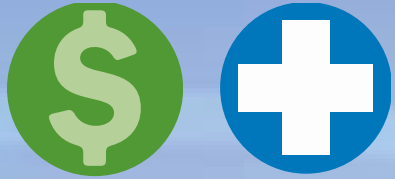# The Integration of Administrative and Survey Data in Support of Medical Expenditure Analyses: Utility and Challenges

Steven B. Cohen

- Measurement and analysis of medical care expenditures

- Data integration model to enhance analytic capacity

- Major sources of administrative data on medical expenditures

- Considerations to assess suitability for integration

- National survey efforts integrated with administrative data

- Opportunities and challenges

Health Policy Significance

Health care expenditures:

- Over one-sixth of the U.S. GDP

- Rate of growth exceeds other sectors of the economy

- Recent moderation in rate of growth

- Expenditure distribution is highly concentrated and subject to visible transitions over time

- Among the largest components of the federal and states' budgets

- Cost containment of continuing concern to private and public payers

Policy Impact
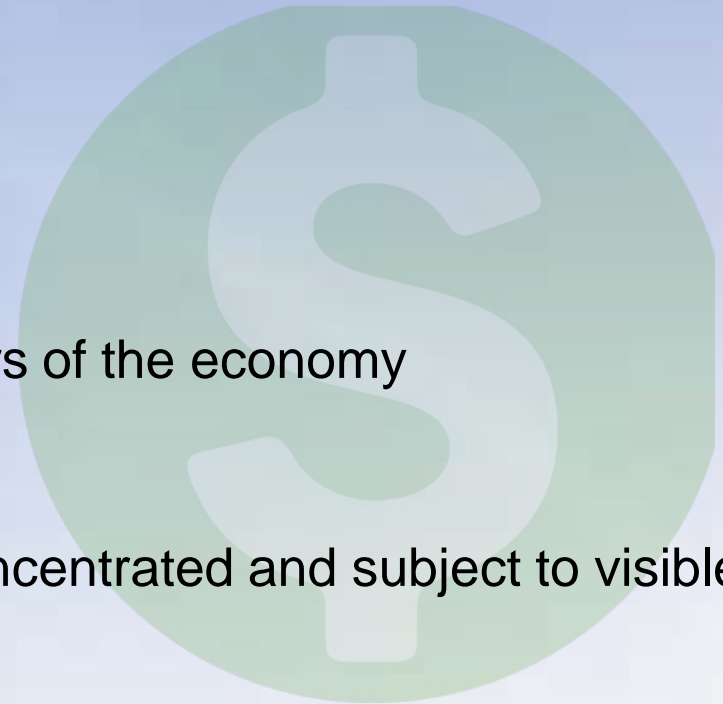
- Inform and evaluate health reform policies; the effect of tax code changes on health expenditures and tax revenue; and proposed changes in government health programs such as Medicare, Medicaid, CHIP and the ACA.

- Identification of health care disparities.

- Measures health care burdens for chronic conditions,

- Serves to measure share of gross domestic product (GDP) attributable to health care and trends over time.

RTI INTERNATIONAL

Data Integration Model

- Data integration is a process in which related and supplemental data from multiple sources are connected into a unified structure.

- The resultant integrated data resource serves as a platform to enhance analytic efforts.

- The data integration model facilitates greater analytic utility for each of the component data sets as a consequence of their "connectivity".

- Data integration is often implemented in a data warehouse or data enclave setting to ensure the extraction, linkage and structure of the combine data resources are presented in a unified manner.

Integrated Survey Design Features

- Direct linkage between sample members in core survey with larger host survey; **administrative records**; or follow-up surveys

- Use of secondary data (e.g. aggregate data at the county/state level) as core component of survey

- Informs sample design, nonresponse and poststratification adjustments, imputation and data supplement for item nonresponse

- Facilitates reductions in measurement error

- Need for greater attention to ensuring confidentiality: limitations in public use data

RTI INTERNATIONAL

Major sources of administrative data on medical expenditures

CMS Medicare and Medicaid claims data

- **Medicare Provider Utilization and Payment Data: Physician and Other Supplier-**administrative claims data for Medicare beneficiaries enrolled in the fee-for-service program

- **Medicaid Analytic eXtract (MAX) data-**person-level data files derived from MSIS data on Medicaid eligibility, service utilization and payments

Major sources of administrative data on medical expenditures

All-Payer Claims Database Systems (APCDs)

- **Healthcare Cost and Utilization Project (HCUP)-** largest collection of longitudinal hospital care data in the US with all-payer, encounter-level information-AHRQ

- **State APCDs-**comprehensive, multi-payer data on the cost, quality, and utilization of health

Medical Billing Systems

- **Patient financial records-**billing/patient accounts; outsourced claims administration

- **Pharmacies-** billing systems at the corporate level; central or regional level; local retail pharmacy

RTI INTERNATIONAL

8

Major sources of administrative data on medical expenditures

Big Data Platforms

- **MarketScan Data Warehouse** – US employer-based patient data; proprietary claims data. Based on a large convenience sample-primarily of large employers.

- **Health Care Cost Institute (HCCI) Data**-health care claims payment data contributed by large national insurers.

- **Optum Labs Data Warehouse (OLDW)-** administrative claims for over 100 million individuals; clinical data from EHRs

- **Electronic Health Records (EHRs)-** Includes administrative and billing data. As adoption grows, so does potential for enhancing the nation's health-care statistics capabilities.

Assessing the quality of administrative data on medical expenditures to enhance the analytic utility of survey data

Considerations to determine suitability for integration:

- Purpose/Content

- Definitions/Documentation/Standardization:

Charges/Payments/Costs

- Compatibility/Comparability

- Clarity

- Representativeness/Completeness

- Relevance

Assessing the quality of administrative data on medical expenditures to enhance the analytic utility of survey data

Considerations to determine suitability for integration:

- Accuracy

- Size/Precision

- Accessibility

- Timeliness

- Cost-efficient

RTI
INTERNATIONAL

Administrative Records Serving as Sampling Frames

Medicare enrollment files: The Medicare Current Beneficiary Survey (MCBS) conducted by the Centers for Medicare & Medicaid Services (CMS) is a continuous, multipurpose survey of a nationally representative sample of the Medicare population: *the sample is selected from Medicare enrollment files*

MCBS provides **expenditures and sources of payment for all services used by Medicare beneficiaries**, including co-payments, deductibles, and non-covered services; to ascertain all types of health insurance coverage and relate coverage to sources of payment

Data Integration: Administrative Records Enhancing Analytic Capacity of Surveys

The Medicare Current Beneficiary Survey (MCBS) Cost Supplement Files link the Medicare claims data to the survey data.

The accuracy of the MCBS expenditure data are enhanced through a **reconciliation process** that combines information from survey respondents and Medicare administrative files.

The process facilitates the production of a comprehensive picture of health services received by beneficiaries, amounts paid, and sources of payment.

Data Integration: Administrative Records Enhancing Analytic Capacity of Surveys

Medical Expenditure Panel Survey (MEPS)

Annual Survey of 14,000 households:
provides national and state estimates (most populous) of health care use, expenditures, insurance coverage, sources of payment, access to care and health care quality

Permits studies of:

- Distribution of expenditures and sources of payment

- Role of demographics, family structure, insurance

- Expenditures for specific conditions

- Trends over time

Sponsored by the Agency for Healthcare Research and Quality

MEPS Integrated Design

Household Component (HC)

Medical Provider Component (MPC)

Medical Organization Survey (MOS)-support provided by *the Robert Wood Johnson Foundation*

Insurance Component (IC)

- Longitudinal design

- Linkage to National Health Interview Survey

- **Linkage to CMS claims data**

- **Linkage to National Death Index**

- Data Supplementation at the state and county levels.

Medical Provider Component

Purpose

- Compensate for household item nonresponse

- Primary source for expenditure estimates

- Greater accuracy and detail

- Imputation source

- Supports methodological studies

Medical Provider Component

Targeted Sample

- All associated hospitals and sample of associated physicians
- Sample of associated office-based physicians
- All associated home health agencies
- All associated pharmacies

Data Collected

- Dates of visit
- Diagnosis and procedure codes
- Charges (except prescriptions) and payments

MEPS-MPC Sample Size

- 5,500 Hospitals

- 13,000 Office-Based Providers

- 600 Home Care Providers

- 100 Institutions

- 20,000 Separately Billing Doctors

- 8,000 Pharmacies

Source for event level expenditures

Household	Provider	MEPS value - Y_{ij}
Reported	reported	$Y_{ij} = Provider \$_{ij}$
Nonresponse	reported	$Y_{ij} = Provider \$_{ij}$
reported	nonresponse	$Y_{ij} = Household \$_{ij}$ [1]
nonresponse	nonresponse	$Y_{ij} = Imputed \$_{ij}$

[1]Recalibrated as necessary based on analyses of concordance between sources

RTI INTERNATIONAL

Uses of Administrative Data

Sample Frames

Permits efficient oversamples of rare populations to inform medical expenditure analyses: individuals with chronic conditions; inpatient stays; specific procedures; high utiliziers of the health care system.

Source to benchmark survey estimates

Issues of representativeness/completeness: uninsured population; individuals in managed care

Integration of Administrative and Survey Data: Opportunities and Challenges

Data Linkage:

- Use of secondary data (e.g. aggregate data at the county/state level) as core component of survey

- Informs sample design, nonresponse and poststratification adjustments, imputation and data supplement for item nonresponse

- Facilitates reductions in measurement error

Issues of access; need for greater attention to ensuring confidentiality: limitations in public use data

Acquisition of necessary permissions, participation.

Accuracy in matches: Type I/Type II error trade-offs

Integration of Administrative and Survey Data: Opportunities and Challenges

Signal Detection:
Large samples available in administrative data sources permit the identification of:

impending health problems/threats

hidden patterns

previously undetected relationships.

Research:
Stimulates more focused survey and research efforts to **assess the validity, intensity and persistence** of the signal in representative populations.

Enhancements in Disclosure Avoidance Methodology:

Development of improved statistical disclosure control (SDC) procedures that ensure greater levels of privacy protection, while controlling for "mosaic" effects.

Advances in methodologies to permit the public release of substantially greater levels of survey data content to meet growing demands for access from the external research community, while ensuring confidentiality provisions are satisfied.

Design integration of non-probability samples in national surveys-development of "hybrid" designs. Identification of the subset of the target population represented by the non-randomly selected supplemental samples.

Potential to facilitate composite estimation in national surveys, yielding greater precision in estimates at less cost.

Targeted applications for surveys that require oversampling of sparse or "rare" population subgroups.

Summary

- Integration of Administrative and Survey Data in Support of Medical Expenditure Analyses

- A model for enhanced analytic capacity and data quality

- Inherent challenges

- Opportunities for future innovations