



HEALTH CARE
COST INSTITUTE

Using HCCL 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 HCCL's History

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

HCCL 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



HEALTH INSURANCE CLAIM FORM

APPROVED BY NATIONAL UNIFORM CLAIM COMMITTEE (NUCC) 02/12

PICA

1. MEDICARE MEDICAID TRICARE CHAMPVA GROUP HEALTH PLAN FECA BLK/LUNG OTHER
(Medicare) (Medicaid) (ID#DoD) (Member ID) (ID#) (ID#) (ID#)

2. PATIENT'S NAME (Last Name, First Name, Middle Initial)

3. PATIENT'S BIRTH DATE MM DD YY SEX M F

4. INSURED'S NAME

5. PATIENT'S ADDRESS (No. - Street)

6. PATIENT RELATIONSHIP TO INSURED
Self Spouse Child Other

7. INSURED'S ADDRESS

CITY STATE CITY

ZIP CODE TELEPHONE (Include Area Code) ZIP CODE

9. OTHER INSURED'S NAME (Last Name, First Name, Middle Initial)

10. IS PATIENT'S CONDITION RELATED TO:
a. OTHER INSURED'S POLICY OR GROUP NUMBER
b. RESERVED FOR NUCC USE
c. RESERVED FOR NUCC USE

11. INSURED'S POLICY OR GROUP NUMBER
a. EMPLOYMENT? (Current or Previous)
b. AUTO ACCIDENT? PLACE (State)
c. OTHER ACCIDENT?

12. PATIENT'S SIGNATURE
13. INSURED'S OR PAYOR'S SIGNATURE

14. DATE OF CLAIM MM DD

15. DATE OF SERVICE FROM TO

16. DATES PATIENT FROM TO

17. NAME OF REFERRING PROVIDER OR OTHER SOURCE
17a. NAME
17b. NPI

18. HOSPITALIZATION FROM TO

19. ADDITIONAL CLAIM INFORMATION (Designated by NUCC)

20. OUTSIDE LAB? YES NO

21. DIAGNOSIS OR NATURE OF ILLNESS OR INJURY (Relate A-L to service line below (24E)) ICD Ind

22. RESUBMISSION CODE

23. PRIOR AUTHORITY

24. A. DATE(S) OF SERVICE FROM TO
B. PLACE OF SERVICE
C. PLACE OF SERVICE
D. PROCEDURES, SERVICES, OR SUPPLIES (Explain Unusual Circumstances)
E. DIAGNOSIS POINTER
F. CHARGES

25. FEDERAL TAX I.D. NUMBER SSN/EIN
26. PATIENT'S ACCOUNT NO.
27. ACCEPT ASSIGNMENT? YES NO
28. TOTAL CHARGE \$

31. SIGNATURE OF PHYSICIAN OR SUPPLIER INCLUDING DEGREE OR CREDENTIALS (I certify that the statements on the reverse apply to this bill and are made a part thereof.)
32. SERVICE FACILITY LOCATION INFORMATION
33. BILLING PROVIDER

SIGNED DATE

NUCC Instruction Manual available at: www.nucc.org PLEASE PRINT OR TYPE APPF

**CMS-1500
(Professional)**

INPATIENT

1. Any Hospital 123 Any Street Anytown NJ 08999
2. Any Hospital 456 Any Street Anytown NJ 08999
3. PAT # 1234
4. TYPE OF BILL 0111
5. FED. TAX NO. 98765
6. STATEMENT PERIOD FROM 11 03 06 TO 11 04 06
7. RESERVED

8. PATIENT NAME Doe, John
9. PATIENT ADDRESS 1234 Main Street Anytown NJ 08999

10. BIRTH DATE 03 20 1971 SEX M ADMISSION DATE 11 03 06 TYPE 08 3 3 3
11. CONDITION CODES 12 01
12. IDENTIFYING EVENTS PA RESERVED

13. OCCURRENCE DATE CODE OCCURRENCE DATE CODE OCCURRENCE DATE CODE OCCURRENCE DATE CODE OCCURRENCE DATE CODE OCCURRENCE DATE CODE
14. Occurrence and Occurrence Span Codes may be used to define a significant event that may affect payer processing

15. VALUE CODES AMOUNT VALUE CODES AMOUNT VALUE CODES AMOUNT
A1 952.00
16. Value Codes and amounts required when necessary to process claim

42 REV. CD.	43 DESCRIPTION	44 HCPCS / RATE / HPPS CODE	45 SE RV DATE	46 SE RV UNITS	47 TOTAL CHARGES	48 NON-COVERED CHARGES	49
0129	Semi-Private			2	400.00	0.00	Future Use
0250	Pharmacy			1	50.00	0.00	
0360	OR Services				100.00	0.00	

50. PAYER NAME AmeriHealth
51. HEALTH PLAN ID Report HIPAA National Health Plan Identifier when mandatory
52. PRIOR PAYMENTS Y Y
53. EST. AMOUNT DUE 222222222
54. OTHER PER V ID 1234567890
55. INSURANCE GROUP NO. 1234

56. PATIENT NAME Doe, John
57. INSURED'S UNI-QUE ID ABC1234567800
58. GROUP NAME Watch Repair, Inc.
59. EMPLOYER NAME Watch Repair, Inc.

60. TREATMENT AUTHORIZATION CODES
61. DOCUMENT CONTROL NUMBER 491234
62. EMPLOYER NAME Watch Repair, Inc.

63. ADMIT DATE 4280
64. PATIENT REASON FOR DRG
65. MAY BE USED TO REPORT EXTERNAL CAUSE OF INJURY Reserved
66. PRINCIPAL PROCEDURE CODE DATE 3749 11 03 06
67. OTHER PROCEDURE CODE DATE
68. ATTENDING NPI 222222222 QUAL 62 1234569822
69. LAST Smith FIRST David

70. REMARKS
71. B3 282N00000X
72. SECONDARY
73. TERTIARY

UB-04 (Facility)



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

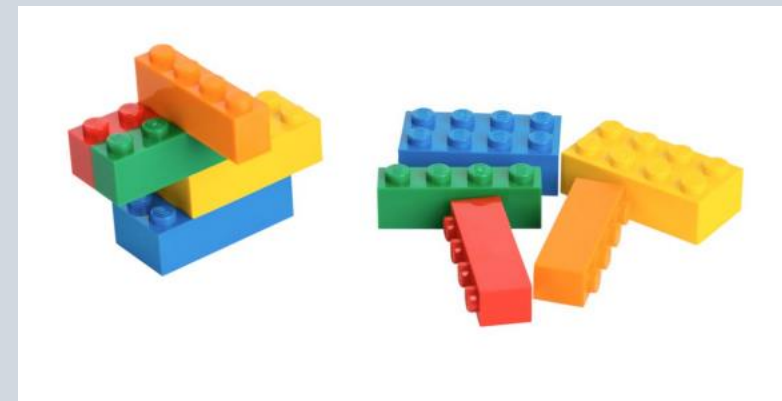


Facility claims

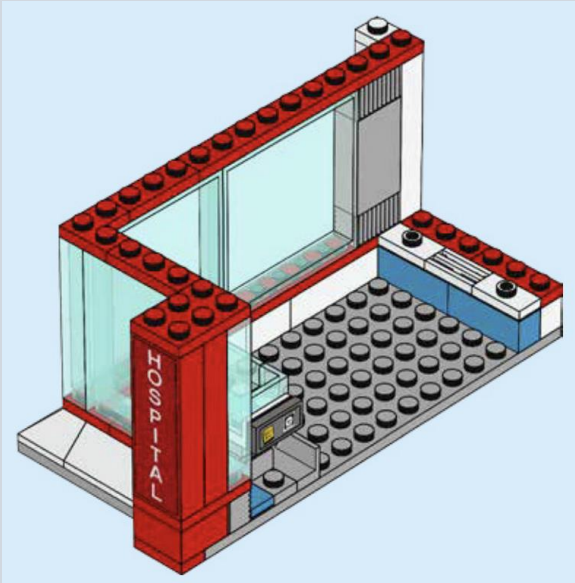


Member Enrollment Records

Professional claims

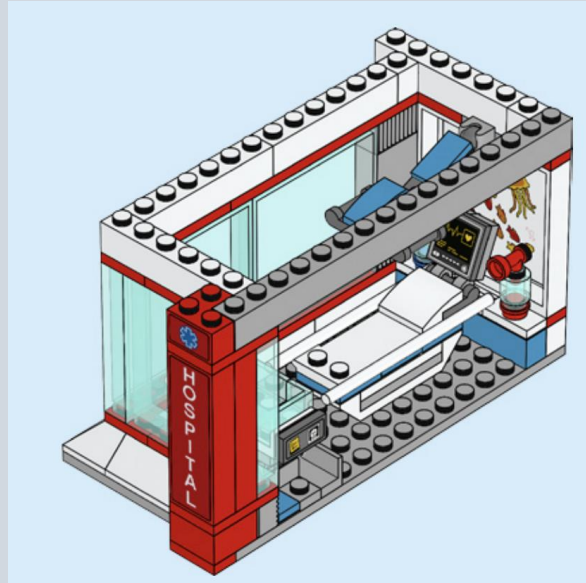


What are our LEGO® pieces (claims data elements) required to construct data for study on **costs of inpatient childbirth**?



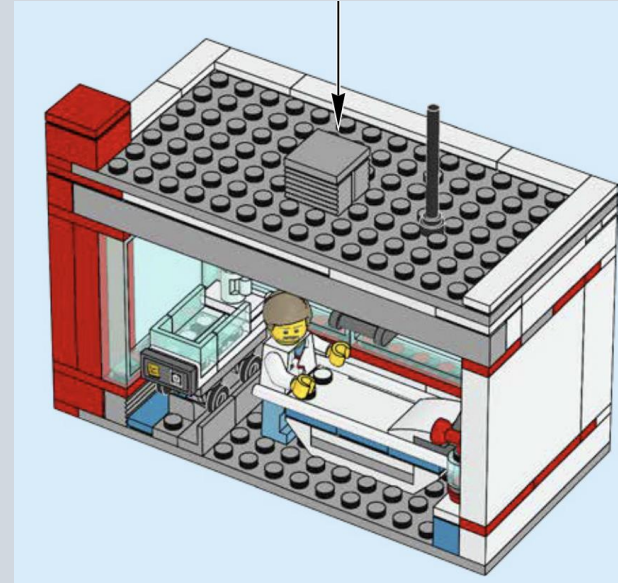
Start with **Facility Claims**

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

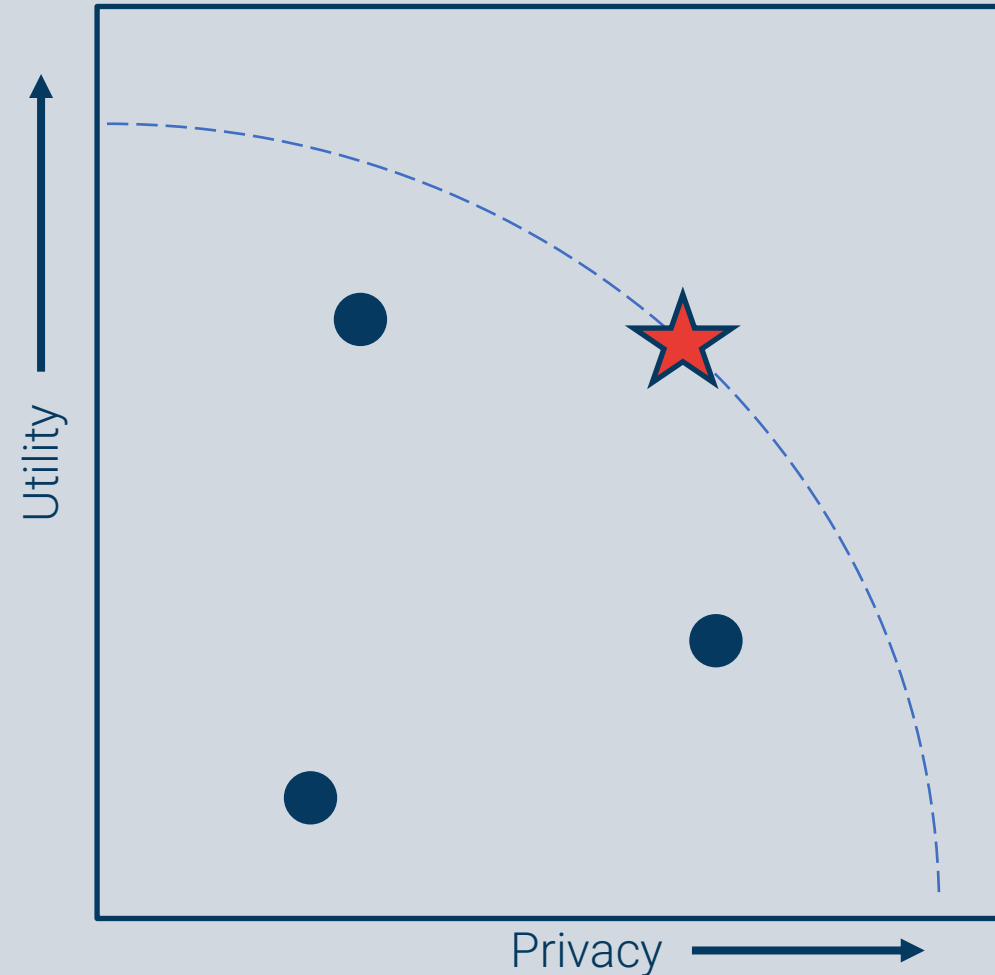


Add geographic or age stratification from **Member Enrollment** records.

Demographic data not found in claims transactions.

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

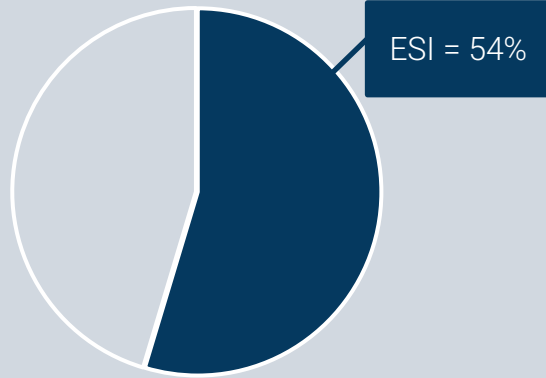


What makes HCCI data so great?



HCCI's Commercial Claims Dataset

More than half of people in the US get health insurance through an employer



Claims From 2012 to 2020



New data added every year!

De-Identified



HIPAA and Anti-Trust Compliant

55 Million Covered Lives Annually



30+ Plans

1 Billion Claims Per Year

More than 30% ESI Population

Includes All 50 States and DC



What studies can HCCI claims data support?*

Utilization and spending over time

- **HCCUR**
- Real world evidence on treatments and disease
- Geographic variation
- Comparative effectiveness of medications and treatments
- Clinical guidelines in practice

Impact of health policies

- Rich geographic spread (urban vs. rural)
- Relaxation of state's scope of practice with telemedicine
- Physician practice changes due to FDA black box warnings
- State laws on opioid prescription limits

Prices and Markets

- **HMI**
- Private equity acquisitions
- Traditional IO microeconomics studies such as market competitiveness

*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

**Per Project Annual
Access Fee**

\$45,000

**Add-On: Extra Seat
Secure Data Enclave**

\$10,000

**Add-on: Encrypted NPI-
Level Dataset Merge**

\$3,000

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

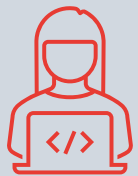


From application to data access usually takes a minimum of 8 weeks.



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?

