



# Claims and Provider Payment Data Gaps for Responding to COVID-19: A Framework for Future Strategies on Improving Access and Use

*The COVID-19 Public Health Emergency (PHE) made visible existing gaps and limitations in claims data that should be addressed proactively, before another PHE*

Rachael Zuckerman, Violanda Grigorescu, Steven Sheingold

## KEY POINTS

- A recent study by ASPE and NORC highlighted the claims data limitations identified during the COVID-19 Public Health Emergency Health (PHE) and provided considerations to address these limitations.
- The study found limited transparency about claims data management and availability, as well as the importance of supplementing claims data with other types of data to study the COVID-19 PHE.
- Improving the utility of claims data may be achieved through better data collection and availability; standardization of claims data across payors and database providers; and active engagement of the health care ecosystem, private sector included.
- Overall, findings revealed that identified limitations related to claims data were existing issues, exacerbated by, but not unique to, the COVID-19 PHE.
- Although the gaps identified are not specific to the COVID-19 PHE, the urgency in implementing solutions is driven by the potential risk of additional PHEs in an unforeseen future.

## INTRODUCTION

Health services researchers rely heavily on claims data to make their work informative and useful for policymakers. These data have many advantages including large sample sizes and patient-level information that can be used for studying costs, utilization, practice patterns, burden of illness and quality issues, as well as forming the basis for policy simulations. There are also well-known gaps in claims data that can limit the scope and usefulness of some research efforts. For example, claims data may fall short of fully describing important health outcomes. Moreover, our multi-payer health care system means that results of studies using a particular claims database cannot be generalized to the U.S. population and typically exclude patients without health insurance. Finally, many claims databases do not capture important information on patients and providers, or for confidentiality or proprietary reasons do not make such information available to researchers.

These gaps were highlighted during the COVID-19 public health emergency (PHE). The magnitude of this PHE created a new demand for more timely data and rapid turnaround research to inform policymakers about the ongoing situation and assist in developing response policies. Although the official declaration of a PHE led to temporary but substantial changes in health care delivery, financing, and public health reporting, the COVID-19 PHE declaration only had minor impacts for health service claims data collection and reporting.

For example, at the start of the COVID-19 PHE, important policy questions included health system capacity, health effects of COVID-19, the financial impact of the PHE on providers, and decision making related to the provider relief fund (PRF). In these cases, claims from the Medicare fee-for-service (FFS) program were identified as a potential resource to address several critical policy and research questions. However, extending same analysis to the Medicare Advantage (MA), Medicaid, and commercially insured populations remains challenging due to the gaps cited above. In addition, claims are collected for the primary purpose of reimbursement and fraud analysis, and therefore other information important to the COVID-19 PHE response—such as socio-demographic and economic information—is not always available for research.

It is also the case that commercial claims are made available through multiple third-party vendors, which results in various requirements for executing data use agreements (DUAs), as well as a variety of access methods and data definitions. These gaps limit the use of claims data for usual research purposes, but present significant obstacles when a PHE requires rapid research response.

ASPE commissioned NORC to conduct a study to detail the data gaps in claims data identified during the COVID-19 PHE and consider solutions that might better prepare HHS and the health care system more generally for research needs in future PHEs. In this paper, we build on the results of the NORC study by further detailing research experience using claims data during the COVID-19 PHE and adding additional considerations for overcoming claims data gaps moving forward.

## SUMMARY OF FINDINGS

The NORC project employed an environmental scan, key informant interviews, and a technical expert panel. The project aims were to identify gaps and challenges in accessing and using commercial claims data (with a focus on employer sponsored insurance), and to delineate strategies to improve data collection and availability for researchers and policymakers in the case of future PHEs.

As the gaps and limitations in claims data identified during the COVID-19 PHE were not specific to this PHE, they can be seen as an example of the larger gaps and limitations in claims data for use in emergencies broadly, whether a declared PHE or other fast-moving public health situation. Moreover, many of the issues identified are applicable to policy development and analysis in even non-emergency situations.

There are three key takeaways from the NORC report. First, there is limited transparency about claims data management and availability—particularly among private insurers and commercial claims database providers. Second, claims data alone had limited research utility for the COVID-19 PHE; answering some questions that arose during the PHE required using claims data along with other data sources, such as electronic health record (EHR) data and state vaccine registries. Finally, most of the identified limitations related to claims data were existing issues, and not unique to the COVID-19 PHE.

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

1. Encouraging private insurers' data sharing
2. Considering ways to incentivize greater claims database transparency
3. Encouraging standardization to enable claims databases' interoperability

It is imperative to encourage adoption of national standards across all payers as a long-term solution, based on lessons learned from EHR data standardization, such as the United States Core Data for Interoperability (USCDI) for EHR exchange maintained by the Office of the National Coordinator for Health Information Technology.<sup>a</sup> Currently, each payor may collect claims using their own unique system, making it difficult to

---

<sup>a</sup> <https://www.healthit.gov/isa/united-states-core-data-interoperability-uscdi>

aggregate claims across payors in the health care system, let alone with information outside of the health care system (i.e., social services or public health). The long-term goal is to transform all claims into a common data model format to facilitate standardization while preserving privacy, which would make claims interoperable with each other and with nonclaims databases as well as facilitate linkages (e.g., EHR data, social services data, survey data) to provide a more complete picture of health care access, quality, and spending. There is work underway to meet that goal, but this process takes time. A potential intermediary solution is to standardize third party vendors' processes for acquiring and making claims data available to researchers.

## ASPE RESEARCH DURING THE COVID-19 PHE

In this section we highlight several areas of COVID-19 PHE-related research conducted at ASPE during 2020 and 2021. Within each of these projects, claims from the Medicare FFS program were exceptionally useful, but not generalizable to other populations. Moreover, gaps and limitations in other claims databases presented obstacles to addressing the research questions for other populations.

During 2020 and 2021, ASPE staff were engaged in COVID-19 PHE-related projects monitoring changes in utilization and payments and the financial impact on providers; estimating COVID-19 related deaths, excess deaths, and their distribution among population groups; and monitoring the use of telehealth services. For each of these projects, claims from the Medicare FFS program provided an excellent source of data.

These Medicare FFS data might be considered a gold standard for claims by having the following properties:

- Available on a near real time basis
- Available at the beneficiary level
- Can be linked over time with little dropout
- Claims for all covered services are available
- Claims can be linked to individual providers
- Claims can be aggregated to multiple geographic and provider levels
- Can be linked to enrollment files with “better than average” demographic information
- Contain good information regarding diagnoses, comorbid conditions, and death
- Can be linked to other sources of mortality data on a timely basis

These projects highlighted many of the same issues raised in the NORC report. First, since Medicare covers aged and disabled individuals, the FFS claims are not generalizable to the full U.S. population. For example, ASPE was able to estimate of the rates of COVID-19 infection, hospitalization, and death for FFS beneficiaries, but private insurance claims could not be used to fully replicate the analysis for other populations. Private insurance claims data were not as timely, did not include race and ethnicity information, and could only account for COVID-19 related deaths that occurred in the hospital. In addition, extending the analysis to MA beneficiaries, now nearly 50% of the Medicare population, was challenging. The MA encounter data that provide similar information to claims was not as timely or complete and, as such, could not be evaluated during the same time periods.

Other issues arose when aggregating data from other sources and linking to Medicare claims for analyses. Although claims are available at the individual level in various commercial claims databases, proprietary and confidentiality issues may limit the usefulness of these data in some cases. The information often cannot be used by researchers at the patient level, and ASPE researchers found that data were not always available at the aggregation or disaggregation level needed, such as being available at a different provider level (i.e., provider tax identification number (TIN) vs. hospital level). A good example of the aggregation issue was an effort to estimate hospital ventilator capacity. ASPE conducted this analysis at the hospital level by using claims to estimate Medicare utilization of ventilators during peak flu seasons. Similar information was available for all

patients through the Healthcare Cost and Utilization Project (HCUP), but ASPE researchers were restricted to using estimates aggregated over state urban and rural areas rather than at the hospital level. Thus, hospital-level or even area-level estimates of capacity, or those for smaller geographic areas, were less precise than would have been desired.

## MOVING FORWARD: PREPARING FOR FUTURE EMERGENCIES

The NORC report and the descriptions of the ASPE efforts detail data gaps that might be summarized in three areas:

1. Claims data do not always contain the information necessary to conduct specific types of research—especially those that focus on measures of health status that cannot be directly determined from billing information.
2. Our current US multi-payer system means that any single claims database reflects the experience of a particular enrolled group, and the results from research using these data cannot be generalized to the population. Moreover, the uninsured will not be represented in claims data that come from insurance companies. As discussed in the bullet above, other sources, such as hospital discharges, may be needed to supplement claims databases to fill these gaps.
3. Each claims database may have important information missing or not available to researchers, such as race and ethnicity or complete diagnostic coding. In addition, some information may be available but only when aggregated to levels (e.g., by geography) that are less useful for addressing specific research questions.

While many types of data (surveillance, registries, health facility reporting, claims, etc.) were used in the COVID-19 PHE response, claims data are uniquely suited to many analyses of health service utilization and individual-level analyses. In non-emergency times, the gaps in claims data present obstacles to fully utilize health services research opportunities that might be valuable from a policy perspective. Perhaps more importantly, these data gaps prevent a well-informed research and policy response to a PHE and its potential impact on population health. While there are ongoing efforts to construct all payer claims databases that may alleviate some of these data gaps, such as the Department of Labor’s State All Payer Claims Databases Advisory Committee (SAPCDAC) and the President’s Fiscal Year 2023 budget request to build a national-level APCD, in the non-emergency research environment, we must also focus on preparing for future PHEs. That is, can these gaps be substantially reduced or eliminated for using claims data during a PHE as opposed to general use, and can we be better prepared for new barriers that are likely to arise in a future PHE?

First, while the current project provides a useful start to considering these important issues, there are still policy objectives to be considered, unanswered questions to address, and missing information to complete to propose solutions to the data gaps identified. As we move away from the COVID-19 PHE, assessing lessons learned and how to act on them to prepare for the future PHEs will be important. In this section, we consider what further information would be needed to craft policies to address data gaps, and what types of subsequent research might be useful in this regard.

One consideration is that the COVID-19 PHE may be unique in terms of its widespread global and national effects, its duration, and its significant public health consequences. Other PHEs may be more localized and of shorter duration. In considering PHE-related data and preparedness policies, therefore, policy makers might weigh a number of factors, including those discussed below.

*To what extent did the data gaps described above delay or prevent COVID-19 PHE-related research in a way that was harmful to policy response, medical treatment, and assessment of public health interventions?*

Claims-based research faced significant obstacles during the COVID-19 PHE, but would research using a complete claims database from all payers have produced a better policy response? The answer is likely mixed but should be further assessed in future projects. It is likely that the distribution of PRF funds would have been more efficient and equitable if full information were available on payments to providers prior to the COVID-19 PHE. On the other hand, key aspects of COVID-19, such as the emergence of variants, geographic variations in exposure, and the development of treatments and vaccines progressed so rapidly that even the most recent available information may not have allowed researchers and policymakers to do some claims-based evaluation of trends in a timely manner.

*What are the benefits, obstacles, and costs of establishing a national all payer claims database for use during a PHE?*

Following from the first question, this suggests the need for some analysis of the costs and benefits of pursuing a national, all payer claims database to be used during a PHE. Such analysis would include careful consideration of the experience during the COVID-19 PHE and an assessment of whether preparedness and response would have been greatly improved with better data. To properly inform emergency preparedness and response for future PHEs, the analysis must also take into account that future PHEs might vary in multiple dimensions from the COVID-19 PHE. The analysis should consider improving specific data elements and interoperability, as well as building a more flexible and nimble system to quickly address the needs of a PHE. Making such improvements in data would carry an infrastructure cost in terms of resources needed for assuring claims could be standardized to allow linking across payers, over time, and to databases that contain more comprehensive sociodemographic and geographic data. In addition, assessing the efforts needed to overcome statutory and regulatory barriers, as well as confidentiality and proprietary concerns, would likely involve substantial effort.

*Are there alternative ways of addressing data gaps that might be more targeted for local emergencies but with the potential to be aggregated to the national level for widespread PHEs such as COVID-19?*

To address this, it may be prudent to first determine whether alternatives would have improved the response to the COVID-19 PHE. If the answer is affirmative, we should then determine what the best alternative to the current state is.

In light of these considerations, further analysis might also explore the extent to which there are alternative, less resource intensive ways of addressing the current claims data gaps. For example, in the absence of an all-payer claims database, improvements in existing claims data and waiving restrictions on their use that data aggregators generally impose during a PHE could provide research results that could be representative of the population while maintaining privacy protections.

During future PHEs, understanding the appropriate level for specific analyses and evaluating whether the data can be provided with needed identifiers while preserving privacy, rather than imposing the same restrictions for all analyses, may allow researchers and policymakers to improve their policy response. Further analyses of populations covered by these existing databases would be needed to assess the viability of this strategy. Could this work build on existing HHS data collection efforts outside of claims data, such as the Centers for Disease

Control and Prevention’s data modernization initiative,<sup>b</sup> the National Institutes of Health’s Rapid Acceleration of Diagnostics Initiative,<sup>c</sup> or the Food and Drug Administration’s Coronavirus Treatment Acceleration Program?<sup>d</sup> As explained below, improvements in public insurance databases would also be necessary.

A first step would be to better understand why some of the data gaps currently exist and the issues involved in overcoming them. For most research, commercial claims are available to researchers through databases such as Marketscan, IQVIA, and HCCI, who aggregate claims from private insurers. We need to better understand the extent to which data gaps occur because information is: 1) unavailable in the original claims; 2) not provided to the data aggregators; or 3) provided but under conditions of restricted use (e.g., only available at certain levels of aggregation). It is crucial to engage insurers to understand these data gaps in a way that helps propose solutions. However, engaging insurers in such conversations proved to be challenging. Representatives of the insurance industry declined to participate in the NORC project through interviews or the technical expert panel.

A second step would be to get a full picture of the insured population in the US. This would require the use of multiple databases which are currently independent and not able to be combined. The current state requires separate contracts and data use agreements for each public and private insurance database, which then arrive in different formats and cannot be consolidated to examine the entire population. A consolidation of the acquisition, curation, and sharing process used by these aggregators of claims is another important step to consider moving forward. That consolidation can be achieved only through open collaboration among experts from different entities representing government and the private sector, the insurance industry included. Open collaboration and conversation would help us to better understand the differences in processes and data provided by private entities, payors or third-party vendors that limit our ability to overcome this fragmentation in data available for research and improve the nation’s overall data infrastructure.

It is also important to address problems with public insurance databases, including MA and Medicaid. While we have presented claims for the traditional FFS Medicare program as a gold standard, getting comparable information for beneficiaries enrolled in MA plans is more difficult. MA encounter data provides service utilization information similar to claims and has been available for research for the past several years. As the Medicare Payment Advisory Commission (MedPAC) has detailed, however, these data are substantially incomplete and therefore less useful for research purposes.<sup>e</sup> MedPAC has made several recommendations to improve these data. Medicaid claims data also provides a potentially rich source of research data but, as a federal database of claims provided by individual states with varying programs, remain less timely and fall short in certain areas of completeness.<sup>f</sup> Improving these two data sources would significantly improve our ability to do research during a PHE.

Finally, engaging the health care ecosystem to improve claims data is the key to both improving data collection and availability and standardizing data across payors and database providers. APCDs have begun to do this work within certain states, but not all states have APCDs and many that do have developed unique systems independent of other states. We had difficulty engaging insurers for this study, a key stakeholder for improving claims data. Finally, additional stakeholders should be engaged, such a self-insured employers, who could also benefit from improved claims data to better manage their health care spending and understand their insured

---

<sup>b</sup> <https://www.cdc.gov/surveillance/data-modernization/index.html>

<sup>c</sup> <https://www.nih.gov/research-training/medical-research-initiatives/radx>

<sup>d</sup> <https://www.fda.gov/drugs/coronavirus-covid-19-drugs/coronavirus-treatment-acceleration-program-ctap>

<sup>e</sup> [https://www.medpac.gov/wp-content/uploads/import\\_data/scrape\\_files/docs/default-source/reports/jun19\\_ch7\\_medpac\\_reporttocongress\\_sec.pdf](https://www.medpac.gov/wp-content/uploads/import_data/scrape_files/docs/default-source/reports/jun19_ch7_medpac_reporttocongress_sec.pdf)

<sup>f</sup> <https://www.gao.gov/assets/gao-21-196.pdf>

population. Building a community with the shared goal of improving health service claims information could improve research now and better prepare us for future PHEs.

## U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES

### Office of the Assistant Secretary for Planning and Evaluation

200 Independence Avenue SW, Mailstop 447D  
Washington, D.C. 20201

For more ASPE briefs and other publications, visit:  
[aspe.hhs.gov/reports](https://aspe.hhs.gov/reports)



### SUGGESTED CITATION

Zuckerman R., Grigorescu V., and Sheingold S. Claims and Provider Payment Data Gaps for Responding to COVID-19: A Framework for Future Strategies on Improving Access and Use. (No. HP-2023-15). Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services. December 2023.

### COPYRIGHT INFORMATION

All material appearing in this report is in the public domain and may be reproduced or copied without permission; citation as to source, however, is appreciated.

### DISCLOSURE

This communication was printed, published, or produced and disseminated at U.S. taxpayer expense.

---

Subscribe to ASPE mailing list to receive email updates on new publications:  
<https://list.nih.gov/cgi-bin/wa.exe?SUBED1=ASPE-HEALTH-POLICY&A=1>

For general questions or general information about ASPE:  
[aspe.hhs.gov/about](https://aspe.hhs.gov/about)