Data Validation

VERMONT
GREEN MOUNTAIN CARE BOARD

Summary

The GMCB is undertaking a comprehensive data verification project in tandem with health care stakeholders to validate Vermont's all-payer claims database (VHCURES), beginning with quantifying the difference between hospital discharges captured in VHCURES and Vermont's hospital discharge data system (VUHDDS). The overarching goal of this work is to improve and enhance the data under GMCB's stewardship. This work will help us to understand the difference between these data resources and those maintained by the payers submitting data to VHCURES as well as the hospitals submitting data to VUHDDS, and will include validation of the financial data available in VHCURES to determine how closely it aligns with the records of the payers submitting data to the system, as well as the providers receiving these payments.

Green Mountain Care Board

The purpose of the Green Mountain Care Board is to promote the general good of the State by:

1. Improving the health of the population;
2. Reducing the per-capita rate of growth in expenditures for health services in Vermont across all payers while ensuring that access to care and quality of care are not compromised;
3. Enhancing the patient and health care professional experience of care;
4. Recruiting and retaining high-quality health care professionals; and
5. Achieving administrative simplification in health care financing and delivery.

18 V.S.A. § 9372

Background

All-payer claims databases (ACPDs) are among the richest sources of health care data available to states.[1] They aggregate medical, pharmacy, and dental claims from major insurers in a single location. They also link claims[2] to other types of health care data[3] so that analysts can investigate trends with statistical precision.

Vermont's APCD is the Vermont Health Care Uniform Reporting and Evaluation System (VHCURES).[4] The Green Mountain Care Board (GMCB) has been responsible for managing VHCURES since 2013. The GMCB contracts with Onpoint Health Data, a nonprofit data services organization based in Maine, to collect claims data directly from public and private health insurers. Onpoint cleans, consolidates, and releases the data to the GMCB in a database format. The result is among the largest and most complex databases managed by the State of Vermont, consisting of billions of claims that each span hundreds of columns of data.

Data Reporting Complexities

APCDs like VHCURES are massive undertakings that require multiple stakeholders to work together. Insurer claims depend on accurate information from those who provide care. Clinicians and patients themselves can make small reporting errors as they supply information and transcribe physical forms into their electronic record systems. Altogether, these issues can lead to differences between the database and the truth on the

[1] Porter, Jo, Denise Love, Ashley Peters, Jane Sachs, and Amy Costello. "The Basics of All-Payer Claims Databases: A Primer for States." Robert Wood Johnson Foundation, January 2014. http://www.rwjf.org/content/dam/farm/reports/issue_briefs/2014/rwjf409988.

[2] Claims data may include information about dates of care, diagnoses, procedures, what insurers pay, what patients pay out of pocket, and whatever else an insurer records as it processes claims.

[3] These tables may include details about health care providers, patient/policyholders, insurers and their policies, and other components of health care

[4] Green Mountain Care Board. "VHCURES Overview: A Guide for Data Users," January 2019. https://gmcboard.vermont.gov/sites/gmcb/files/documents/VHCURES%20Overview%20Jan2019%20FINAL.pdf.

ground. How do these differences impact APCDs? The federal Agency for Health Care Research and Quality[5] published a high-level evaluation of three states' database quality in 2017.[6] [7] It found that the variables used to calculate cost (initial price, paid price, insurance type, copays, coinsurance, out of pocket costs) were seldom missing from APCDs. However, the Agency identified several areas of concerns that pertain to one or more of the APCDs:

1. Single individuals may have more than one unique identifier, complicating linkage of claims.
2. Inpatient admission date, admission type, and discharge date variables were reported at variable rates, suggesting "use of fields for non-inpatient claims or missing values."
3. Between 34% and 47% of claims lacked secondary diagnosis codes. Outpatient claims lacked secondary diagnoses at greater rates than inpatient claims. A lack of secondary diagnosis codes impacts the usefulness of APCD data "for adequate risk adjustment without linking across multiple encounter and pharmacy records to establish comorbidities for a patient."
4. There was high variability in the inclusion of procedure or service codes, particularly secondary codes. Outpatient encounters almost always included a code for an office visit as, or in addition to, the primary procedure or service code.
5. It was difficult to attribute all encounters to a health care provider. This is because providers may use multiple Medicaid or National Provider Identification numbers. Providers may also provide care across different locations "based on administrative and logistic features of their practice" There were often missing or incorrect values about the geographic location of care.
6. Data did not provide definitive markers for either assigned or utilized primary care providers, medical homes, or other constructs.
7. The sampled APCDs lacked consistent race or ethnicity data.

Data Quality Improvement as Standard Practice

It is standard practice for APCD stewards or their contractors to verify the quality of new data after it has been delivered. Such reviews are primarily concerned with the structure of the data, such as whether the new data fit the rows and columns of the database, and whether data are missing at rates beyond established parameters.[8] [9] [10] By contrast, these reviews do not verify whether the data match the delivery of health care on the ground. Such reviews are difficult to scale, as they must consider the specific circumstances of each claim. According to the Minnesota Department of Health,

> "[MDH] does not attempt to verify accuracy of medical coding by providers or their clinic staff or assess accurate recording of diagnoses during the encounter; ***those activities typically require***

[5] The Agency of Health Care Research and Quality is part of the US Department of Health and Human Services.
[6] Data Evaluation of All-Payer Claims Databases. Content last reviewed April 2017. Agency for Healthcare Research and Quality, Rockville, MD. https://www.ahrq.gov/data/apcd/backgroundrpt/data.html
[7] The Agency does not name the three states in its report, in accordance with the project-specific Data Use Agreement for each APCD.
[8] Center for Improving Value in Health Care. "Colorado APCD Data Quality Process: Intake, Processing, Release," n.d. https://www.civhc.org/wp-content/uploads/2018/09/CIVHC-Data-Quality-Process_FINAL9.2018.pdf.
[9] Virginia Health Information. "Overview of the VA APCD Data Validation Process," n.d. https://www.nahdo.org/sites/default/files/Data%20Quality%20Forum%20VHI%20%20Presentation%206.13.18.pdf.
[10] Arkansas Ceter for Health Improvement. "The Arkansas All-Payer Claims Database (APCD) Data Validation Reporting," December 2017. https://www.arkansasapcd.net/Docs/205/.

complex chart review. [emphasis added] MDH's data aggregation vendor does, however, ensure that submitted diagnosis codes are valid and populated to the state-mandated thresholds."[11]

Such reviews are also difficult to scale because they require input from health care insurers, who have the raw claims data, and health care providers, who have patient records. Even then, the high volume of claims per year compels stakeholders to compare their records at a summary scale. For example, the State of Washington gave providers an opportunity to verify median *all-payer allowed amounts* for select inpatient and outpatient procedures prior to the publication of a price transparency dashboard.[12] [13]

GMCB's Enhanced Data Validation Project

GMCB's project to validate VHCURES will begin with an exercise to quantify the difference between hospital discharges captured VHCURES and Vermont's hospital discharge data system (VUHDDS). This will help us to understand the difference between these data resources and those maintained by the payers submitting data to VHCURES as well as the hospitals submitting data to VUHDDS. The project will also include validation of the financial data available in VHCURES to determine how closely it aligns with the records of the payers submitting data to the system, as well as the providers receiving these payments. The work will account for how billing, payments, and administrative data may differ depending on the type of provider (e.g. hospital, FQHC, independent practice). The project will culminate in a report documenting the process and providing guidance for analysts using the data to improve results, including recommendations to account for measurement error. The work is also the first step in the GMCB's effort to produce a price variation report, which was identified as a priority by Board members. Successful validation will depend on stakeholders who are willing to provide comparative information to attempt to quantify the differences between data systems. With this invitation we are including GMCB's analytic vendors, Vermont hospital systems association, payers, hospital systems, DVHA, FQHCs, and independent clinicians.

The project objectives are to:

- Develop a quality improvement guidance document for analysts who use GMCB's data sets, including recommendations to account for measurement error in financial data.
- Produce reports to inform GMCB decision-making (e.g. price variation).
- Develop tools or reports to allow ongoing validation for those submitting data and establish goals for improving data quality.
- Determine priorities and feasibility of enriching GMCB's data assets by linking it to other data (e.g. HIE, vital statistics).

Key Dates & Timeline

On July 28, 2020 GMCB held the first meeting of interested stakeholders to launch this project. By the end of 2020 GMCB will produce a document outlining findings from initial data validation work and will present a progress report to the legislature during the 2021 session. By the end of 2022, GMCB expects to publish a price variation dashboard which will help the public to better understand prices in Vermont.

Last Updated: August 2020

---

[11] Minnesota Department of Health. "Minnesota's All-Payer Claims Database Frequently Asked Questions," February 2015. https://www.health.state.mn.us/data/apcd/docs/faqapcd.pdf.
[12] Washington State Hospital Association. "Verification of State's Data in All Payer Claims Database." December 18, 2017. https://www.wsha.org/articles/verification-states-data-payer-claims-database/.
[13] All-payer allowed amounts are average price ceilings, not the actual price paid by different payers.