

May 8th, 1:30 PM - 3:00 PM

## Big Data: Medical Claims Data for Translational Research: Medicaid

Bruce A. Barton

*University of Massachusetts Medical School*, [Bruce.Barton@umassmed.edu](mailto:Bruce.Barton@umassmed.edu)

Follow this and additional works at: [http://escholarship.umassmed.edu/cts\\_retreat](http://escholarship.umassmed.edu/cts_retreat)

 Part of the [Health and Medical Administration Commons](#), [Health Services Administration Commons](#), and the [Translational Medical Research Commons](#)



This work is licensed under a [Creative Commons Attribution-Noncommercial-Share Alike 3.0 License](#).

---

Bruce A. Barton, "Big Data: Medical Claims Data for Translational Research: Medicaid" (May 8, 2013). *UMass Center for Clinical and Translational Science Research Retreat*. Paper 4.

[http://escholarship.umassmed.edu/cts\\_retreat/2013/presentations/4](http://escholarship.umassmed.edu/cts_retreat/2013/presentations/4)

This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in UMass Center for Clinical and Translational Science Research Retreat by an authorized administrator of eScholarship@UMMS. For more information, please contact [Lisa.Palmer@umassmed.edu](mailto:Lisa.Palmer@umassmed.edu).

# **Big Data: Medical Claims Data for Translational Research**

## **Medicaid**

Bruce A. Barton, Ph.D.

Team Leader, Research Methods (Biostatistics)

Research and Evaluation Unit

Commonwealth Medicine

# Overview

- Description of Medicaid datasets
- Most commonly used variables on each datasets
- Examples of translational analyses using Medicaid

# Eligible Populations

- In Massachusetts, Medicaid and CHIP programs are combined under one “roof” (MassHealth)
- Groups with incomes below 133% of Federal Poverty Level
  - Adults (employed or not)
  - Children
  - Disabled adults
  - Pregnant women
  - HIV+ individuals
  - Dept of Mental Health clients
  - Senior citizens (also eligible for Medicare)
  - Women with breast/cervical cancer
  - People in need of long-term care

# Individual-level data

The available Medicaid data (MAX files) include:

- Personal summary
- Other therapy claims
- Inpatient claims
- Prescription claims
- Long-term care claims
- Provider characteristics
- Medicare-Medicaid linked enrollee
- Mini MAX

# Personal Summary

- Available from Fiscal Year 1999 to 2009 (~ 3 year lag)
- One record per person in system for at least one day
- Contents:
  - Demographics
  - Eligibility category and Zip code
  - Maintenance assistance status
  - Monthly enrollment status
  - Utilization summary
  - Open and close reasons
  - Program
  - Mortality information

# Other Therapy Claims Data

- Contents:
  - Physician services
  - Lab/radiology
  - Clinic/ambulatory care
  - Home health
  - Hospice
  - Durable medical equipment

# Inpatient Claims Data

- Contents:
  - Admission information
  - Diagnoses
  - Procedures
  - DRG (Diagnosis-Related Group)
  - Revenue codes
  - Discharge status
  - Length of stay

# Prescription Drug Claims Data

- Contents:
  - Date of prescription/fill
  - Provider
  - NDC code
  - New/refill
  - Days supply
  - Drug class/therapeutic class
  - OTC indicator

# Other Therapy Claims Data

- Contents:
  - Diagnoses
  - Procedures
  - Dates of start/end of service
  - Medicare payment information (for dual elig)
  - Provider

# Long-term Care Data File

- Nursing facilities
- Independent care facilities
- Independent psychiatric facilities
- Adult foster care
- Home health care
- Hospice

# Long-term Care Long-Term Data File

- Contents:
  - Date of admission and dates of service
  - Institution and type
  - Days by type of institution

# Mini-MAX

- 5% cross-sectional national sample of Medicaid data
- Summary of data for each patient in sample, including characteristics, diagnoses, and utilization
- Oversample small states and select subgroups to increase stability of estimates

# Types of Projects

- Utilization studies for disparities research:
  - Homeless
  - Recent immigrants
  - Disabled (physically/mentally)
- Effect of policy changes (state or federal) on utilization (and prescribing) patterns
- Utilization of different service types for low-income families

# Types of Projects (Cont.)

- Utilization of services for patients with particular diagnoses (e.g., diabetes, asthma, substance abuse)
- Cost of new vs. old treatments/medications
- Adverse events from treatments – Phase IV / surveillance studies?
- Utilization of cancer screening services – and outcomes

## Types of Projects (Cont.)

- Although many patients/individuals will come and go in the Medicaid program, some subgroups (such as the disabled, homeless, or those with substance abuse problems) are traceable across time – so that it is possible to follow longitudinal treatment patterns (and cost) across multiple years
- Have to be careful of biases, such as a cohort effect, but could provide beneficial information

# How to get started

- Contact Robin Clark or Wen-Chieh Lin at CHPR to discuss idea of project and how to frame it in the context of Medicaid data
- Contact RESDAC ([www.resdac.org](http://www.resdac.org)) about Medicaid data availability, cost, and requirements to receive data
- RESDAC (Research Data Assistance Center) is gateway to CMS data – forms, procedures, advice, costs

# Cautions

- After RESDAC acceptance (may require one or more revisions to data request), can take several months to receive data
- Data files can be quite large and would be best to utilize UMMS Regulatory Environment (Linux servers) to store/analyze data
- Three layers of privacy: research identifiable files, limited data sets, and public access files

➤ Thanks!!

➤ Questions?