health Interview Survey (NHIS) provides a rich source of data for studying relationships between income and health and for monitoring health and health care for persons at different income levels. However, the nonresponse rates are high for two key items: total

family income in the previous calendar year and personal earnings from employment in the previous calendar year. To handle the problem of missing data on family income and personal earnings in the NHIS, multiple imputation of these items—along with personal earnings status and ratio of family income to the Federal poverty threshold (derived from the imputed values of family income)—was performed for the survey years 1997–2002. This presentation describes the approach used in the multiple imputation project and evaluates the methods via analyses of the multiply imputed data. The analyses suggest imputation corrects for biases that occur when estimates are based on just the complete cases and that multiple imputation results in gains in efficiency.

I-6 (Invited): Statistical Issues in the Hospital CAHPS (HCAHPS) Survey

Overview of HCAHPS, Design of the HCAHPS Instrument, and Political Issues

Paul Cleary, Harvard Medical School*

This presentation will begin with an overview of the HCAHPS project, including the role it plays in providing information to consumers about the quality of care at hospitals. The cyclical process of designing the instrument also is described. The discussion of survey content will include the choice of appropriate scales for the items and the use of screener items, criterion variables, demographic items, and rating items. Common pitfalls, such as misleading and inappropriately ordered questions, also will be discussed. In the latter part of the talk, the roles played by the various statistical methods used in designing a survey will be presented. Particular emphasis will be placed on psychometric concepts such as validity and reliability, as these often are not included in statistics programs. The constraints placed on the survey design by political forces also will be discussed.

Issues Concerning Sample Size Calculation and Reporting

Marc Elliott, RAND Corporation*

This presentation will encompass the finite population versus infinite population controversy, the difficulties with constructing reports for hospitals when there is substantial variation in sample size (particularly several hospitals with small sample size), and whether case-mix adjustment should be used when making reports. The issue of shrinkage, where hospitals with small sample sizes are pulled toward the mean more than hospitals with big sample sizes, and the concerns of the hospitals about this phenomenon also will be discussed.

Hierarchical Factor Analysis for Survey Data with Structured Nonresponse

James O'Malley, Harvard Medical School*

Health care quality surveys in the United States are administered to individual respondents (hospital patients, health plan members) to evaluate performance of health care units (hospitals, health plans). Due to both planned item nonresponse (caused by screener items and associated skip patterns) and unplanned nonresponse, quality measures—such as item means—are based on different subsets of the survey respondents. For better understanding and more parsimonious reporting of dimensions

of quality, we analyze relationships between quality measures at the unit level by applying techniques such as factor analysis to covariance structure estimated at the unit level in a hierarchical model. At the lower (patient) level, we first fit generalized variance-covariance functions that take into account the nonresponse patterns in the survey responses. A between-unit covariance matrix is then estimated using a hierarchical model, which evaluates the fitted generalized variance-covariance functions to account for sampling variation. Maximum quasilielihood and Bayesian inferential procedures are used for model fitting. At the second (plan or hospital) level, we propose comparing two analytic strategies: estimating an unstructured covariance matrix and applying an exploratory factor analysis to summarize relationships and estimating a factor analytic structure integrated into the model (thus more closely related to confirmatory factor analysis). The latter strategy allows specific hypotheses concerning the number of factors and the grouping of items into related factors that define composite items to be tested.

C-7 (Contributed): Advanced Methods for Assessing Intervention Effects

Estimating Drug Effects in Claims Data Using the Prescribing Physician as an Instrumental Variable

M. Alan Brookhart, Brigham and Women's Hospital; Sebastian Schneeweiss, Brigham and Women's Hospital*

Post-marketing observational studies of the safety and effectiveness of prescription medications are critically important but fraught with methodological problems. The data sources available for such research often lack information about indications and other important confounders for the drug exposure under study. Instrumental variable (IV) methods have been proposed as a potential approach to control confounding by indication in non-experimental studies of treatment effects; but good instruments are hard to find. I propose a new IV based on a prescribing physician's preference for one drug relative to a competing therapy. I illustrate the use of this IV in a study comparing the GI effects due to exposure to COX-2 inhibitors relative to non-selective NSAIDs.

Conventional multivariable regression found no protective effect due to COX-2 use within 120 days from the initiation of treatment. The proposed IV method, however, attributed a protective effect to COX-2 exposure compatible with randomized trial results. Future work is needed to examine how modest violations of the exclusion assumption bias the IV results.

Prescription Drug Insurance and Its Effect on Prescription Drug Utilization and Health of the Elderly

Nasreen Khan, University of Illinois at Chicago; Robert Kaestner, University of Illinois at Chicago; Swu Jane Lin, University of Illinois at Chicago*

Approximately 30% of the elderly do not have prescription drug coverage. To remedy this problem, the Medicare Modernization Act was passed recently and the government plans to provide limited drug coverage to the elderly beginning in 2006. Surprisingly, little is known about how drug coverage will affect health. We examine the

effect of drug coverage on drug use, use of other medical services, and health of the elderly. Data are from the 1992–2000 Medicare Current Beneficiary Surveys. I use two empirical approaches: fixed-effects and instrumental variables (IV). The fixed-effects approach uses longitudinal data and controls for unmeasured, person-specific effects that may confound the relationships of interest. The IV approach uses arguably exogenous variation in prescription drug coverage to obtain estimates of the relationships of interest. Estimates indicate prescription drug coverage has a significant effect on drug use and health after controlling for individual characteristics. However, once controls for unmeasured heterogeneity were included, I did not observe an effect of drug coverage on health; drug coverage still had an effect on drug utilization.

Evaluation of a Nursing Home Informatics Tool To Reduce Adverse Outcomes: Methodological Issues

*Richard Gardiner**, *New York Association of Homes and Services for the Aging*

The evaluation of a clinical informatics tool to reduce adverse events in nursing homes is complicated by many factors. Treatment nursing homes (n=91) were not randomly selected; outcome measures were influenced by patient mix, pre-existing trends, and high within-home variation; and effective use of clinical information varied between homes. Predictive accuracy was evaluated by comparing predicted against actual adverse events. Experimental controls were chosen using a variant of the nearest neighbor clustering methodology using home performance, size, and other demographic characteristics to identify closely matched homes. Experimental homes were classified by their level of use of the information system. A mixed model, using repeated measures, was used to evaluate the impact upon adverse event rates between homes throughout time.

Professional Caregiver Insurance Risk: Implications of Health Care Provider Insurance Risk Assumption

*Thomas Cox**, *Seton Hall University College of Nursing*;
Colene Bryne *New York Association of Homes and Services for the Aging*

Professional Caregiver Insurance Risk refers to insurance risk transfers from insurers or government to health care providers (HCPs) in capitation contracts, DRGs, prospective payment plans, and profit-sharing mechanisms. Risk-disaggregating insurance risk transfers to inadequately capitalized HCPs force them to act as inherently and necessarily inefficient mini-insurers. HCP-managed insurance portfolios are more variable than insurers' portfolios, leading to reduced service capacity when HCPs manage insurance risks. While some health care providers will benefit, some will not, experiencing higher than average operating losses due solely to the disaggregation of risk. Risk disaggregation means HCPs must plan to deliver lower levels of service than anticipated in insurers' premium rates in order to manage their inappropriate and potentially illegal insurance risk assumption. This paper demonstrates that HCPs must inefficiently manage insurance risks using the Law of Large Numbers and the Central Limit Theorem and typical actuarial risk theoretic assumptions, demonstrating the adverse impact of using risk transferring health care finance mechanisms as matters of public policy.

Using CHAID for Instrument Development and Practice Guidelines

*James Bost**, *University of Pittsburgh*

This presentation will focus on the use of CHAID (Chi-squared Automatic Interaction Detector) in health services and outcomes research. CHAID uses a statistical programming algorithm to classify individuals based on a set of potential predictor variables. CHAID is used to construct trees that split based on the best predictors at each level to optimize prediction. For binary outcome variables, it uses the chi-squared test to determine the best split; for continuous dependent variables, it uses F-tests. CHAID has been used extensively in market research and has been used recently as a methodology for health care research, due, in part, to advances in statistical software. This presentation will focus on its use in two areas: item reduction in the development of an instrument to assess medication adherence for schizophrenics and determining which patients with a particular condition will most benefit from a particular treatment intervention.

A Case Example of Data Mining and Causal Analysis—Surprising? True? Useful?

*Andrew Brunskill**, *University of Washington*

I used two relatively novel approaches to analyzing a classic survey-based dataset that related reported access to health care to other variables. The two approaches are “Orange” for data mining program and “Tetrad” to infer directed acyclic graphs (causal networks). I present the results and argue that each program makes a particular and valuable contribution to data understanding—especially hypothesis generation—that may not be produced by conventional analytic approaches. The generation of novel, useful, testable hypotheses appears to be of substantial benefit for health care policy and practice.

I-7 (Invited): Causal Inference with Longitudinal Data

The Analysis of Sequential Treatments: Practical Issues

*Babette Brumback**, *University of Florida*

Treatment for chronic conditions often requires sequential modification, and this presents interesting challenges for evaluating efficacy. This talk will focus on two modeling approaches useful for specifying meaningful targets of inference: marginal structural models and structural nested mean models; it will also review why some standard targets are not as meaningful. Because practical constraints invariably limit our ability to estimate the targets, the modeling process inevitably involves compromises. As with most statistical analyses, the compromises can be interpreted as aggregating data across ‘similar’ individuals and/or time periods. Related to the problem of determining similarity are other common statistical problems such as choice of outcome, modeling the treatment process, selecting a baseline, and evaluation of competing models. These problems will be illustrated with an analysis of longitudinal data on treatment for HIV.

Random Effects Logistic Models for Analyzing Efficacy of a Longitudinal Randomized Treatment with non-Adherence

Dylan Small*, *University of Pennsylvania*

We present a random effects logistic approach for estimating the efficacy of treatment for compliers in a randomized trial with treatment non-adherence and longitudinal binary outcomes. We use our approach to analyze a primary care depression intervention trial. The use of a random effects model to estimate efficacy supplements intent-to-treat longitudinal analyses based on random effects logistic models that are commonly used in primary care depression research. Our estimation approach is an extension of Nagelkerke et al. (2000, *Statistics in Medicine*)'s instrumental variables approximation for cross-sectional binary outcomes. Our approach is easily implementable with standard random effects logistic regression software. We show through a simulation study that our approach provides reasonably accurate inferences for the setting of the depression trial under model assumptions. The sensitivity of our approach to model assumptions is evaluated for the depression trial. This is joint work with Tom Ten Have, Marshall Joffe and Jing Cheng.

Inferring Causal Effects in Clustered Longitudinal Data: The Effect of Publicly Reporting Outcomes in Cardiac Surgery

Mary Beth Landrum*, *Harvard Medical School*; Robert S. Huckman, *Harvard University*; David M. Cutler, *Harvard University and NBER*

Provider profiling has been urged by many as a fundamental step in medical care reform. Despite the building momentum for this strategy, the underlying premise of profiling - that disseminating information about provider quality will lead to improved quality of care - has yet to be established. Using a longitudinal data set containing all individuals who underwent CABG in a hospital in New York State over a 9-year period, we consider how report cards affect the behavior of individual providers, specifically providers who are publicly identified as being significantly better or worse than their peers. Our analysis addresses three intended goals of provider profiling. First, we examine whether report cards lead poor-performing providers to improve their performance relative to their peers, adjusted for changes in patient severity. Second, we consider whether report cards lead to increased volume for highly rated providers and lower volume for poorly rated providers. Finally, we consider how report cards influence the allocation of patients to particular physicians, in particular whether report cards lead to more severe patients being operated on by higher quality physicians. We compare several approaches for making causal inferences in this context including longitudinal random effects and fixed effects models and propensity score approaches. We find estimates of the effect of publicly reporting quality data are sensitive to inferential approach and highlight assumptions underlying each approach to explain the differential effects. Understanding the true effects of report cards on access and quality of care becomes increasingly important as more states and other organizations are publicly releasing quality report cards for an increasing number of conditions and as insurers are starting to use provider-specific quality in determining appropriate reimbursement.

I-8 (Invited): Population Needs-based Funding Models

Health Care Funding Models Based on Population Needs: The UK Experience

Peter C. Smith*, *University of York*

Almost all developed nations with statutory health insurance try to secure a fair allocation of funds between insurers. In social insurance countries, the emphasis is on creating a level playing field so social insurers can offer comparable insurance coverage for similar insurance premiums. In systems funded by public taxation, the emphasis has been on securing geographical equity through the funding mechanism. Among publicly funded systems, the United Kingdom has been in the vanguard of countries seeking a fairer allocation of funds between geographical areas. In the 1970s, the seminal report by the Resource Allocation Working Party (RAWP) proposed a formulaic approach to allocating funds between regions that was implemented gradually across England. The RAWP approach sought to make operational the principle of horizontal equity (equal access for equal need) and has been steadily refined over the succeeding decades. Significant technical developments have included the use of small-area data, and deployment of multilevel (hierarchical) and instrumental variable econometric modeling techniques to improve the robustness of the empirical modeling. This paper reports developments in England and the other countries of the UK.

National Population Needs-based Funding: The New Zealand Experience

Peter Crampton*, *University of Otago*

New Zealand first introduced national population needs-based funding for health services in 1983. Since then, a number of funding formulas have been used at regional and subregional levels. This presentation gives a brief overview of health services structure and funding in New Zealand, describes the philosophical basis for New Zealand's approach to needs-based funding, and explores how this approach differs from insurance-based funding. The presentation then provides an overview of the funding models currently used in New Zealand, followed by a more detailed description of the components of the regional and subregional population-based funding formulas, which include weighting factors for age/gender, socioeconomic deprivation, ethnicity, and 'unmet need'. Finally, the implications for health policy of New Zealand's approach to population needs-based funding are discussed, including the impact of the funding model on equity of resource distribution and the incentive structures for district health boards and primary health organizations.

Allocating Health Care Resources According to Need: An Approach to Developing Needs-based Formula Using Linked Health Survey and Administrative Data

Jeremiah Hurley*, *McMaster University*

A common goal of health policy is to allocate public health care resources according to need. This paper presents an approach to developing needs-based funding formulae using individual-level linked health survey and utilization data. Needs-based funding shares

are developed in three basic stages: (1) estimate the full utilization model, including both need-related and non-need-related adjusters; (2) predict individual-level needs-based home care utilization holding all non-need factors constant; (3) use individual-level estimates and sample weights to develop regional needs-based allocations. The approach is then applied to the allocation of the home care budget in Ontario, Canada. The results suggest methods based on such data offer considerable advantages while raising several new challenges.

C-8 (Contributed): Multilevel Models and Bayesian Methods

Hierarchical Modeling of Inpatient Rehabilitation Facility Responses to Prospective Payment

Susan Paddock, RAND Corporation*

The Centers for Medicare and Medicaid Services (CMS) implemented the Inpatient Rehabilitation Facility Prospective Payment System (IRF PPS) for hospital inpatient rehabilitation care for Medicare beneficiaries in 2002. The goal of the IRF PPS is to enhance access to inpatient rehabilitation care by compensating providers based on their actual case mix and thereby ensuring beneficiary access to care while controlling Medicare's inpatient rehabilitation expenditures. The implementation of prospective payment systems can lead to a range of effects among providers, namely reducing the amount of care delivered (e.g., reducing length of stay), selecting patients in such a way to maximize profitability, and altering coding practices. In this talk, I illustrate how I am using Bayesian hierarchical models to examine provider-level responses to the implementation of the IRF PPS and characterizing variation in these responses with respect to facility-level characteristics. I also describe our strategy to derive a measure of facility-level case mix change in the presence of changes in the practice of coding patient severity.

Extending the Capture-recapture Methodology To Estimate Subpopulation Sizes

Ulysses Diwa, University of Connecticut; Dipak Dey, University of Connecticut*

The general capture-recapture methodology is gaining popularity in epidemiological studies. It is used primarily to estimate the size of an underlying population of interest. However, it does not readily provide solutions to questions about sizes of subpopulations. This paper attempts to extend the methodology to answer such questions by presenting four methods of estimating the sizes of (sub) populations. One is the direct application of the basic capture-recapture methodology. The second presents a kind of conditional maximum likelihood estimator. A third approach is the Bayesian analog of the second. Finally, a Bayesian approach that models the reparameterized probabilities is presented.

GEE and Summary Measures Analysis in Medication Use over Time in Asthmatic Pregnant Women

Tebeb Gebretsadik, Vanderbilt University; Patrick Arbogast, Vanderbilt University; Tina Hartert, Vanderbilt University; Ayumi Shintani, Vanderbilt University*

We compared two analytical approaches to describe patterns in asthma

medication use over time in a stable cohort of asthmatic pregnant women: the summary measure method—a widely used method in health utilization research that uses summary measures, such as weekly proportions of medication users, as a unit of observation—and the Generalized Estimating Equations (GEE) method, which uses individual observations as a unit of observation. In a cohort of 8,142 pregnant women enrolled in the Tennessee Medicaid program, both methods were used to analyze asthma medication use during three predefined time intervals: pre-pregnancy, early pregnancy, and the second trimester to the date of delivery. For the three classes of asthma medications, the summary measure method had wider confidence intervals compared to the GEE method by a difference of 2 to 4 fold. The estimations from the summary measures method were less efficient in measuring the decline in medication use shortly after pregnancy.

Methods for Profiling the Value of Hospital Care following Acute Myocardial Infarction

Justin Timbie, Harvard Medical School; Sharon-Lise Normand, Harvard Medical School*

In developing tiered provider networks and pay-for-performance programs, health plans have attempted to estimate and compare the “value” of treatment across hospitals by combining measures of quality and cost-efficiency. This analysis compares three methods to assess the value of hospital care for AMI in 69 Massachusetts acute care hospitals in 2003. For simplicity, two outcomes are considered: in-hospital survival and cost-per-episode of care. The first approach models these outcomes independently using hierarchical logistic regression models and hierarchical linear models, respectively. Two methods for computing hospital-specific estimates of the value of AMI care are obtained by estimating the joint probability of exceeding a distributional threshold and combining the two outcomes using a cost-effectiveness ratio. In the third approach, survival is modeled as a function of spending in each hospital using a hierarchical linear model with random intercepts and slopes. Hospitals classified in the high value tier and outlying hospitals are compared across the three methods and the advantages and disadvantages of each modeling approach are discussed.

Model Selection versus Information Selection for Performance Evaluation: Where Best To Invest?

Kevin L. Sloan, U.S. Veterans Health Administration/University of Washington; James F. Burgess, Jr., U.S. Veterans Health Administration; Xiao-Hua (Andrew) Zhou, University of Washington; Chuan Zhou, Vanderbilt University; Paul Fishman, Group Health Cooperative; Li Wang, U.S. Veterans Health Administration*

Profiling techniques promise to provide valid provider performance evaluation results by taking into account differential patient populations. However, little attention in the literature has been paid to the tradeoffs involved in investing in more sophisticated statistical models vs. gathering salient provider-specific information and their impact on the results of provider-performance evaluation. Furthermore, although Bayesian and hierarchical approaches to risk adjustment expand analytic options for a proper leveling of the playing field, such approaches also require careful consideration of the policy consequences of the decisions made when choosing models.

In examining variations in outpatient cost for 144 hospitals in the U.S. Veterans Health Administration system, we start with results from the

standard, frequentist profiling of outpatient cost. We then consider the hierarchical arrangement of patients into facilities, apply alternative analytic methods, and compare the facilities flagged as cost outliers. In a stepwise fashion, we examine how increasingly elaborate models impact the assessment of facility performance. Contrasts include non-hierarchical vs. hierarchical models, frequentist vs. Bayesian techniques, and ignoring vs. including facility-level information—all interpreted from a health care management policy perspective.

We find that the choice of information to include—particularly relevant, facility-level information—is more important than choice of statistical model specification, although more complex model specifications offer a better foundation to respond to criticisms related to failure to account for special characteristics of poorly performing programs.

Hidden Markov Models for Longitudinal Comparisons

Steven Scott*, *University of Southern California*

Medical researchers interested in temporal, multivariate measurements of complex diseases have begun developing health state models recently that divide the space of patient characteristics into medically distinct clusters. Health services researchers sometimes use k-means clustering to form the health states and a first-order Markov chain to describe transitions between the states. This procedure ignores information from temporally adjacent observations and prevents the correct propagation of uncertainty through the analysis. A natural way to address these issues is to combine clustering and longitudinal analyses using a hidden Markov model. I fit hidden Markov models to longitudinal data using Bayesian methods that account for all the uncertainty in the parameters, conditional only on the underlying correctness of the model. Potential time inhomogeneity is accounted for by embedding transition probabilities into a hierarchical model that provides Bayesian shrinkage across time. I illustrate this approach by comparing two antipsychotic medications for schizophrenia.

I-9 (Invited): Methods in Longitudinal Data Analysis

Validation of Life Table Approaches to Estimating Population Health Status

Liming Cai*, *National Center for Health Statistics*

Longitudinal databases are ideal for describing the health histories of individuals. Unfortunately, there are no national databases that have followed individuals for an extended time to track health events and health changes at the personal level. Therefore, life table techniques have been used to simulate individuals' health experiences by building simulated populations. But, how accurately do the results from simulated populations reflect actual health experiences? The Cardiovascular Health Study offers an opportunity to address this question because it has 14 years of annual follow-up data on a 6,000-person study group. We will report the results of simulation cohorts from fitting two life table models—the multistate life table (MSLT) and the semi-Markov process model (SMP)—to the actual CMS data. The SMP model incorporates new approaches to dealing with left-censored data. We will focus on comparing distributional statistics for

estimates such as active life expectancy, age at onset of disability, and years in various health. A split sample of the CHS will be used with estimation and verification subsamples.

New Findings on non-Response from the Medicare Current Beneficiary Survey (MCBS)

John Kautter*, *RTI*

The MCBS is a continuous survey of about 12,500 Medicare beneficiaries begun in 1991. Respondents remain in the survey for 3.5 years. Because there are data on the use and cost of health care services and on mortality for all Medicare beneficiaries—both respondents and non-respondents—the MCBS offers an unusual opportunity to study the effect of non-response, both at the onset of the survey and throughout the 3.5-year follow-up period. A recently completed study by RTI International, under contract with the Centers for Medicare & Medicaid Services, analyzed the extent and impact of non-response in the MCBS for 1977–99. The study found that initial refusers were healthier than respondents (perhaps a busy, active senior phenomena). Correction for non-response, using traditional approaches, succeeded in bringing estimates for respondents and non-respondents very close on measures such as per capita Medicare spending.

Handling Incomplete Data in Longitudinal Clinical Trials

Geert Molenberghs*, *Limburgs Universitair Centrum*

The Randomized Controlled Trial (RCT) is used often to establish a causal effect of a new treatment on a response. The allocation of the different treatments must be truly randomized. By keeping the groups as similar as possible at baseline, the effects from factors other than the intervention will be minimal and differences in response on a clinical outcome can be ascribed solely and entirely to differences in treatment allocation. However, in practice, this paradigm is jeopardized in two important ways. First, some patients may not receive the treatment as planned in the study protocol because they are sloppy. Some may take more than planned at their own initiative. In rare cases, patients may even gain access to medication allocated to the other treatment arm(s). Second, some patients may leave the study—some rather early after their enrollment in the trial, some at a later stage. In such cases, virtually no data or, at best, partial data are available. This is bound to happen in studies that run over a relatively long period and/or when the treatment protocol is highly demanding. Thus, in reality, missing data is an almost ever-present problem in clinical trials. In RCTs, these missing data undermine the randomization basis for estimates of treatment efficacy. Regardless of the cause, inappropriate handling of the missing information can lead to bias. In the RCT setting, a commonly used method to analyze longitudinal data with non-response is based mostly on setting a subject's missing response equal to their last observed response (last observation carried forward, LOCF). There are several objections to the use of this method. Other methods include complete case analysis (CC) or simple forms of imputation. This is done often without questioning the possible influence of these assumptions on the final results, even though several authors have written about this topic. We will contend that analyzing the data as if it were complete after carrying the last observation forward is an unscientific approach. One should shift from a LOCF analysis to the

scientific methods at hand, which make the assumptions transparent and thus allow the sensitivity of the conclusions to these assumptions to be assessed and reported.

C-9 (Contributed): Health Care Cost and Payment Systems

Specification of Regression Models in the Development of Inpatient Case Mix Grouping Methodology

*Qian Yang**, Canadian Institute for Health Information; *Jeff Hatcher*, Canadian Institute for Health Information

The main objectives of developing Case Mix Grouping methodology are to establish clinically relevant and resource homogeneous groups of patient cases. To establish the resource homogeneity, regression equations are set up to model the relationship between resource consumption, represented by either patient case cost or length of stay (LOS) and explanatory factors such as diagnoses, interventions, and patient age. Under the classic linear regression framework, the random errors of the model are assumed to be independent and normally distributed with a mean of zero and have constant variances. This study fits the inpatient data with the regression models, analyzes the random errors to verify the assumptions, and proposes adjustments to the model when the assumptions are not met. The study found that the random errors follow lognormal distributions with unequal variances. The regression models therefore used log of total cost or log of LOS as a dependent variable and solved the equations using the weighted least squared approach. The study also found that the effect of any additional diagnosis on resources is multiplicative, rather than additive.

Estimating Incremental Cost-effectiveness Ratios from Cluster-randomized Intervention Trials

*Mohammad Chaudhary**, Johns Hopkins Bloomberg School of Public Health; *Mohamd Shoukri*, King Faisal Specialist Hospital and Research Center

Reliable and accurate interval estimates of incremental cost-effectiveness ratio (ICER) of a health care intervention facilitate policy decisions. Because ICER is generally positively skewed and far from normal and the variance of the ratio estimator is intractable, the health economics literature has suggested a number of alternative approaches to estimating confidence intervals for cost-effectiveness ratios. These approaches have been restricted to individual-level data. The trials randomizing clusters have become particularly widespread in the evaluation of non-therapeutic outreach interventions. The methods for the analysis of cluster-randomized data have been dealt with extensively, however the extension of these methods to cluster randomized trials still remains a challenge. This paper attempts to evaluate the interval estimation methods for ICER for individual-level data when applied to cluster-randomized trials. I simulate the cost and effectiveness data from cluster-randomized trials under alternative scenarios and evaluate the performance of parametric and non-parametric interval estimation methods for ICER.

Estimating the Effects of Drug co-Payments on Statin Adherence Using Cross-sectional Time Series Data

*Teresa Gibson**, Medstat; *Kirsten Axelson*, Pfizer Global Pharmaceuticals; *Tami Mark*, Medstat; *Kimberly McGuigan*, Pfizer Global Pharmaceuticals; *Shaohung Wang*, Medstat

Estimating the effects of prescription drug co-payments on adherence to statin therapy within a cross-sectional time series framework allows evaluation of overall price effects. Furthermore, the effects of the initial co-payment can be separated from the effects of changes in co-payments over time. For this paper, medical and pharmacy claims for statin users were selected from the 2001–2003 Medstat MarketScan database and Generalized Estimating Equation logistic models were used to estimate the effects of co-payments on statin adherence (based upon % of days on therapy) for new statin users ($n=142,341$) and continuing statin users ($n=92,344$). It was found that co-payments were a larger financial barrier to adherence for new users than continuing users (OR: .911 new users $p<.01$, .996 continuing $p=.45$) and the size of the initial co-pay had a larger effect on adherence (OR: .814 new $p<.01$, .916 continuing, $p<.01$) than the change in co-pay over time (OR: .919 new $p<.01$, .999 continuing, $p=.80$). While co-payments serve as a financial barrier to adherence to statin therapy, the size of the initial co-payment plays a significant role in adherence and may be a focus of interventions.

Effect of Health Interventions on Longevity, Morbidity, Years of Healthy Life, and Costs

*Paula Diehr**, University of Washington; *Liming Cai*, U.S. Centers for Disease Control and Prevention; *Ann Derleth*, University of Washington; *Anne Newman*, University of Pittsburgh

Public health interventions can improve population health by changing the proportion of persons initially in each health state or by changing the probabilities of transition from one health state to another. Multi-state life table methods are used to estimate the impact of nine types of interventions on years of healthy life, years of morbidity, longevity, and estimated lifetime medical expenditures. Compared to the status quo, all the interventions improved longevity and years of healthy life, and some improved years of morbidity and lifetime medical expenditures as well. The interventions that improved longevity the most, however, also increased morbidity and medical expenditures. Results differed by the age at which interventions were initiated. The effects of combinations of intervention types were not always additive.

Competing Risk Analysis Applied to Health Economic Evaluations

*George Carides**, Merck Research Laboratories; *Shannan Allen*, Merck Research Laboratories

Health economic evaluations often utilize the cumulative incidence of events in a clinical trial to assign costs associated with one or more treatments. For example, the incidence of end-stage renal disease or stroke observed during a clinical trial may be used with external estimates of cost to estimate the downstream costs associated with these events. A common practice is to apply standard survival analysis methods, such as the Kaplan-Meier method, and Cox proportional hazards regression to estimate these cumulative incidences. The biases inherent in these approaches whenever there exists the possibility that some patients may die without experiencing the cost-generating event(s) are illustrated. An alternative method is presented and illustrated with a case study in diabetic nephropathy.

Health Care Costs Following Treatment Initiation for Alzheimer's Disease (AD) in Managed Care

Ann Harada*, *Prescription Solutions*; Ann Vanderplas, *Prescription Solutions*

A retrospective claims analysis examined five AD regimens: donepezil (DO), galantamine (GA), rivastigmine (RI), memantine (ME), and memantine+AChEI (acetylcholinesterase inhibitor) combination (M+A). Members initiated therapy from January 1, 2004, to June 30, 2004, excluding those with previous AChEI use or those not continuously enrolled in the six-month pre- and post-treatment initiation. Change in pharmacy, medical, and total (pharmacy+medical) costs was analyzed, adjusting for age, gender, comorbidities, and pretreatment total costs. A total of 2,461 were newly started on AD therapy; 65% were female and mean age was 80 years (SD 7.2). Sixty-two percent initiated DO, 14% ME, 11% GA, 8% RI, and 6% M+A. While adjusted pharmacy costs increased for each treatment, all medical costs decreased. Total adjusted cost (PMPM) changes were significantly lower in all regimens except RI: M+A (-\$1,106), ME (-\$1,079), GA (-\$883), DO (-\$723), and RI (-\$150). Initiation of AD therapy was associated with overall PMPM cost decreases in the first six months following treatment initiation.

I-10 (Invited): Selection Bias in Observational Studies

More Ado about Two: Endogenous Switching, Sample Selection, Endogenous Treatment Effects, and the Modified Two-part Model

Joseph V. Terza*, *University of Florida*

The modified two-part model (MTPM), discussed by Mullahy (*Journal of Health Economics*, 1998, 17, 247-281), is extended to account for endogenous switching and its two common incarnations: sample selection and endogenous treatment effects. A minimally parametric version of the model, based on a conditional mean restriction only, is derived. Corresponding method of moments (MM) and two-stage method of moments (TSM) methods are introduced. A fully parametric version of the model, based on an appropriate conditional distribution assumption, also is offered. Corresponding full information maximum likelihood (FIML) and conditional maximum likelihood (CML) estimators are derived. Possible generalized method of moments (GMM) approaches for models based on sets of conditional moment restrictions that lie at an intermediate point along the spectrum between the minimally parametric and fully parametric formulations are discussed. Using simulated data, the MM, TSM, FIML, and CML estimators are compared with simple linear approaches: conventional Heckman-type Mill ratio correction and instrumental variables. The criteria for comparison are small and large sample bias and efficiency. The methods are applied to data from the 2002 National Survey on Drug Use and Health in analyses of the effect of drug abuse on earnings.

Two Approaches to Adjusting for Selection Bias in Cross-sectional Data: Instrumental Variables and Propensity Score

Matthew L. Maciejewski*, *University of Washington, Seattle*; Song Wang, Xiao-Hua (Andrew) Zhou, *University of Washington*

Medicare claims data were obtained on a 2% sample (N=59,089) of Medicare beneficiaries with diagnosed diabetes in 1995. Managed care benefits and out-of-pocket premiums also were obtained from Medicare to generate county-level measures of plan generosity. Cross-sectional analysis of 59,089 patients is performed. Two measures over quintiles for treatment effects are undertaken to compare sample selection modeling strategies: instrumental variables and propensity score matching. A nonparametric bootstrap is performed to calculate confidence intervals for the treatment effects by quintile. Instrumental variables and propensity score matching generate marginal treatment effects that are consistent on average but differ by quintile. Overall, Medicare HMO enrollees have lower mortality rate than FFS enrollees.

Causal Estimation Using Quasi-experimental Designs

Sharon-Lise Normand*, *Harvard Medical School*; Richard G. Frank, *Harvard Medical School*; Thomas G. McGuire, *Harvard Medical School*

Quasi-experiments refer to studies in which randomization is not used to assign subjects to treatment groups and the untreated response for both the treated and control subjects is the baseline measurement. The most common estimator of treatment effect is the difference-in-difference (DID) estimator, defined as the population average difference over time in the control group subtracted from the population average difference over time in the treated group. The DID estimator is popular because it remains unbiased under simple violations of assumptions for the time trends in outcomes and of the comparability of the treatment groups. In this paper, we examine several estimators for inferring causal effects, including regression estimators, inverse weighted estimators, and matching estimators. We describe the robustness of the estimators to violations of the common assumptions, such as ignorability and separability. We illustrate methods using observational data to assess the impact of the introduction of managed behavioral care on treatment utilization and costs.

C-10 (Contributed): Health Disparities and Access to Care

Access to High-cost Medicines in Australia: Evaluating Health Outcomes Using National Claims Data

Christine Lu*, *University of New South Wales*; Ric Day, *University of New South Wales*; Ken Williams, *University of New South Wales*

Access to high-cost medicines, for example tumour necrosis factor inhibitors (TNFIs) for rheumatoid arthritis, is restricted under the Pharmaceutical Benefits Scheme. Prior application for initiating or continuing TNFIs requires detailed information on each patient. Here,

I examine patient health outcomes from using TNFIs. De-identified data on patients using a TNFI (etanercept, infliximab, and adalimumab) were requested from the claims agency: patient demographics, other anti-rheumatic therapies (such as item, dose, treatment duration, reasons of withdrawal), and changes in clinical outcomes (levels of inflammatory markers and joint counts) from August 2003 to March 2005. Only a subset of information collected is entered into an administrative database (such as date, item, patient demographics, cost) and data on health outcomes were unavailable. Also, medical services used by patients are not linked to medicine use. I find that comprehensive medicine use and clinical outcome data need to be accessible to review the appropriate use of these medicines. Urgent enhancement of national administrative databases to accomplish this is recommended.

Racial Disparities in Primary Care and Health Care Utilization at the End of Life

*Andrea Kronman**, Boston University Medical Center;
Arlene Ash, Boston University School of Medicine; *Karen Freund*, Boston University Medical Center

At the end of life, racial minorities spend more time in the hospital and have higher expenditures. I hypothesized that racial disparities in primary care may mediate these disparities at the end of life. I utilized retrospective analysis of Medicare decedents. Blacks and Hispanics were over-sampled. I measured hospital days during the last six months of life and primary care (visits, Ambulatory Care Sensitive Condition (ACSC) admissions, continuity of care) in the preceding 12 months. Multivariate analysis adjusted for demographics, county variation, and comorbidity. Sample (N=140,407) characteristics: mean age 81, female 56%, Black 42%, Hispanic 12%. Whites had fewer hospital days and lower costs (all $P < .001$). Whites had more primary care visits, greater continuity of care, and fewer ACSC admissions. Primary care visits were inversely associated with hospital days. More primary care visits are associated with fewer hospital days at the end of life. Non-whites spent more time in the hospital, which was partially explained by less primary care. Increasing primary care to Medicare beneficiaries could improve end-of-life care and decrease costs.

Health Insurance Availability and Racial Disparities in Total Knee Arthroplasty (TKA)

*Amresh Hanchate**, Boston University School of Medicine;
Arlene Ash, Boston University School of Medicine

Little is known about why racial disparities in TKA are so large. I use a rich longitudinal survey to quantify the role of health insurance availability. The Health and Retirement Study (HRS) is an ongoing biennial survey of persons born before 1948. Data from survey rounds 1995/96 to 2002 were used. Outcome is a binary indicator of first TKA. I used multivariate longitudinal logistic regression analysis to obtain odds ratios (OR) of TKA adjusted for illness burden, functional health, income, education, and geographic region. Six hundred sixty three TKAs were observed (N=21,254). Racial disparity in TKA rate (per 100) mirrored earlier studies (women: white=3.5, black=3.1, Hispanic=2.4; men: white=3.2, black=1.3, Hispanic=1.9). Among the pre-retired (age 64 or younger), lack of health insurance was a strong deterrent to TKA (OR: 0.36, 95% CI 0.21-0.61) compared to those with health insurance. Among the retired, availability of secondary insurance to Medicare significantly facilitated TKA (OR: 1.24, 95% CI 0.99-1.58). This study indicates health insurance availability or unaffordable co-payments may underlie racial disparities in TKA.

Health Infrastructure and Rural Immunization in India

*Arnab Mukherji**, Pardee-RAND Graduate School; *Ashlesha Dator*, RAND Corporation; *Necaj Sood*, RAND Corporation

We examine the role of health infrastructure in expanding immunization coverage in rural India. Using detailed measures of village health infrastructure from the National Family Health Surveys (1993 and 1998 waves), I estimate separate multinomial logit regression models for polio and non-polio vaccines to estimate the probability that a child will receive “no cover,” “some cover,” or “full age-appropriate cover.” The estimates from the regressions show availability of health infrastructure in a village significantly improved vaccination coverage for non-polio vaccines; larger and better-equipped facilities had bigger effects on coverage; and association between the health facilities in the village and polio vaccine coverage was small and statistically insignificant. Our findings suggest health infrastructure, especially primary health care centers and hospitals, play an important role in expanding vaccine coverage for non-polio vaccine preventable diseases. As a significant proportion of children reside in villages with no health facility, this suggests increased availability of health infrastructure is an important tool for expanding immunization coverage in rural India.

Medicare Beneficiaries' Access to Physician Services in Local Markets in 2003

*Anne Ciemnecki**, Mathematica Policy Research, Inc.; *Tim Lake*, Mathematica Policy Research, Inc.

To assess Medicare beneficiaries' perceptions of access to physicians following cuts in Medicare fees, I targeted 11 “hot spots,” where possible declines in physician participation may be causing access difficulties. I conducted a survey of 3,280 beneficiaries. Areas were selected based on evidence of access problems and concerns about declining physician participation. I over sampled beneficiaries who were more likely to encounter problems. Few beneficiaries reported any problems with access to care, and only a small percentage had a problem they attributed to physicians not taking new Medicare patients or limiting Medicare participation. Access problems were more common among subgroups, including beneficiaries who recently moved to the area or recently enrolled in Medicare fee-for-service. The results from this study suggest the reduction in physician fees did not lead to widespread access problems, even in areas with a high potential for problems. Nonetheless, some of the findings—such as higher rates of problems for vulnerable beneficiaries and signs of worsening access for some beneficiaries—provide grounds for continuing to watch for emerging difficulties.