Using HCCI commercial claims data for research (non-technical)

Jess Chang, Senior Researcher
Tanya Natwick, Senior Data Analyst

November 2022
Goals for attendees

• To learn about HCCI as an organization
• To understand basic claims data concepts
• To see how claims (specifically HCCI data) can be leveraged for research
• To know what you need to do to next if you want to work with HCCI’s data
What is HCCI?
HCCI is a mission-driven, non-profit at the nexus of data, analytics, and action.

We were founded in 2011 to leverage unique, high-quality data and extraordinary expertise in research and policy to give decisionmakers the tools they need to lower health care costs and increase value in the health care system.
Context and HCCI’s History

HCCI’s unique, multi-payer employer-sponsored insurance claims dataset can help inform public and private decisionmakers as they create policy.

HCCI was created to:

- Promote independent, nonpartisan research and analyses on the causes of the rise in U.S. health spending
- Inform the public policy process and assist in developing new solutions to long-term problems confronting the health care system
HCCI only licenses our ESI data for non-commercial purposes

- Academic Researchers at any college or university
- Government Agency Researchers
- Some research-focused non-profits

Commercial Purposes
Data cannot be used to generate profits
What are claims data?
Claims data are records detailing financial transactions between providers and insurers. Claims were designed for the administrative purposes of billing. However, they can also be used as a valuable source of data for research on health outcomes, health care use, and cost.
Claims data are financial transactions between providers & insurers

- Member ID
- Service dates, admission dates
- Diagnosis ICD and DRG codes
- Procedure codes (CPT/HCPCS)
- Drug codes (NDC) and days supply (Rx claims-specific)
- Revenue center, place of service, type of bill
- Provider ID, provider zip, provider specialty
- Allowed amounts and patient out-of-pocket costs
Other administrative data: enrollment file

- Each year during Open Enrollment period, enrollees chooses health plans for the next plan year
- There is no standardized enrollment form, but most insurers collect information as:
  - zip code of residence
  - Sex
  - Age
  - type of plan
  - Subscriber vs. dependent
  - Other coverage i.e. Rx, mental health, vision
Medicare FFS vs. Commercial Claims

**Medicare FFS**
- 1 payer
- 1 plan
- Very little churn
- No provider networks
- Uniform coverage
- 1 repository for billing guidelines

**Commercial/ESI**
- N payers (data provenance: many claims processing engines)
- N plans (POS, PPO, HDHP, HMO, Workers’ comp)
- Significant churn
- N provider networks
- Different benefits administrators
  - Medical, mental health, pharmacy
- Varying coverage
  - Federal and state regulations
  - Funding, employer sizes
  - Differential insurer guidelines
How can we use claims data to answer research question?
Claims data elements are akin to LEGO® pieces
Suppose you are interested in costs of childbirth... now what?

- Is it possible to identify in claims data?
  - Yes for inpatient and ? for home births
- What insurance segment would childbirth be found in?
  - No for Medicare FFS (>= 65 population)
  - Yes for ESI and Medicaid (< 65 population)
- Define costs
  - Facility or Facility + Professional
- Now we elected to look at **inpatient** childbirth
Pharmacy claims

Facility claims

Member Enrollment Records

Professional claims
What are our LEGO® pieces (claims data elements) required to construct data for study on costs of inpatient childbirth?

Start with **Facility Claims**
Find inpatient hospital admissions with start and end dates

How to identify inpatient stays for **childbirth**?
Diagnosis, procedure, and/or DRG codes

Consider **Professional Claims** for full picture of childbirth costs
Anesthesiologists (maybe OON) may submit separate claim outside of facility claims

Add geographic or age stratification from **Member Enrollment** records.
Demographic data not found in claims transactions.

Source: https://www.lego.com/en-us/service/buildinginstructions/60204
Claims data are a rich source of health care information

- Retrospective, observational, real-world data
- Real transactions, real people.
  - Unlike survey data, no recall bias.
- Statistically de-identified data maximizes both research utility and privacy
What makes HCCI data so great?
HCCI’s Commercial Claims Dataset

- More than half of people in the US get health insurance through an employer (ESI = 54%)
- 55 Million Covered Lives Annually
- 30+ Plans
- 1 Billion Claims Per Year
- More than 30% ESI Population

- Claims From 2012 to 2020
- New data added every year!
- Includes All 50 States and DC
- De-Identified
- HIPAA and Anti-Trust Compliant
What studies can HCCI claims data support?*

<table>
<thead>
<tr>
<th>Utilization and spending over time</th>
<th>Impact of health policies</th>
<th>Prices and Markets</th>
</tr>
</thead>
<tbody>
<tr>
<td>• HCCUR</td>
<td>• Rich geographic spread (urban vs. rural)</td>
<td>• HMI</td>
</tr>
<tr>
<td>• Real world evidence on treatments and disease</td>
<td>• Relaxation of state’s scope of practice with telemedicine</td>
<td>• Private equity acquisitions</td>
</tr>
<tr>
<td>• Geographic variation</td>
<td>• Physician practice changes due to FDA black box warnings</td>
<td>• Traditional IO microeconomics studies such as market competitiveness</td>
</tr>
<tr>
<td>• Comparative effectiveness of medications and treatments</td>
<td>• State laws on opioid prescription limits</td>
<td></td>
</tr>
<tr>
<td>• Clinical guidelines in practice</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Not an exhaustive list
Limitations and Considerations

- Claims data does NOT contain clinical measurements, lab results, EHR data elements, or services not reimbursed by health insurers
- Real world observational data requires lots of cleaning
- Cannot track enrollees over payers over time (a switch in health insurers)
- Health equity implications (better data stewards for impactful research)
  - No race/ethnicity data or gender identity or protected class information
- HCCI data is an ESI population
  - Some questions may not be relevant to ESI population (end-of-life care would mostly involve Medicare)
  - Patterns in use/spending may be different than uninsured/Medicaid/Medicare
  - Different insurers have different plans, coverage, provider networks, etc. Not as homogenous as Medicare FFS.
How to take the next step
How to ensure a successful research project using commercial claims data

- Defining a clear research question
  - Health care system institutional details matter!
- Reach out to HCCI staff!
- Dataset assembly will make up over 90% of project
  - Requires careful planning
  - Practical data decisions during course of project
- **BIG DATA!** Billions and billions of records
  - Find graduate assistants or staff with appropriate skill set such as strong technical database background
- **Best language to wrangle big data is SQL**
- **Stata is not your friend to wrangle big data into analytical dataset**
# Pricing for Access - Research Team

<table>
<thead>
<tr>
<th>Per Project Annual Access Fee</th>
<th>Add-On: Extra Seat Secure Data Enclave</th>
<th>Add-on: Encrypted NPI-Level Dataset Merge</th>
</tr>
</thead>
<tbody>
<tr>
<td>$45,000</td>
<td>$10,000</td>
<td>$3,000</td>
</tr>
</tbody>
</table>

Annual access fee includes two seats with access to the full HCCI Commercial Claims Dataset, along with licenses for SAS, Stata, Python, and other statistical software.
What to Expect

From application to data access usually takes a minimum of 8 weeks.

Data Licensing Agreement usually takes the most time
• We provide a standard agreement; Approval is faster when changes are minimal
• If your organization has licensed HCCI data before, suggest to your legal office that they consult the previous agreement

The application requires a well-defined research question and scope and a concise but specific research approach/methodology.
• We need enough detail to confirm that you will be able to answer your research question with our data.

Contact us with questions along the way – we’re here to help!
Available Resources and Updates

- Data dictionary and additional resources available on HCCI’s website
- Sign up for HCCI’s newsletter to receive periodic updates
- Check out HCCI’s public use files, briefs, and research studies

HCCI Website:
www.healthcostinstitute.org

Email Us:
data@healthcostinstitute.org
Thank you!

Questions?