ASPE REPORT

Summary Report
Claims and Provider Payment Data Gaps for Responding to COVID-19

Prepared for
The Office of the Assistant Secretary for Planning and Evaluation (ASPE)
at the U.S. Department of Health and Human Services

By
NORC at the University of Chicago

September 2022

Office of the Assistant Secretary for Planning and Evaluation

The Assistant Secretary for Planning and Evaluation (ASPE) advises the Secretary of the U.S. Department of Health and Human Services (HHS) on policy development in health, disability, human services, data, and science and provides advice and analysis on economic policy. ASPE leads special initiatives; coordinates the Department's evaluation, research, and demonstration activities; and manages cross-Department planning activities such as strategic planning, legislative planning, and review of regulations. Integral to this role, ASPE conducts research and evaluation studies; develops policy analyses; and estimates the cost and benefits of policy alternatives under consideration by the Department or Congress.

The Office of Science and Data Policy

The Office of Science and Data Policy is the departmental focal point for policy research, analysis, evaluation, and coordination of department-wide public health science policy and data policy activities and issues. The Office provides authoritative advice and analytical support to the ASPE and departmental leadership on public health science policy and data policy issues and initiatives, coordinates science and data policy issues of interagency scope within HHS, and manages interagency initiatives in science policy and data policy. The Office works closely with staff from across the Department on strategic plan development and implementation efforts. The Offices also carries out a program of policy research, analysis, evaluation, and data development in these issues.

The Office of Health Policy

The Office of Health Policy (HP) provides a cross-cutting policy perspective that bridges departmental programs, public- and private-sector activities, and the research community to develop, analyze, coordinate, and provide leadership on health policy issues for the Secretary.

This report was produced by NORC at the University of Chicago under Contract No. HHSP233201500048I, Task Order No. 75P00121F37021 for the Office of Science and Data Policy.

Suggested Citation

Dunville, R., Grenen, E., Cotter, M., Hunt, M., Koltun, S., MacLean, K., Montazer, J., Ramirez, E., and Caro, V. (2022). *Final Summary Report: Claims and Provider Payment Data Gaps for Responding to COVID-19*. Report for the Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services. September 2022.

CONTRIBUTING AUTHORS

Richard Dunville, MPH, NORC at the University of Chicago
Emily Grenen, MSc, NORC at the University of Chicago
Megan Cotter, MPH, NORC at the University of Chicago
Meaghan Hunt, BA, NORC at the University of Chicago
Sarah Koltun, MPH, NORC at the University of Chicago
Kai MacLean, BS, NORC at the University of Chicago
Jaleh Montazer, MPH, NORC at the University of Chicago
Elizabeth Ramirez, BS, NORC at the University of Chicago
Valentina Caro, BA, NORC at the University of Chicago

Violanda Grigorescu, MD, MSPH, ASPE
Steven Sheingold, PhD, ASPE
Rachael Zuckerman, PhD, ASPE

PROJECT OFFICERS AND PROJECT LEADERSHIP
Sonal Parasrampuria, PhD, ASPE
Oluwarantimi Adetunji, PhD, ASPE
Rachael Zuckerman, PhD, ASPE

Table of Contents

Introduction

Addressing the COVID-19 public health emergency (PHE) required the United States Department of Health and Human Services (HHS) to act promptly using timely data illustrating COVID-19 risk factors, cases, hospitalizations, deaths, and SARS-CoV-2 vaccination coverage at the population level. In addition, HHS had to use different data sources to monitor available ventilators, hospital staffing levels, personal protective equipment, and distribution of provider relief funds.

Throughout the COVID-19 PHE, many internal HHS projects were conducted primarily using Medicare fee-for-service (FFS) claims data that was readily available within HHS. Medicare FFS claims data can be accessed in nearly real time (usually, approximately one month after the service was rendered) and at the beneficiary level. These claims data also can be linked across time and health service providers and can be aggregated to many geographic and provider levels. Further, FFS claims include relatively complete information on beneficiary address, race/ethnicity, dual enrollment in Medicare and Medicaid (which may be used as a proxy for socioeconomic status), age, gender, spending, diagnoses, and comorbidities. These attributes led HHS staff to consider Medicare FFS data as a "gold standard" for claims.

In comparison, there was a dearth of timely data sources available to HHS from commercial insurance payers, including payments from payers to providers. When commercial claims data were available, their utility was further challenged by incomplete records and lack of standardization across multiple payers that hindered the ability to conduct studies for a holistic view of the commercially insured population.

To this end, the Office of the Assistant Secretary for Planning and Evaluation's (ASPE) Office of Science and Data Policy, in collaboration with its Office of Health Policy, contracted NORC at the University of Chicago (NORC) to conduct a study examining the availability and use of other, non-federal sources of claims and provider payment data during the COVID-19 PHE.

Goal

The goal of the project was to identify gaps and challenges in accessing and using claims data and to delineate strategies to improve data collection and availability for researchers and policymakers in the case of future PHEs.

Methods

A three-pronged approach was used to achieve the project goal:

(1) an environmental scan (e-scan) of commercial claims data availability and use *before* and *during* the COVID-19 PHE;

(2) key informant interviews (KIIs) with stakeholders who could speak to unmet data needs and new data made available *during* the COVID-19 PHE; and

(3) a technical expert panel (TEP) to describe gaps and limitations of datasets and strategies for addressing these limitations.

This report provides an overview of project findings, including suggestions to address gaps identified during the COVID-19 PHE. In addition, examples of next steps that may improve availability and use of claims data are highlighted.

Summary of Cross-Cutting Findings

Following is a summary of findings that cut across the three activities implemented for this project. Icons denote which research methods corroborate each finding (Table 1). Additional discussion of these findings is included in the appendices.

Table 1. Cross-Cutting Findings

Key		
Key Informant Interviews	Environmental Scan	Technical Expert Panel
Benefits of health care claims data for research		
Environmental Scan, Technical Expert Panel	**Large sample size** of insured individuals was the most commonly cited benefit of claims databases—particularly commercial claims databases that include data from multiple payers and health systems and all-payer claims databases (APCDs) that aggregate both public and private claims data.	
Environmental Scan, Technical Expert Panel	Claims data often enable researchers to **track individuals longitudinally**—this is especially true of Medicare data, which has a stable population that is easy to track over time.	
Health care claims data use for COVID-19 PHE-related research		
Key Informant Interviews, Environmental Scan	During the COVID-19 PHE, claims data were frequently used to explore **changes in health care utilization.** For example, researchers used claims to study telehealth, and their findings noted increased telehealth use during the COVID-19 PHE relative to before.	
Key Informant Interviews	Timely data were needed to monitor and report in real time the impact on service utilization and health outcomes during the COVID-19 PHE. The **lag in time** from when the services are provided to accessing and using the claims for research (i.e., adjudication processes that may take six months or more) was an existing issue that became more visible during the fast-evolving COVID-19 PHE. There were **no significant changes** to the types of data made available during the COVID-19 PHE for most researchers. Medicare claims were made available more quickly to internal HHS researchers.	

Government versus commercial claims data sources

	Medicare data face similar issues of **limited generalizability** as commercial claims data sources. First, Medicare data only cover a specific population of adults aged 65 and older and a subset of people with disabilities, and Medicare FFS claims exclude the large and growing proportion Medicare Advantage enrollees. Similarly, APCDs have limited generalizability because self-insured employer health plans cannot be mandated to contribute their claims data to APCDs. For entities outside the federal government, **access to Medicare data for research is often delayed**. Medicare data are especially valuable for **tracking individuals over time** because they include a stable population with continuous coverage that makes changes far less frequently than the commercially insured or Medicaid populations. *Note: APCDs were created to address this issue and have a consolidated approach to accessing and using both government and nongovernment claims; however, APCDs have their own challenges and limitations. More information on APCDs can be found in other ASPE reports.[1]*
	Insurers are not forthcoming about their own data availability and limitations. Insurers were also least forthcoming among the study participants—both in willingness to be interviewed and to make data available to external entities.

Limitations in using health care claims data for research *Limitations are denoted in **bold blue font.***

	Health care claims data may **lack relevant health and demographic data**; some data are not collected in claims, and other data are captured inconsistently. For example, some researchers mentioned that claims data lack important clinical information, which they need to obtain through other sources, such as hospital discharge and hospital observation data. In some cases, relevant data are available but proprietary, and confidentiality issues require levels of aggregation that are less useful for many studies. *Note that data availability varies depending on the database, therefore, some researchers reported having access to certain types of data that other researchers reported were unavailable.* Health care claims **data are not collected and/or intended to be used for public health surveillance and research** but rather for financial purposes. As a result, some clinical data are not typically collected in claims, and claims data are not regularly linked to clinical data sources, such as electronic health record (EHR) data. Additionally, **some services or diagnostic information are not reported as claims**. For example, COVID-19 tests and vaccines are often obtained outside of traditional health care settings and are therefore not consistently submitted as claims. Finally, due to a **lack of transparency among insurers and commercial claims database owners** (e.g., not publicizing codebooks online), researchers may simply not have full knowledge of what is and is not available.

---

[1] Carman K.G., Dworsky M., Heins S., Schwam D., Shelton S., & Whaley C. (2021). *The History, Promise and Challenges of State All Payer Claims Databases: Background Memo for the State All Payer Claims Database Advisory Committee to the Department of Labor.* RAND Health Care (Document Number: PR-A1396-1). https://aspe.hhs.gov/sites/default/files/migrated_legacy_files//200696/apcd-background-report.pdf

👥		🗨	Health care claims data are subject to **time lags** between date of service and when data are available for research. In addition, researchers cited the amount of time needed to issue and adopt new COVID-19 International Classification of Diseases (ICD) diagnostic codes.
👥	📄	🗨	As with any data source, health care claims data are subject to **coding discrepancies or inaccuracies**. For example, inconsistent methods of classifying provider type data or inaccuracies in classifying facility type can make claims data difficult to analyze.
👥		🗨	**Urgent needs and competing priorities can lead to a lack of time and shortage of staff** (i.e., expertise), particularly in the public sector, to clean and analyze health care claims data.

Potential solutions for addressing health care claims data limitations
*Limitations are denoted in **bold blue font.***

👥	📄	🗨	**Lack of relevant health and demographic data** may be addressed by encouraging **data sharing** among insurers, health care claims database owners, and other types of databases, such as EHRs or vaccine registries. Combining other data sources with claims data can help answer research questions.
👥	📄	🗨	**Lack of relevant health and demographic data** and **lack of representativeness and generalizability** may be addressed through **data standardization methods** to enable interoperability among data sources. For example, implementing new or expanded common data specifications for claims submitters or identifying and disseminating ways to layer common data models (e.g., Observational Medical Outcomes Partnership (OMOP), Fast Healthcare Interoperability Resources (FHIR)) onto existing data collection streams. Additionally, improve accessibility of standardized, linked data by developing a central repository with linked data to pull from during the next PHE.
👥	📄	🗨	**Lack of relevant health and demographic data** and **lack of representativeness and generalizability** may be addressed through **incentivized and/or mandated data standardization** for all data owners, including government and nongovernment entities, including private commercial insurers. Researchers specifically mentioned the need to create a standard format that makes integrating Medicare data into state APCDs easier. However, it is important to consider the time and money needed to change standards.
	📄	🗨	**Lack of relevant health and demographic data** and **lack of representativeness and generalizability** may be addressed through targeted **primary data collection (e.g., surveys) to reach populations that may not be included in secondary claims data.** In addition, it is important **to encourage researchers and the broader public health workforce to use and become familiar with claims data before the next PHE**. One approach could be developing a rapid contracting mechanism to expedite the connection between researchers and data providers to ensure more researchers can gain experience working with data.

	Some limitations may be acceptable to researchers. Instead of creating the "perfect" dataset at the outset, anticipate what research questions will need to be answered in future PHEs and then determine which limitations must be addressed, versus which are acceptable. For example, it may be unnecessary to have a fully representative population dataset for a PHE that primarily affects a subset of the population (e.g., Zika virus).

Discussion

Following are **key takeaways** from the project findings:

✓ **Lack of transparency about claims data availability and limitations—particularly among private insurers and commercial claims database providers.** Insurers were least responsive to interview requests; some cited that their organizations prohibited them from speaking to government agencies. In addition, those who participated in interviews were often reluctant to answer all questions, especially those about limitations of their databases. Additionally, there was limited information on public-facing websites regarding claims database pricing, timeliness, and included data elements.

✓ **Claims data alone had limited research utility for the COVID-19 PHE.** Researchers noted that answering questions that arose during the COVID-19 PHE required using claims data along with other data sources, such as EHR data and state vaccine registries. During the COVID-19 PHE, claims data were missing some essential elements, including many vaccination and testing services and their pricing data. Additionally, COVID-19 diagnostic codes were unavailable at the start of the COVID-19 PHE, and in some instances, COVID-19 cases were not reported in claims due to self-testing and self-management. In supplementing claims data with other data sources, many researchers noted that the lack of interoperability among claims and other databases prevented desired research.

✓ **Most limitations identified during the COVID-19 PHE were not unique to the COVID-19 PHE but rather were highlighted by COVID-19 PHE research needs.** Most limitations identified, such as the limited generalizability of individual databases, lack of standardization among databases, and time lag between health service provision and availability in claims databases, were known issues before the COVID-19 PHE. However, the PHE research needs to understand a new condition and provide timely information to policymakers, exacerbated the impact of these limitations on addressing important research and policy questions. A few limitations were unique to the COVID-19 PHE, including vaccinations and testing funded by the federal government and thus not available in claims and the need to issue and adopt new ICD codes. Preparation for a future PHE will need to address both the longstanding limitations and improved flexibility to address limitations unique to the needs of a new emerging PHE.

The findings underscored the need to consider **solutions** proactively—before another PHE—including:

1. **ENCOURAGE PRIVATE INSURERS' DATA SHARING**

   It may be beneficial to **develop creative ways to bring insurers to the table alongside other claims database providers.** For example, insurers may be more likely to engage in data sharing if incentivized through payments for the creation of shared datasets.

2. **ENCOURAGE DATABASE TRANSPARENCY**

Considering ways to incentivize greater transparency—for example, by facilitating access to data that commercial claims databases currently lack (e.g., COVID-19 testing data in exchange for databases taking specific steps to increase transparency) could address this issue. In addition, a forum could bring together claims data providers and researchers to foster greater collaboration and transparency.

3. **ENCOURAGE STANDARDIZATION TO ENABLE DATABASE INTEROPERABILITY**

Finally, it is imperative to pursue implementation of **national standards across all payers based on lessons learned from EHR data standardization.** Transforming claims into a common data model format to facilitate standardization would make claims interoperable with each other and with nonclaims databases and facilitate linkages (e.g., EHR data, social services data, survey data) to provide a more complete picture of health care access, quality, and spending.

Additional discussion of the potential solutions identified by the TEP are included in Appendix B.

There are several key **limitations to this project approach and activities**. First, information gathered from some sources may be **subject to bias**. Specifically, claims database websites reviewed for the environmental scan and KII and TEP participants from commercial claims databases or insurance companies may have been inclined to share information that made their database seem more marketable to potential buyers.

In addition, data gathered here are **not comprehensive**. The environmental scan was not a systematic review of the literature. Instead, NORC used a broad lens to capture findings from a mix of sources, including peer-reviewed literature, gray literature, and database websites. In addition, KII and TEP participants did not represent all sources of claims data; in particular, there was limited representation from private insurance companies in the KIIs; some cited that their organizations prohibited them from speaking to government agencies. Moreover, there was not a single representative from an insurance company who agreed to participate in the TEP. Therefore, this research may not be fully representative of the perspectives of the insurer audience.

Conclusions and Next Steps

This project identified several key issues related to availability and use of claims in general and commercial claims databases specifically. Most of these issues existed and were known before the COVID-19 PHE; however, the urgent need for real-time and as-complete-as-possible data during the COVID-19 PHE increased the visibility of these issues and highlighted others.

The project findings suggest that improving availability—in as close to real time as possible—and use of commercial claims data may be achieved through an enhanced collaboration between government and nongovernment leaders and payer entities. Jointly applying lessons learned and implementing agreed-upon solutions hold promise in filling the gaps identified here and achieving timely access to all types of claims when most needed during a future PHE.

The following could be achieved through this collaboration:

1. **Standard data sharing agreements** between payers and third-party entities (i.e., commercial claims databases and APCDs) that compile and curate commercial claims data to make it available during a PHE.

   Currently, there are third-party vendors that aggregate and curate commercial claims data to make them available, but there is no consensus on the data acquisition process from commercial insurers, which leads to variation across databases. Therefore, studies with similar or the same objectives may yield different results when using different commercial databases, leading to questions about the quality of research informing decision-making. A consensus data sharing process between payers and third-party entities is an important step toward improving availability and use of commercial claims. Longer term, standardization across claims data sources may be achieved through the implementation of a new federal board with government and nongovernment members.

2. Sustained and structured **technical and financial support** for projects and initiatives aimed at transforming claims into a common data model format as a preliminary step to data standardization to make claims interoperable and linkable with EHRs and other data sources.

3. Continued and improved **longitudinal research** using government and nongovernment claims, as well as linked claims with other data such as EHRs, social services, surveys, etc.

4. Consistent and structured **partnerships and collaboration** among government and nongovernment researchers for knowledge sharing.

Actions must be pursued now to fill the gaps and overcome existing barriers to ensure accurate, timely, and representative claims data are available for use when most needed.

Appendix A. Findings from the Key Informant Interviews and Environmental Scan

Background

The primary objectives of the KIIs and e-scan were to:

- Gather information about access to and use of commercial claims and provider payment data for research before and during the COVID-19 PHE.

- Capture key perspectives from those closest to the data, including those working for or with claims database providers, private health insurers, and state agencies with a focus on public health, Medicaid, and APCDs, as well as HHS agency employees.

Data collection and analyses for the two tasks were conducted in parallel, and the e-scan incorporated findings from the KIIs. Collectively, the two activities aimed to address the following research questions:

- What claims and provider data were available for research prior to 2020? What variables were available (e.g., patient demographics, pharmacy data)?

- How were health care claims and provider payment data used for research? What were the strengths and weaknesses of using health care claims and provider payment data for research?

- What new claims and provider payment data were made available during the COVID-19 PHE in response to federal, state, local, and other research needs?

- What data elements were inconsistently or unreliably captured in claims that would have been useful to inform policymaking?

- What was the quality of health care claims and provider payment data with respect to timeliness and completeness? Were these data standardized to facilitate linkages with other data for longitudinal research?

Summary of Findings

Availability and use of claims in general

1. **Prior to COVID-19, claims data were used to explore trends in health care use and pricing**, such as trends in medication dispensing and/or refills, screening, and vaccinations, and variation in prices for services, for example, by payer type.

2. **During the COVID-19 PHE, claims data were most frequently used to explore changes to health care utilization**, such as changes in utilization of telehealth, normal preventive services (e.g., cervical cancer screening, routine vaccination), and prescription medication fills. Most KII participants reported using claims data to identify changes in service utilization from before and during the

COVID-19 PHE. Some articles identified through the e-scan also explored trends in COVID-19 treatment (e.g., proportion of patients who undergo dialysis or a tracheotomy) or trends in COVID-19 testing.

3. **During the COVID-19 PHE, access to claims data largely did not change.** KII participants reported no significant changes to the types of data available within their databases or to the time lag from data collection to availability.

4. **The most regularly cited benefit of claims data for research was large sample sizes.** Researchers noted that commercial databases typically include data from multiple payers and health systems. Other benefits are that claims data are quick and relatively cost effective to analyze, enable researchers to track individuals longitudinally, and are not reliant on self-report.

Limitations of using claims data for research

KII participants and e-scan publications noted several common limitations of claims data for public health research, including COVID-19 PHE-related research. Several individuals commented that the root cause of these limitations is that claims data are simply not designed to be used for public health surveillance or research but rather for payment.

Although generally applicable for all claims databases, these limitations can be more complex and difficult to overcome for commercial claims databases, mainly due to how they are currently compiled and made available for use during a PHE and for research.

1. **Lack of data on relevant health and social factors (e.g., lack of clinical information, lack of sociodemographic data).** The KIIs and e-scan found that lack of clinical and diagnostic data, particularly related to COVID-19 test results, vaccination data, and long-term COVID-19 outcome data, as well as frequent lack of other important sociodemographic data (notably, race/ethnicity and language) currently limits the utility of claims data to address PHE-related research priorities. Participants also noted that the time needed to issue and adopt new ICD diagnostic codes limited their ability to use claims data to identify COVID-19 cases during the first months of the COVID-19 PHE.

   This lack of data stems from the fact that health care claims are generated for administrative purposes. Therefore, claims typically only include data necessary for payment and were initially used only for fraud analysis.

   COVID-19 testing and vaccination data, for example, are often incomplete because these services are not consistently submitted as claims to payers. Some relevant clinical factors (e.g., stage of diagnosis, surgical complexity) are also not collected for claims databases because they are not needed for payment purposes. In addition, claims databases are not typically linked to clinical data sources that do have this data, such as EHRs. Moreover, only diagnoses associated with services are included in claims, such that undocumented conditions and conditions for which services were unreported to insurance (e.g., those managed with over-the-counter medications) are not collected.

2. **Time lags.** All KIIs discussed the issue of time lags between health service provision, claims data collection, and availability of claims data to researchers.

Generally, time lags occur because processors must fully adjudicate claims before they are available within a database.

3. **Coding discrepancies or inaccuracies.** As with any data source, claims data include errors. The KIIs and e-scan found that claims data are subject to human error because they rely on people to report and document information. For example, inconsistent methods of classifying provider type or inaccuracies in classifying facility type can make claims data difficult to interpret.

4. **Limited time, capacity, and resources to collect, access, and analyze data.** State agency and HHS interviewees noted that limited staff time and expertise to analyze claims data, particularly during the COVID-19 PHE, limited their research using claims data. HHS staff also noted that costs to access commercial claims can be prohibitively high.

   Limited staffing capacity, competing priorities, staff shortages during the COVID-19 PHE, and insufficient funding were among the reasons noted by participants.

5. **Inadequate representativeness of claims data.** The KIIs and e-scan found that a significant, general limitation of claims databases is their representativeness. In particular, no databases include uninsured individuals. In addition, government and nongovernment claims must be linked to capture all insured individuals in a population. APCDs were created to address this issue and have a consolidated approach to accessing and using both government and nongovernment claims; however, APCDs have their own challenges and limitations. More information on APCDs can be found in other ASPE reports.[2]

   Participants acknowledged that commercial claims databases are a product of a multi-payer system where payers largely are siloed—each only collects data on its own enrollees or beneficiaries. In addition, claims are not in a common data model format to allow standardization, which translates into a lack of interoperability (i.e., multiple data formats, lack of a universal person-level identifier) that makes linking databases to obtain a more complete picture of insured individuals in the United States difficult.

   However, claims data are generally available at the person level and are linkable to other data sources. Insurers, commercial claims database providers, and APCDs all have person-level data with person-level identifiers. Linking claims data to other sources is sometimes possible using variables like date of service or national provider identifiers (NPIs). However, there are persistent challenges to linking data due to restrictions on access to person-level data from other databases and interoperability (i.e., common data formats) among many databases.

Availability of and access to data in claims databases

Following are key differences between claims databases explored in the KIIs and e-scan.

1. **Data ownership varies by data steward.** Health insurers own commercial claims data only from their own enrollees. Conversely, commercial claims databases and APCDs typically include data from

---

[2] Carman K.G., Dworsky M., Heins S., Schwam D., Shelton S., & Whaley C. (2021). *The History, Promise and Challenges of State All Payer Claims Databases: Background Memo for the State All Payer Claims Database Advisory Committee to the Department of Labor.* RAND Health Care (Document Number: PR-A1396-1). https://aspe.hhs.gov/sites/default/files/migrated_legacy_files//200696/apcd-background-report.pdf

multiple payers. APCDs have data from public insurers (e.g., Medicare and Medicaid) and select commercial insurers for individuals within their state. Commercial claims databases have claims data from select health insurers from states where the insurers operate. While HHS staff can readily access Medicare and Medicaid data from the Centers for Medicare & Medicaid Services (CMS), they have limited access to commercial claims, in part due to the cost of purchasing commercial claims databases.

2. **The degree to which claims data are made available varies by database provider.** *Some* insurers and *all* commercial claims database providers make their data available to external health organizations (e.g., federal agencies; pharmaceutical companies) at a cost. States (i.e., APCDs) sometimes charge a fee for use and tend to limit use of their data to those who can demonstrate a public benefit (e.g., academic researchers, other state agencies). HHS makes claims data (e.g., Medicare and Medicaid) available to external organizations at a cost.

   In addition, there are key differences in data availability among several large, multi-payer claims databases explored in the e-scan. Note that this information is limited only to what was described on database websites.

   a. **Health Care Cost Institute (HCCI) data may be more transparent and accessible than other claims databases.** For example, HCCI makes a data dictionary available on its website, lists which commercial insurers are included in its data, and describes its licensing process and pricing structure. However, unlike other claims databases, HCCI does not house its own, linkable nonclaims data. In addition, HCCI's sample is smaller than other databases reviewed here.

   b. **MarketScan data may be available for a longer time frame and have a shorter time lag than other claims databases.** In addition, MarketScan's claims data can be linked with its nonclaims datasets, including EHR, laboratory, and dental data. However, MarketScan does not make data variables or licensing and pricing information available on its website.

   c. **IQVIA data may have a larger geographic scope and sample size than other claims databases.** In addition, IQVIA claims data can be linked with its other nonclaims datasets, including EHR, pharmacy, and hospital data. However, like MarketScan, IQVIA does not make data variables or licensing and pricing information available on its website.

Appendix B. Findings from the Technical Expert Panel

Background

The objective of the TEP was to review and discuss the current state of health care claims data and limitations identified through the e-scan and KIIs while exploring potential solutions and the feasibility of implementation.

The following study research questions guided the conversation:

- What are the current gaps in and limitations of claims datasets when attempting to respond to PHEs?

- What strategies could be used to reduce existing barriers to improve these data for future PHEs?

Summary of Findings

Following is a summary of the primary findings from the TEP, including reported gaps and limitations in current claims data, barriers to addressing these gaps and limitations, and potential strategies to reduce barriers and improve claims data for future PHEs.

Limitations of using claims data for research

1. **Access and quality of the various data sources differ across those who collect and use claims data for public health and public policy purposes.** Medicare FFS data, the "gold standard" within HHS, only covers a specific population of adults aged 65 and older and people with disabilities, limiting the generalizability of using these data. Though federal agencies' receipt of Medicare data can be timely and allow for near real-time analyses and action, for non-federal researchers, Medicare FFS data may become available later and only after Medicaid, Medicare Advantage, and other commercial data are made available. In addition, there may be time lags to receipt of Medicare data due to internal CMS processes which delay the distribution of these data to a large demand of external users.

2. **Claims data lack information for some populations (e.g., individuals who are incarcerated, individuals without insurance) and nonclaims-based payments (e.g., episode-based payments).** Moreover, state APCDs are limited in obtaining self-insured group health plans' data, and commercial claims databases have challenges in gathering data from private health plans.

The TEP noted that there are broader issues with data availability for certain populations and health care processes across claims and other data sources. Truly addressing the limitations identified during the COVID-19 PHE would require addressing both the limitations to claims data and these limitations in health services data more broadly.

Potential Solutions

Following are several potential solutions proposed by the TEP to improve claims data access and quality for future PHEs.

1. **Consider legislative changes to compel data sharing.** TEP members suggested the federal government could introduce legislation that would require self-insured employer plans (currently exempt from reporting requirements) to share their claims data. Similarly, the federal government could establish a national APCD and/or compel health plans to provide data to one or more existing multi-payer data efforts. TEP members emphasized that private data providers have a financial interest in protecting their data and are only likely to share data if there is a more compelling financial incentive or if they are mandated to do so. Importantly, TEP members noted that federal agencies would need to have solutions in place to mitigate privacy concerns if data sharing were mandated.

2. **Standardize data collection and reporting requirements.** TEP participants stressed that a single data source is unlikely to be fully comprehensive. In addition to compelling data sharing, standardizing data collection and reporting methods—for example, by implementing a universal, patient-level identifier—would ultimately improve consistency and allow linkages across datasets.

3. **Alternatively, identify approaches for linking data that do not require a universal, patient-level identifier.** Identity resolution is critical for ensuring data are usable, but it must be done in a privacy-compliant manner to protect the data of individuals included in the dataset. One TEP member's organization had implemented a workaround; an internal team of privacy analysts used probabilistic matching to link patient-level records across data sources based on patient-level characteristics (e.g., birthday), enabling them to track individuals' service utilization and care outcomes over time with a high level of accuracy and *without* the need for a universal patient-level identifier.

4. **Strengthen the analytic workforce.** Investing in workforce development programs would familiarize more researchers with analyzing claims data and foster analytic skills before the next PHE. These investments could include training opportunities for public health departments and state agencies so they are prepared to quickly analyze data during the next PHE. Including historically black colleges and universities and other minority-serving institutions in these efforts would enable more communities to analyze and respond to a crisis with local data tailored to communities with the highest needs.