State All Payer Claims Databases: Identifying Challenges and Opportunities for Conducting Patient-Centered Outcomes Research and Multi-State Studies

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

October 2023

The 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. For more information, visit https://aspe.hhs.gov/.

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. HP carries out this mission by conducting policy, economic, and budget analyses; assisting in the development and review of regulations; assisting in the development and formulation of budgets and legislation; and assisting in survey design efforts, as well as conducting and coordinating research, evaluation, and information dissemination on issues relating to health policy. For additional information, visit https://aspe.hhs.gov/office-health-policy-hp.

Office of the Secretary – Patient-Centered Outcomes Research Trust Fund

The Office of the Secretary Patient-Centered Outcomes Research Trust Fund (OS-PCORTF) was established as part of the 2010 Patient Protection and Affordable Care Act and is charged to build data capacity for patient-centered outcomes research. Coordinated by ASPE on behalf of the Department, OS-PCORTF has funded a rich portfolio of projects to meet emerging U.S. Department of Health and Human Services policy priorities and fill gaps in data infrastructure to enhance capabilities to collect, link, and analyze data for patient-centered outcomes research. For more information visit https://aspe.hhs.gov/collaborations-committees-advisory-groups/os-pcortf.

Background

This cover memo is a companion piece to the report which follows, "State All Payer Claims Databases: Identifying challenges and opportunities for conducting patient-centered outcomes research and multi-state studies," the third in a series of reports commissioned by the Office of the Assistant Secretary for Planning and Evaluation (ASPE) from the RAND Corporation addressing state all payer claims databases (APCDs). APCDs include medical, pharmacy, and dental claims, as well as enrollment and provider files collected from private and public payers by states, usually as part of a State mandate.[i] As of January 1, 2023, a total of 23 states have either a mandatory APCD (with statutorily-mandated reporting from covered payers) or a voluntary APCD,[ii] and an additional eight states are currently developing mandatory APCDs.[iii] APCDs have an important role to play in supporting health services research, informed policy making, and health care system transparency, as they can address the need for, among other things, reliably updated data on health care utilization and the cost of care that can be tracked longitudinally across payers, Their data can also be linked to other databases, which is especially important when considering the social drivers of health.

The first commissioned report, released in June 2021, was prepared to help inform the deliberations of the Department of Labor's State All Payer Claims Database Advisory Committee, a committee created by the No Surprises Act.[iv,v] That report provided information on the status of APCDs across the states, how they have been used, and their strengths and limitations (https://aspe.hhs.gov/sites/default/files/private/pdf/265666/apcd-background-report.pdf) .[vi] The second report, released in June 2022, provided additional detail on ACPD data collection and access procedures, further described use cases, and discussed some of the most important challenges associated with operating an APCD and working with APCD data (https://aspe.hhs.gov/sites/default/files/documents/96f34fd0474b3da4884836c4341f1bbe/Linking-State-Health-Care-Data.pdf).[vii] This third report focuses on issues associated with two particular aspects of APCD research: the use of APCDs to conduct patient-centered outcomes research (PCOR) and the challenges of multi-state APCD research. PCOR aims to generate evidence about the effectiveness of treatments, services, and other health care interventions on the full range of outcomes that patients, caregivers, clinicians, policymakers, and other stakeholders have identified as important. The following report explores the use of APCDs for these two purposes, including presenting findings from a series of discussions with officials at six state APCDs, data vendors, and researchers who have used data from multiple state APCDs; example research plans for two multi-state use case studies; and a prototype data request application that could be used for the purposes of creating a standardized application for requesting APCD data from multiple states.

ASPE's interest in APCDs is heightened by our charge to coordinate across relevant federal health programs to build data capacity for PCOR, including administering the Office of the Secretary's Patient-Centered Outcomes Research Trust Fund (OS-PCORTF). Established by the Patient Protection and Affordable Care Act, the PCORTF supports the work of the Patient-Centered Outcomes Research Institute (PCORI), Agency for Healthcare Research and Quality (AHRQ), and the Office of the Secretary of HHS to conduct, disseminate, and expand capacity for PCOR and comparative clinical effectiveness studies (CER), respectively.[viii] In December 2019, Congress reauthorized the PCORTF through 2029 and expanded the range of outcomes that should be considered in PCOR studies to include the "potential burdens and economic impacts of the utilization of medical treatments, items, and services on different stakeholders and decision-makers respectively. These potential burdens and economic impacts include medical out-of-pocket costs, including health plan benefit and formulary design, non-medical costs to the patient and family, including caregiving, effects on future costs of care, workplace productivity and absenteeism, and healthcare utilization."[ix] APCDs address the broad needs emphasized in the 2022 report from the National Academies of Sciences, Engineering, and Medicine, *Building Data Capacity for Patient-Centered Outcomes Research: Priorities for the Next Decade*, for data that can be tracked over time across jurisdiction and are well-positioned to support the expanded scope of PCOR studies arising from the PCORTF reauthorization. [x]

However, researchers face a number of challenges in using APCDs for PCOR studies, including multi-state analyses. APCDs are rooted in state legislation, with state governments or their agents determining the content of their APCDs and procedures and authorities for the collection and release of data. Particular studies using APCDs are often legislatively authorized, respond to specific state concerns, focus on health care costs and utilization, and, significantly, are limited to a single state. Having multi-state data would allow for expanded research on relatively rare medical conditions and smaller population subgroups important to assess health equity. Such data would also allow for cross-state comparisons of trends. However, gaining access to data from multiple states can be challenging, and even when access is granted, the data may not be readily usable for regional or national research. For example, data from different states may not share a common format, or critical variables may be defined differently across APCDs. The difficulties in conducting multi-state APCD research are illustrated by the lack of a single multi-state effort among the more than 25 research studies included in the recent APCD Showcase (a collection of research studies, dashboards, and related APCD products from 2022-2023)[xi].

Given the common interests among states related to the use of APCD data (e.g., setting targets for health care cost growth[xii] or specifying a target percentage of total health care costs to be devoted to primary care[xiii] there have been emerging efforts to address these challenges. For example, the APCD Council, an organization of state APCD stakeholders, has developed technical specifications and file layouts for APCD data collection, commonly known as the APCD-CL[xiv] [xv], which has been adopted by some states, especially states newly establishing APCDs[xvi]. However, as discussed in the following RAND report, much work remains to be done to improve the availability, quality, accessibility, and ability to link APCD data for PCOR studies and multi-state analyses.

Looking Forward

Increased interest from states in using APCD data for common purposes as well as researchers seeking to leverage the data for PCOR studies and multi-state analyses, presents opportunities for collective action. As we noted in the introduction to the second RAND report,[xvii] OS-PCORTF has funded a project to develop and pilot a prototype database from several volunteer states with a goal of providing greater transparency into the outcomes, effectiveness, and costs of our health care system, building on a base of health care claims data being collected at the state-level.[xviii] The project builds on the work of the agency for Healthcare Research and Quality's (AHRQ) Healthcare Cost and Utilization Project (HCUP) to establish a common data model across state hospital-related discharge databases – an effort that began with a handful of participating states and is now the largest and most comprehensive hospital-based data resource, representing 48 states and the District of Columbia and over 95 percent of all hospital discharges in the U.S. In addition to being an asset to individual states, HCUP is regularly used by health policy analysts to address a wide variety of topics at the national level.

Building a model for a national level APCD is critical to supporting the rigorous and robust research needed to address some of the nation's most pressing health care challenges, including in ambulatory care settings. ASPE hopes this report and related efforts will provide a foundation for coordination and collaboration across stakeholders to further advance APCD infrastructure and the use of APCD data in research and policy.

[i] https://www.ahrq.gov/data/apcd/index.html#:~:text=All-payer%20claims%20databases%20%28APCDs%29%20are%20large%20State%20databases,States%2C%20usually%20as%20part%20of%20a%20State%20mandate.

[ii] Note that California, Washington, and Texas have both voluntary and mandatory efforts, as voluntary efforts were in place prior to mandatory efforts began.

[iii] https://www.mdclarity.com/blog/all-payer-claims-databases-apcd

[iv] Consolidated Appropriation Act 2021, Division BB, Title I, Section 115 https://www.congress.gov/116/plaws/publ260/PLAW-116publ260.pdf

[v] The Act required the Department of Labor to establish an Advisory Committee to produce a report with recommendations for a standardized reporting format for ERISA group health plans to voluntarily report to state APCDs and to offer guidance to the states on the use of the standardized reporting format, The Committee's report can be found at https://www.dol.gov/agencies/ebsa/about-ebsa/about-us/state-all-payer-claims-databases-advisory-committee,

[vi] Carman K, Dworsky M, Heins S, Schwam D, Shelton S , Whaley C. 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. The RAND Corporation, June 2, 2021. https://aspe.hhs.gov/sites/default/files/private/pdf/265666/apcd-background-report.pdf

[vii] ASPE staff, Carman K, Dworsky M, Heins, Quershi N S, Schwam D, Shelton S , Whaley C Linking State Health Care Data to Inform Policymaking: Opportunities and Challenges https://aspe.hhs.gov/sites/default/files/documents/96f34fd0474b3da4884836c4341f1bbe/Linking-State-Health-Care-Data.pdf

[viii] Patient Protection and Affordable Care Act, Publ. L. No. 111-148, 124 Stat. 119 (2010). https://www.govinfo.gov/content/pkg/PLAW-111publ148/html/PLAW-111publ148.htm

[ix] Further Consolidated Appropriations Act, 2020, Publ. L. No. 116-94, 133 Stat. 2534 (2019). https://www.congress.gov/116/plaws/publ94/PLAW-116publ94.pdf

[x] Washington, DC: The National Academies Press. https://doi.org/10.17226/26489.

[xi] https://www.apcdshowcase.org/case-studies?field_category_tid=4&page=2

[xii] "States Setting Health Care Spending Growth Targets Experienced Accelerated Growth in 2021", Health Affairs Forefront, June 29, 2023. https://www.healthaffairs.org/content/forefront/6-29-angeles-piece

[xiii] https://www.aafp.org/dam/AAFP/documents/advocacy/payment/apms/BKG-PrimaryCareSpend.pdf

[xiv] https://www.apcdcouncil.org/apcd-common-data-layout-apcd-cdl%E2%84%A2

[xv] The State All Payer Claims Data Bases Advisory Committee identified the APCD-CDL as "an immediately available starting point for a uniform data layout for adoption by APCDs" and submitters APCD data. State All Payer Claims Data Bases Advisory Committee Report with Recommendations. https://www.dol.gov/sites/dolgov/files/ebsa/about-ebsa/about-us/state-all-payer-claims-databases-advisory-committee/final-report-and-recommendations-2021.pdf

[xvi] See, for example, California: https://www.law.cornell.edu/regulations/california/22-CCR-97300

[xvii] ASPE staff, Carman K, Dworsky M, Heins, Quershi N S, Schwam D, Shelton S , Whaley C Linking State Health Care Data to Inform Policymaking: Opportunities and Challengeshttps://aspe.hhs.gov/sites/default/files/documents/96f34fd0474b3da4884836c4341f1bbe/Linking-State-Health-Care-Data.pdf

[xviii] https://aspe.hhs.gov/panoramic-view-patient-care-data-innovations-kages

Project Report

State All Payer Claims Databases

Identifying Challenges and Opportunities for Conducting Patient-Centered Outcomes Research and Multi-State Studies

Ashley M. Kranz, Michael Dworsky, Jamie Ryan, Sara E. Heins, Mallika Bhandarkar

RAND Health Care

PR-A2886-1
October 2023
Prepared for the U.S. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation

About This Project Report

State all payer claims databases (APCDs) collect data on health insurance enrollment and claims across public and private insurance plans within a single state. The further development of APCDs, including steps to facilitate use cases that combine data from multiple state APCDs, could play an important role in building data capacity for patient-centered outcomes research (PCOR). However, the utility of APCDs for PCOR studies has been limited by missing data and by missing populations not captured in state APCDs. Differences in APCD data completeness and processing across states, differences in the data available for release, lack of harmonized data, and the lack of clear information about these state-to-state variations also limit multi-state applications of APCD data, including for patient-centered outcomes research.

The U.S. Department of Health and Human Services (HHS), other federal government departments, and Congress have indicated interest in improving the utility of state APCDs for research. The Office of the Assistant Secretary for Planning and Evaluation (ASPE) contracted with the RAND Corporation to conduct discussions with state APCD leaders and other stakeholders about the uses of APCDs for patient-centered outcomes research and multi-state analyses, including barriers to these uses of APCD data and opportunities to address these barriers.

This research was funded by the Office of the Secretary Patient-Centered Outcomes Research Trust Fund through ASPE (task order # HHSP23337008T under contract #HHSP233201500038I) and carried out within the Payment, Cost, and Coverage Program in RAND Health Care.

Acknowledgments

We are grateful to the stakeholders who shared their perspectives included in this report, including leaders from state APCDs in Oregon, Utah, Colorado, Virginia, Maryland, and Massachusetts; data vendors that provide services to state APCDs; and researchers. Kerry Reynolds at RAND provided feedback related to connections between APCDs and patient-centered outcomes research, and Rosa Maria Torres at RAND provided helpful administrative assistance with the report. Kevin McAvey of Manatt Health served as a consultant on the project and shared comments on this report. Valuable feedback on this report was provided by ASPE. We thank Petra Rasmussen, Christine Eibner, and Paul Koegel of the RAND Corporation for reviewing this report.

Disclaimer

The findings and conclusions in this project report are those of the authors and do not necessarily represent the official position of HHS.

Summary

Background and Approach

All payer claims databases (APCDs) collect data on health insurance enrollment and claims across public and private insurance plans within a single state. The further development of APCDs, including steps to facilitate use cases that combine data from multiple state APCDs, could play an important role in building data capacity for patient-centered outcomes research (PCOR). The Office of the Assistant Secretary for Planning and Evaluation (ASPE) contracted with the RAND Corporation to identify challenges and strategies for leveraging APCDs for patient-centered outcomes research and multi-state use cases—with a focus on missing data, missing populations, and data linkages. A series of semistructured discussions and focus groups with stakeholders were held, including state APCD leaders, data vendors that provide services to state APCDs, and researchers who have used multiple state APCDs, and a draft common data application and two research plans for use cases relying on multi-state APCD data were developed.

Key Findings

All stakeholders indicated that APCDs have the potential to be used for patient-centered outcomes research. Their ideas regarding the potential use of APCDs were aligned with goals included in the Strategic Plan from the Office of the Secretary Patient-Centered Outcomes Research Trust Fund, especially in expanding longitudinal data resources (through data standards and linkages) to generate evidence on diverse patient populations (ASPE, 2022).

Stakeholders, however, noted major limitations related to missing data, missing populations, and opportunities to further expand the use of APCDs for patient-centered outcomes research. Major challenges highlighted by stakeholders included missing information about race and ethnicity, missing information on individuals covered by self-insured plans regulated under the Employee Retirement Income Security Act (ERISA), and missing information on individuals without insurance or covered by federal and military insurance (i.e., Federal Employees Health Benefits program and TRICARE) and other health care programs (i.e., the Veterans Health Administration and Indian Health Service). Data linkages between APCDs and other administrative data were identified as an opportunity to address some of these limitations (e.g., fill in missing information about race and ethnicity) and address research questions that cannot be fully addressed using the data available in APCDs alone (e.g., questions on the outcomes and effectiveness of specific health care interventions and/or requiring context-specific data, such as opioid overdoses and maternal health).

The use of APCDs from multiple states was recognized as an approach with significant potential benefits, including for policy evaluations, studying small population subgroups, and rare conditions. Although some multi-state use cases have been successfully completed, stakeholders mentioned barriers, including the burden of submitting separate data requests to multiple states, differences across states in allowed uses of the data, variation across states in file layout and data structures, and differences in intake and data processing procedures that could limit comparability even between states that receive similar files from payers and store data in similar layouts. Some

researchers also noted that the complexity and expense of a multi-state analysis might deter them from pursuing such work in the future given missing populations (e.g., limited information on ERISA plans) and uncertainty about the comparability of data across states.

Facilitating Use of APCDs for Patient-Centered Outcomes Research and Multi-State Use Cases

Below, we summarize issues identified by the stakeholders and possible actions suggested by the stakeholders to facilitate the expanded use of APCDs. All items in this table reflect perspectives mentioned by stakeholders during discussions (Table S.1). More detailed discussion of issues and possible actions is presented in Chapter 4 of this report.

Table S.1. Stakeholder-Suggested Options to Facilitate Patient-Centered Outcomes Research and Multi-State Analyses Using APCDs

Issues	Possible Actions (Implementer)
Missing race and ethnicity data in APCDs	• Link APCDs and hospital admission records, vital records, or other public health data (states)[a] • Probabilistically impute this information in APCDs (states)[a]
Confusion related to 42 Code of Federal Regulations Part 2 and the sharing of substance use disorder data	• Clarify and improve communication to payers (federal government or states)[a]
Missing populations: Access to ERISA plan data	• Offer more outreach to ERISA plans, with customized reporting and benchmarks to encourage voluntary data submission (states)[a] • Create legislation or regulation to require ERISA submission to state APCDs (federal government)[a] • Collect ERISA plan data in a nationwide, federally supported database, which could then be shared with states and researchers (federal government)[c]
Missing populations: Access to federal health plan and health system data	• Create a dataset with these data that could be shared with states and researchers (federal government)[c]
Missing populations: Lack of data on the uninsured	• Link APCDs with hospital records, electronic health records, and other datasets to capture information on patients without continuous insurance coverage or those who self-pay for some care (states)[a]
Missing populations: Timely access to Medicare data	• Establish more-regular Medicare data submission to APCDs, potentially including using the All Payer Claims Database Common Data Layout (APCD-CDL™) as a format for regular submission of Medicare data to state APCDs (states working with federal government)[b]

Issues	Possible Actions (Implementer)
Uncertainty regarding what is included in APCD data across states and how these data can be used	• Gather and centrally disseminate current information about what populations and data are included in each state's APCD and what uses are allowed (states)[a] • Define metrics for APCD data completeness and support state reporting of these metrics (federal government working with states)[a]
Barriers to pursuing multi-state APCD collaborations and research	• Refine and share a draft common data application (federal government)[a] • Produce and maintain a mapping between state APCD data layouts and the APCD-CDL™ (federal government)[a] • Develop open-source software for APCD data intake and processing to help make data more comparable across states (federal government)[a] • Create a learning network for users of multiple states' APCD data (federal government or states)[a] • Expand federal funding opportunities to encourage multi-state APCD work (federal government)[a] • Expand federal funding opportunities to encourage linking of APCDs to other data to conduct PCOR studies (federal government)[a] • Incentivize states to harmonize APCD datasets and data layouts, either with financial support or as a requirement to access federal health plan data (federal government)[b]
Lack of national or multi-state APCD	• Encourage states to submit their own APCD data to a federal data provider for purposes of building a multi-state resource (federal government)[c]

NOTE: The table lists issues identified by stakeholders and their suggestions of possible actions to address those issues. When possible, a party that might take a given action is identified in parentheses.
[a] These actions do not require data collection or data-sharing from the federal government.
[b] These actions potentially require data collection or data-sharing from the federal government but do not involve submission of state APCD data to a federally supported multi-state database.
[c] These actions involve establishment of a federally supported multi-state database.

Conclusions and Next Steps

APCDs can be used to conduct patient-centered outcomes research that addresses a range of questions important to patients, caregivers, clinicians, and policymakers. The use of data from multiple state APCDs may further facilitate these studies by providing control groups to evaluate state policies and allowing the examination of population subgroups and rare conditions. However, many barriers to the use of APCDs in PCOR studies and multi-state use cases need to be addressed. This report explores these challenges and potential paths forward to facilitate the use of APCDs for patient-centered outcomes research and multi-state use cases.

Contents

Tables

Chapter 1. Introduction

All payer claims databases (APCDs) collect data on health insurance enrollment and claims across public and private insurance plans within a single state. Today, 25 states have a mandatory APCD currently operating or in implementation (APCD Council, undated-b). Creation of APCDs has been motivated by a wide variety of goals, including improving the health of the population, reducing or controlling the growth of costs, and promoting price transparency (McCarthy, 2020).

State APCDs have already been used to support a wide range of policymaking and operational activities in line with the objectives envisioned by state legislatures that established APCDs. Some examples include informing legislation on surprise billing and prescription drug prices, monitoring network adequacy, providing consumers with price transparency information, understanding population health, and evaluating the impacts of state policy and reforms (Carman et al., 2022).

State APCDs offer an opportunity to support patient-centered outcomes research (PCOR), which can be defined as follows:

> Patient-centered outcomes research aims to generate high-quality evidence about the effectiveness of treatments, services, and other health care interventions on the full range of outcomes that patients, caregivers, clinicians, policymakers, and other stakeholders have identified as important. (Office of the Assistant Secretary for Planning and Evaluation [ASPE], 2022)

APCDs can provide data on health care utilization needed for PCOR studies.[1] For example, APCDs have been used to study such patient outcomes as opioid overdoses (Burke et al., 2020), delivery of appropriate care (Haakenstad et al., 2019), and out-of-pocket (OOP) costs (Steenland et al., 2019). Moreover, the expanded use of APCDs can support the U.S. Department of Health and Human Services' (HHS's) efforts to achieve the goals set forth in the new Strategic Plan from the Office of the Secretary Patient-Centered Outcomes Research Trust Fund (OS-PCORTF) (ASPE, 2022):

1. *Data Capacity for National Health Priorities:* Build data capacity for patient-centered outcomes research that informs the needs of federal health programs, providers, and the people served by these programs.
2. *Data Standards and Linkages for Longitudinal Research:* Expand longitudinal data resources that enable patient-centered outcomes research to advance evidence generation.
3. *Technology Solutions to Advance Research:* Leverage leading technology solutions to improve data capacity for patient-centered outcomes and comparative clinical effectiveness research.

4. *Person-Centeredness, Inclusion, and Equity:* Expand the collection and analysis of socioeconomic, environmental, and other data so all people making health care decisions have the evidence they value about the outcomes and effectiveness of health care. (ASPE, 2022)

Many of the goals cited by states for establishing APCDs are connected to patient outcomes, such as improving population health and providing quality and price information to inform patients' choice of providers.

However, limitations of existing state APCDs have been extensively documented. Readers interested in an overview should consult Miller et al. (2010) and Carman et al. (2022). These limitations include the payers and populations captured (or not captured) and the completeness of data elements that are submitted.

The state-by-state nature of the APCD landscape also means that research and analyses using APCD data often include only single-state data. While single state analyses are useful for many applications (and may be preferable for some applications because of differences in states' health care systems and policy environments), there are also numerous use cases in which cross-state comparisons or multi-state analyses that pool data from multiple state APCDs could offer advantages. Key examples include research on statewide policy changes (which may require data from another state to provide a suitable comparison group) and research on population subgroups and rare conditions (which may require multi-state data to obtain a sufficiently large sample).

Addressing limitations of existing APCDs, which arise from a variety of sources, could increase the utility of APCDs for policymakers, researchers who focus on patient-centered outcomes and other topics, and public health. As noted in the OS-PCORTF strategic plan, further development of APCDs, including a database combining multiple existing state APCDs, could build and strengthen longitudinal data capacity by providing information on patients' health care utilization and outcomes over time, across geographic boundaries, and across multiple care settings (ASPE, 2022).

Objectives and Approach

This report was commissioned to identify challenges and strategies for using APCDs for patient-centered outcomes research and multi-state use cases—with a focus on missing data, missing populations, and data linkages. This was accomplished through several activities:

- a series of semistructured individual and group discussions with stakeholders
- preparation of a draft common data application that could potentially be used to request APCD data from multiple states
- creation of research plans for two use cases involving multiple state APCDs.

Stakeholder Discussions

Individual and group discussions with stakeholders were held in the spring and summer of 2023. Table 1.1 summarizes the topics and number of participants by stakeholder group and discussion modality. This report describes findings from those discussions.

Table 1.1. Overview of Stakeholder Discussions and Major Topics

Topics Discussed	State Discussions	Vendor Discussions	Researcher Discussions	State Group Discussions
Use cases	Yes	Yes	Yes	No
PCOR studies	Yes	Yes	Yes	No
Multi-state analyses	Yes	Yes	Yes	No
Missing payers and populations in APCDs	Yes	Yes	Yes	No
Missing data in APCDs	Yes	Yes	Yes	No
Longitudinal data linkage	Yes	Yes	Yes	No
Value of common data application	Yes	Yes	Yes	No
Data linkages with non-APCD data	No	No	No	Yes
Sample Size	**State Discussions**	**Vendor Discussions**	**Researcher Discussions**	**State Group Discussions**
Number of participants	6[a]	3[a]	4	5[a]

NOTE: "Yes" indicates that the topic was included in the discussion guide for the discussion type indicated in column header. "No" indicates that this topic was not included in the discussion guide. Two group discussions were held that included representatives from five states.
[a] Indicates the number of states or organizations that participated.

For this project, we reached out to a small group of state APCD leaders willing to participate. States recruited for this project had long-standing APCDs, contributed to a geographically diverse sample, and were deemed likely to participate given their engagement with previous ASPE-supported studies on APCDs or participation in other multi-state initiatives. This outreach began with a set of one-hour semistructured discussions with leaders from six state APCDs: Colorado, Maryland, Massachusetts, Oregon, Utah, and Virginia. These discussions were conducted separately with individual state representatives.

We then facilitated two one-hour group discussions on the topic of linking APCD data to additional data sources, such as vital records, hospital discharge data, and cancer registries. Each group discussion included representatives from two or three of the states that participated in the initial discussions, for a total of five states. State participants included state APCD leaders and/or state experts in linking APCD data.

In addition, we conducted one-hour semistructured discussions with two additional stakeholder groups: (1) data vendors that provide services to state APCDs and (2) independent researchers from academic and nonprofit organizations who have conducted research using multiple state APCDs. Data vendors were identified via suggestions from state APCDs and a recent report (McAvey, 2022). Researchers, all from universities and nonprofit, nongovernmental organizations, were identified based on suggestions from state APCDs and a review of the annotated bibliography of a recent literature review focused on studies using

data from more than one APCD (Carman et al., 2022). Invitations were extended to four data vendors and five researchers; discussions were completed with three data vendors and four researchers. The discussion guides used in this project are included in Appendix A in this report.

All discussions were conducted virtually as one-hour video meetings on Microsoft Teams. Participants were not paid. Discussions were led by a member of the project team, with a second team member taking contemporaneous notes. After finalizing the notes, we reviewed these notes and summarized key themes that emerged across the discussions.

This project was determined to be exempt from further human subjects review by the RAND Human Subjects Protection Committee.

Example Multi-State Use Cases and Prototype Common Data Application

Concurrently with the stakeholder outreach described above, we produced three documents requested by ASPE, which are included as appendixes in this report.

Example Use Cases: Insulin Out-of-Pocket Costs and Long COVID

To help illustrate the value of multi-state APCD use cases, two preliminary research use cases were developed: one for a multi-state analysis of OOP costs paid by insulin users and another for a multi-state analysis of health care costs attributable to long coronavirus disease (COVID). Without confining the analysis to any specific set of states, we identified a set of research questions of potential interest to policymakers and that could be addressed through a one-year study using APCD data. We then developed a research plan for addressing the questions identified in each use case, including a definition of the study population, high-level definitions of key measures in terms of data elements captured in APCDs, a plan for the analysis (including illustrative table shells), and a discussion of study limitations and other interpretive considerations. Finally, noting that neither example use case fully specifies all the procedures and choices that would be needed to implement these analyses, we discuss additional steps that would be necessary to execute each use case. In the insulin use case, for example, we identify resources to identify insulin prescriptions using National Drug Codes (NDCs), but we do not include the code lists necessary to implement the study.

Both use cases offer guidance on two possible approaches to a multi-state study. First, we describe a possible study approach that could be implemented in a single state alone as an initial step toward a multi-state analysis. For these analyses, a multi-state research plan would promote comparability of findings across states by supporting standardization of data processing, selection of the population of interest, measure definition, and analytic methods—steps that are necessary but not sufficient to allow meaningful cross-state comparisons. Second, we discuss research questions of potential interest (such as comparisons of state-average outcome measures) that could only be addressed using multi-state data. Our intent in describing these multi-state use cases is to illustrate the potential value of multi-state APCD use cases while also surfacing challenges to the execution and interpretability of such studies.

Example use cases were circulated to state partners for written feedback. Revisions motivated by the states' feedback were incorporated into the example use cases included in the appendix, and we discuss state perspectives on the value of these example use cases and other possible multi-state applications of APCD data in Chapter 2.

Prototype Common Data Application

A prototype common application was constructed that could potentially be used by researchers or other data users to request data from multiple states. We developed this prototype by reviewing elements that appear in the existing APCD data applications from the six states that participated in the project, the Agency for Healthcare Research and Quality's Healthcare Cost and Utilization Project (HCUP) purchase request, the Centers for Medicare & Medicaid Services (CMS) Research Data Assistance Center's (ResDAC's) data application, and conversations with key stakeholders. The resulting application includes data items common to the participating states and captures common themes observed across data applications.

The exercise of developing a prototype common data application yielded insights about areas of convergence and divergence across states with respect to the information requested in their application forms. These findings are presented as inline annotations in the prototype common data application included in the appendix of this report. While discussing barriers to and potential facilitators of multi-state APCD applications, we asked states for their perspectives on the value of a common data application.

Organization of This Report

The body of this report is organized thematically into two chapters that synthesize findings from the discussions. In Chapter 2, we present findings on the suitability of currently existing APCDs for patient-centered outcomes research and discuss options for improving on existing APCD data, including challenges and strategies related to missing data and missing populations. In Chapter 3, we present findings on multi-state uses of APCDs and discuss state and other stakeholder perspectives about current uses of multi-state APCD data, the potential value of multi-state APCD data, barriers to using multi-state APCD data, and strategies for facilitating multi-state APCD use cases. In Chapter 4, we conclude with a recap of high-level findings regarding expanded use of APCDs for patient-centered outcomes research and multi-state analyses, focusing on issues and possible actions identified by stakeholders.

Chapter 2. Use of APCDs for Patient-Centered Outcomes Research

In this chapter, we summarize findings from discussions with leaders from state APCDs, discussions with data vendors that provide services to state APCDs, and discussions with researchers related to building and strengthening data capacity for using APCDs for patient-centered outcomes research. Potential and realized use cases using APCDs, as well as challenges and opportunities related to facilitating use of APCDs for PCOR studies, including missing data, missing populations, and linkages, are discussed in the context of the four goals of the OS-PCORTF Strategic Plan (ASPE, 2022), as presented in Chapter 1.

Stakeholder Perspectives on Uses of APCDs for Patient-Centered Outcomes Research

During discussions, we asked stakeholders to share their perspectives about potential and realized uses of APCDs for patient-centered outcomes research. As related to Goals #1 and #2 of the OS-PCORTF Strategic Plan, stakeholders noted that APCDs cover a range of health insurers, making them well-suited to study state and federal health programs and the populations served by them. Researchers also indicated that APCDs can be used to examine disparities of care between Medicaid and commercially insured patients and to study children's health, because Medicaid is generally included in all APCDs and is a major payer of children's health care.

Stakeholders emphasized the complexity of APCD data and highlighted the need to increase knowledge across a variety of users and potential users, consistent with Objective 1.4 of the OS-PCORTF Strategic Plan, which focuses on engaging end users throughout the process of building data capacity. One vendor stated that a limiting factor in analysis of existing APCD data is users' familiarity with the data and the availability of people with expertise to work with complicated data. This vendor emphasized that:

> Anything is possible if you try; it comes from real investment of human capital. You have to have smart people looking at it, making business decisions based on what the data tells you. We look at it, drop fields when not reliable, or fix it as possible.

Some stakeholders further emphasized the challenge of working with APCD data and the need to develop an in-depth understanding of APCD data. Researchers appreciated state APCDs that were responsive to their questions and allowed them to better understand particular datasets. One researcher suggested that a workgroup or consortium could be formed to allow the sharing or archiving of the experiences and lessons learned from previous researchers.

Aligned with Goal #4 (Objective 4.3) of the Strategic Plan, stakeholders mentioned that APCDs were uniquely valuable for examining economic outcomes (and generating evidence to inform patient decisions). For example, data vendors mentioned that APCDs provide an opportunity to compare prices and allow comparisons of service use and quality of care across payers or

across providers. One researcher emphasized that a strong APCD infrastructure can help a state to implement policies that promote the health and well-being of its residents—noting that research using APCDs has been used to benchmark premiums for a public option health plan and to cap facility fees for hospital services. Multiple state participants and vendors emphasized that APCDs are an important resource for measuring the total cost of care for specific health conditions or populations, which can be relevant to understanding patient OOP costs.

Challenges and Strategies

Although stakeholders recognized that APCDs were well suited to support research and analyses on a range of questions within patient-centered outcomes research, they noted a number of challenges and strategies for addressing them, which are described below.

Missing Data

Researchers and state participants noted a common barrier to using APCDs for PCOR studies and examining health disparities: missing information about race and ethnicity, an issue affecting all administrative health care data. Addressing this missingness may help to achieve Goals #1 and #4 of the OS-PCORTF Strategic Plan. One state is exploring imputing missing information on race and ethnicity using patient surname and residential address. Another state participant discussed an equity initiative in their state that sought to encourage more accurate reporting of race and ethnicity in patient-level hospital encounter data; the state participant suggested that this effort might have positive impacts in reducing missing race and ethnicity data in their APCD (this initiative is in its early stages). One state participant mentioned, however, that even when this information was filled in, they had no way to know whether the information was self-reported by the patient or completed by a third party via observation, so there remained uncertainty about provenance and accuracy (which affects the quality of the data).

Related to Goal #1 of the OS-PCORTF Strategic Plan (ASPE, 2022), stakeholders mentioned that APCDs were a promising resource for studying substance use disorder (SUD). However, participants from multiple states also mentioned that data submitters continued to redact claims data related to SUD treatment and that ambiguity around 42 Code of Federal Regulations (CFR) Part 2, which provides federal requirements for the confidentiality of SUD patient records, was a barrier to data completeness. Although SUD records can be disclosed for research purposes and included in APCDs, some state participants mentioned variation in submitters' interpretation of the statute as a challenge, noting that some payers conservatively withhold or omit even permissible data in an effort to ensure that they do not run afoul of regulations. One state participant provided an example of this, indicating that one data submitter might remove only SUD diagnosis codes from encounter records, whereas another data submitter might remove the entire patient encounter. This participant noted that these inconsistencies across payers in how data submitters understand 42 CFR Part 2 make it difficult to rely on the APCD for studying SUD treatment (for example, to measure what proportion of behavioral health spending is related to SUD).

Missing Populations

All stakeholders mentioned the need to improve the representativeness of data, which aligns with the emphasis on equity and inclusion of data from all communities in Goal #4 of the OS-PCORTF Strategic Plan. Several states and researchers cited the lack of data from self-insured plans covered by the Employee Retirement Income Security Act (ERISA) as their largest missing population. With the end of the federal coronavirus disease 2019 (COVID-19) Public Health Emergency, which is expected to shift many patients from Medicaid to private insurance, the lack of data from self-insured plans is coming to the forefront. Without these data, participants from multiple states indicated that they can neither quantify what proportion of their population is employer-insured nor characterize their employer-sponsored insurance (ESI) market in terms of risk and utilization. Furthermore, the share of the market covered by ERISA plans varies state to state, and the share of ERISA plans missing (or not missing) from APCDs also varies by state. Researchers highlighted this variation as a challenge to working with multiple state APCDs, with one researcher saying:

> There is a sense that you are always missing the same people across states, but the rates of [that missingness] vary a lot across states.

Researchers encouraged APCDs to share information about who is and is not included in each dataset when possible. Stakeholder perspectives on barriers posed by the unavailability of this information are discussed in Chapter 3 in the "Variation in Missing Data and Missing Populations" section.

States have pursued and continue to consider a variety of approaches to encourage voluntary submission of data from plans covered under ERISA. One state participant shared that they have found success by reducing the reporting burden for these plans and by being more lenient with the submission guide, working from the perspective that any data, even if incomplete or not comprehensive, are better than no data. This process involved a significant upfront effort coordinating with plan administrators for about a year. Asked why some ERISA plans opt into reporting, the state participant said:

> We assume they see the value and importance of having this data available to researchers and our other approved stakeholders to improve health and identify gaps and hotspots.

Another state participant indicated that their potential interest in accepting incomplete data from ERISA plans depends on the intended uses of the data; for example, if the focus of research is utilization, detail on payments may not be necessary.

However, state participants also highlighted limitations of the current situation with ERISA plan data and noted that some time-consuming efforts to encourage voluntary submission had yielded very little in return. Another state participant mentioned that they instituted a pre-processing tool to deidentify all individual patient identifiers in an effort to ease plans' concerns about data privacy and encourage data submission. This effort, however, did not gain traction among ERISA plans, and it created more work for the APCD on the back end. It also prevented the linkage of APCD data to other administrative data, such as death records. Another state participant argued that individually asking each self-insured plan to opt into reporting will never result in comprehensive data and instead suggested that federal legislation would be much more effective.

Other state participants, meanwhile, mentioned that their states are still thinking about how to encourage submission by ERISA plans. Participants from two states mentioned an interest in demonstrating to ERISA plans the value of an APCD by sharing output that these plans could use to inform benefit design or compare themselves to similar groups. A data vendor mentioned that this has been a successful strategy in one state where they operate.

Stakeholders also flagged other populations missing from APCDs. Researchers mentioned that the lack of information about the uninsured makes it challenging to study insurance churn using APCDs. State participants also mentioned lack of data from insurers, such as the Federal Employees Health Benefits (FEHB) program and TRICARE (the health care program for military service members, retirees, and family members) as challenges. States with a large share of federal employees mentioned the lack of data submitted by the FEHB program as a significant challenge. Some states also mentioned lack of data from health care systems, such as the Veterans Health Administration (VHA) and the Indian Health Service (IHS), as a challenge because of missing information about both enrollment and utilization of care in those health care systems. One state participant noted that they have a significant American Indian population but that services provided at IHS clinics and paid for with IHS funds are not shared with the APCD, thus resulting in data that are missing not at random (MNAR), which threatens the internal validity of research on this population (Little and Rubin, 2019). This state participant indicated that a multi-state analysis of health status and health care use for a tribe would be particularly interesting, because tribal organizations often cross state boundary lines, so understanding this population as a whole might require data from multiple state APCDs.

Longitudinal Analyses

A strength of APCDs mentioned by state participants and researchers is the ability to follow individuals over time as they receive care in different settings and change health plans, which aligns with Goal #2 in the OS-PCORTF Strategic Plan: to expand longitudinal data resources for patient-centered outcomes research. Nearly all state participants indicated that their APCD supports longitudinal linkage, although one state participant noted that they could currently do this only for individuals with private health insurance. Researchers reported that it was very helpful that states made efforts to link individuals across plans, but one researcher noted that the validity of the linkage of an individual's records from different insurers and over time was not always clear. This researcher's comment was offered as part of a larger appeal for state APCDs to share more information with researchers about the strengths and limitations of their APCD data. Additionally, state participants indicated that they struggle with tracking individuals who may receive care across state lines, particularly in multi-state metropolitan areas, because states typically have enrollment and claims for state residents but lack out-of-state claims for nonresidents who utilize care in their state.

Data Linkages

The five states participating in group discussions on the topic of data linked to APCDs had linked APCD data with a variety of different data sources, including disease registries, hospital discharge data, vital records, criminal justice data, social service data, and electronic health record data. State participants indicated that they sometimes linked datasets on a regular basis and other times linked data on an ad hoc basis in response to specific requests.

State participants described several different approaches to data linkage, which varied based on the specific datasets linked. These methods generally involved probabilistic matching on multiple different variables, such as name, date of birth, and address components. Some states could use Social Security number (SSN) or a deidentified equivalent in the matching process, which greatly increased the match rate and ease of matching. However, one state participant noted that, by state law, their agency could not collect SSNs and thus could not use this variable in their linkages. Many states either used or were developing common master patient identifiers that would allow matched data to be used internally without SSNs or other identifiers.

Participants cited a few key challenges in the technical processes of linking other data to APCDs. One state participant noted that sometimes Medicare Advantage plans and pharmacy benefit managers (PBMs) would submit data with a single member ID for an entire family, causing the claims data to link to multiple records in the other datasets. Another key challenge with linking PBM data noted by a participant was that PBMs sometimes submit data separately from insurers. If a research question requires linking both pharmacy claims and medical service claims to another source of data, this can generate multiple record numbers for the same person, complicating the linkage process. Another state participant noted a similar challenge, but they were able to address the issue by using an encrypted user ID from its health information exchange (HIE).

While all state participants acknowledged technical challenges, they noted that the larger challenges were often legal or organizational. Linking data proved difficult when not supported by statute or when trying to link across separate organizations within a state. Once the data were linked, end users of the data sometimes also faced challenges, including computational intensity. One state participant mentioned that because they regularly link many datasets to the APCD, the volume of data can be computationally difficult to handle.

Several state participants noted that their states were first motivated to begin linking APCD data based on a specific research question or use case important to the state. In one state, the initial investment in data linkage infrastructure was motivated by legislation to examine trends in opioid overdoses. This initial investment and demonstration of use cases helped to expand to linkages with additional data sources to address such topics as maternal and child health, COVID-19, and the impact of climate change on public health. In another state, a participant indicated that they first linked their APCD data with vital statistics data to facilitate a research study on end-of-life care and then used these linked data in additional studies. Another participant noted that a major motivation to begin linking APCD and hospital discharge data was to enhance race and ethnicity data that were frequently missing from the APCD.

Overall, discussion group participants saw great promise in using APCD data linked with other sources to answer important research questions. Participants from multiple states agreed that linking APCD data with other data sources was important for studies of health and health care inequities. In addition to pursuing person-level data linkages to improve race and ethnicity data, participants also noted the potential of geographic-level (e.g., county or census tract) data linkages to understand health care issues in rural or socially disadvantaged areas. Another participant noted that linked data could be useful to observe self-pay care among insured individuals (i.e., through linkage of APCD data with hospital discharge or electronic health records that capture self-pay service use).

While most state APCDs began by using linked data within the state government, they are increasingly sharing their data with external organizations, particularly state universities, allowing exploration of a broader set of research questions. Participants from multiple states noted many use cases for linked data within their state. While participants expressed interest in conducting multi-state studies using linked data, they noted that this was generally not feasible given that states have linked different datasets using different methods. State participants noted that making resources available to expand and harmonize linkage efforts could facilitate patient-centered outcomes research using multi-state data.

Chapter 3. Multi-State Uses of APCDs

In this chapter, we summarize findings from our discussions with leaders from state APCDs, data vendors that provide services to state APCDs, and researchers related to the use of APCDs from more than one state, including potential and realized use cases, challenges, and strategies to encourage use of APCDs.

Stakeholder Perspectives on Uses of APCDs for Multi-State Research

We asked stakeholders to share their perspectives about potential and realized uses of APCDs for multi-state research. Stakeholders from all groups mentioned that they had used APCD data from one or more states for benchmarking, including comparing the share of total medical payments spent on primary care, pediatric quality of care, and hospital prices across states. Researchers also mentioned using multi-state APCD data to study transitions in insurance coverage, cancer care, and vaccination and to compare quality of care among physicians affiliated and not affiliated with health systems.

Stakeholders also mentioned many promising use cases of interest using APCD data from multiple states. State participants mentioned a continued interest in using multi-state APCD data for benchmarking—specifically, being able to know how their state compares on prices, spending, or quality to both nationwide averages and neighboring states. Specific examples mentioned were comparisons of per member per month costs of primary care and behavioral health between neighboring states and rates of post-acute care utilization.

Examining policy changes using multi-state APCD data was mentioned as a promising use case by researchers and state participants. All types of stakeholders expressed an interest in examining balance billing and in-network and out-of-network spending across multiple states following the No Surprises Act (passed in 2020 as part of the Consolidated Appropriations Act of 2021, with out-of-network balance billing provisions effective in 2022), a potential multi-state APCD use case focused on economic impacts. Trends in pharmacy utilization and prices due to different policies were also mentioned as multi-state use cases of interest across all stakeholder types. Multi-state APCD data were mentioned by researchers as particularly important for providing a control group to examine the impact of statewide policy changes.

State participants also mentioned the value of using APCD data from multiple states to better capture and understand state border-crossing for health care. Combining data from multiple states' APCDs was mentioned as a potential way to understand patterns of care-seeking across state borders and for learning more about facilities in neighboring states—for example, to analyze network adequacy or to include out-of-state providers in consumer-facing price comparison tools. A multi-state dataset could help illuminate where there are gaps or where each state is strong in a particular specialty—or whether providers in a particular state are delivering more low-value care than providers in other states.

Multi-state APCD data are particularly useful for analyses that are difficult to do with only single-state data. Researchers and state participants identified pooling data from multiple

APCDs as a strategy to allow examination of small populations and rare conditions. For example, one state participant suggested pooling data across states and examining the per member per month cost of leukemia and other blood disorders.

Example Use Cases

Building on insights shared from stakeholders, we developed two use cases that could be used by APCD staff or external researchers as a starting point for multi-state analyses:

- **OOP spending for insulin users (Appendix B).** State APCD data offer several advantages for studying insulin OOP spending. Inclusion of multiple payers and coverage types from both public and private insurance sources enables comparison of OOP spending across coverage types. Having medical *and* pharmacy claims for the same individual facilitates more-complete measurement of OOP costs and total costs of care. These data may support state-specific estimates, which are not generally feasible with other health care datasets. Finally, APCD data are often available sooner than other datasets (e.g., household survey or commercial claims datasets), though the time lags relative to other datasets vary widely across states.
- **Costs for individuals with long COVID-19 (Appendix C).** Many APCDs can capture individuals as they churn across different types of insurance over time, which enables a lengthy period of follow-up post-diagnosis. Furthermore, in states with a linkage to an HIE or state laboratory testing data, it might be possible to identify total costs of long COVID dating back to the initial positive COVID-19 lab test result.

The example use cases identify specific data files and data elements to be used and are based on the All Payer Claims Database Common Data Layout (APCD-CDL™) from the APCD Council, the National Association of Health Data Organizations, and the University of New Hampshire (APCD Council, National Association of Health Data Organizations, and the University of New Hampshire, 2021). Not all states use the APCD-CDL™, however, and states and vendors noted that it could require substantial effort to map data elements from different states to the APCD-CDL™. One state participant suggested that multi-state research could be facilitated by an effort to map state data elements to the APCD-CDL™ to highlight direct matches and state-specific variations. This could reduce barriers to the development of multi-state use cases and help to ensure that results from use cases such as these are comparable across states.

Although these use cases do not provide sufficient detail to support immediate implementation, they are intended to provide examples of how APCDs can be used for patient-centered outcomes research and how a multi-state approach might further enhance the value of APCD data. For example, pooling data across multiple states may allow for the exploration of economic outcomes, including among diverse patient populations (consistent with Goal #4 in the OS-PCORTF Strategic Plan [ASPE, 2022]). Using data from multiple states could allow for the examination of the impact of different state policies, such as state policies related to drug price transparency or state-level reforms that have capped OOP insulin prices. That said, analyses using more than one state APCD should be approached with care, because they require understanding the similarities and differences of APCDs across states, including obtaining data, allowed uses of data, and the populations included (e.g., varying shares of ERISA plan reporting).

Challenges and Strategies

Despite high interest in using APCD data from multiple states, stakeholders identified challenges, which we describe below, along with their suggestions for addressing them.

Variation in Processes for Obtaining Data

Researchers using data from more than one state APCD indicated that having to obtain data separately from each state was often a confusing and time-consuming process. One researcher stated that "the process for obtaining data can vary from state to state and even vary year to year within a state." Another researcher noted that they experienced "huge variation in obtaining data from states, in helpfulness, timeliness, cost, and what data was available."

Researchers indicated that the information requested in data applications was reasonable, but it was tedious to fill out multiple similar applications. Some states had additional processes for researchers to obtain Medicaid data and additional questions about external data linkages. We heard from one stakeholder who participated in a multi-state APCD analysis that there were challenges throughout their project, including obtaining permission to directly analyze each state's data and to utilize comparable data elements across all participating states. States also had different timelines for data release, which researchers indicated made it challenging to obtain data for the same years from multiple states at the same time. The cost of purchasing APCD data was also mentioned as a barrier to multi-state APCD research by some researchers. However, one researcher noted that the state that charged them the most for an APCD was the most responsive to their questions, which they indicated was extremely helpful, suggesting that these higher costs may be supporting staff time to interact with data purchasers.

Another barrier mentioned by multiple state participants was the occasional inaccessibility and extra restrictions associated with public insurance data in comparison to the commercial claims data in the APCD. Specifically, the approval and release processes for Medicaid and Medicare data could be challenging and costly, leading to inconsistency in the breadth of data contained in different states' APCDs.

Common Data Application

We specifically asked stakeholders if a common data application—that is, a single data request form that could be submitted by external researchers to request data from multiple states simultaneously—might facilitate research using APCD data from multiple states. Stakeholders generally agreed that this would facilitate multi-state use of APCDs. Participants from two states suggested that the common data application could include the common fields across all the APCDs as mandatory fields while leaving the more expansive data collected by some states but not others as separate, optional fields or state-specific add-ons.

We heard from both state participants and researchers that a common data application is a good idea in theory but that it would be difficult to implement until all APCDs had a common data layout (CDL) with harmonized variables. Variation in data request processes was also mentioned as a potential barrier to a common data application. One state cautioned that state review committees would likely be sensitive to being overstepped and will still want to have a say in what gets released via a common data application.

After reviewing elements that appear in six existing state APCD data applications, the HCUP data application, and ResDAC's data application and holding conversations with key stakeholders, we developed a prototype common data application (Appendix D) and an annotated common data application (Appendix E). The annotated application includes text describing why certain elements are included and deviations observed across applications, and it highlights outstanding issues—such as the challenge of using a common data application when datasets are not standardized across states. Some instances of variation are highlighted below:

- **Study aims:** While all applications asked applicants to describe their study aims, the level of detail requested varied. This could be a brief 150-word abstract or may include a more extensive explanation of project objectives, a brief summary of the literature, specific research question(s), individual specific aims, project methodology, and a description of intended products or reports to be derived from the requested data.
- **Purpose:** Applicants were typically instructed to check a box to indicate the purpose of the project. This item could be used to distinguish between research, health care operations, or public health activities. There was variation in the level of detail provided, with some applications including potential purposes for the data request: for example, assess utilization of health care services, observe cost trends, compare providers/health plans, create or enhance a commercial product or service, or assess population health.
- **Data linkages:** About half of the applications reviewed asked about data linkages. One state strongly prefers to conduct any data linkages in house prior rather than providing the dataset in order to prevent potential reidentification by linkage.
- **Standard limited dataset:** Sometimes a standard limited dataset was available, and requesting this may allow applicants to complete a shorter application. When a standard limited dataset was not available, the application typically included many questions to determine the least amount of data that can be shared to achieve the project aims.
- **Data security:** The amount of information about data security requested by state APCDs varied widely. Some applications had more requirements related to data security, and some had less.

Variation in Allowed Uses for Data

Some states have legislative restrictions regarding the uses for which their APCD data can be released. For example, despite high interest in public health surveillance, some APCD data can only be released for official statistical or research purposes, not public health surveillance. Some APCDs do not allow for the identification of prices for individual providers, and others are not available to external researchers. Additionally, the inability or unwillingness of some states to pool data with other states was noted as a barrier to multi-state research—this was not described as a major barrier, as researchers and data vendors indicated that there are times when pooling data would not be appropriate (e.g., when examining states that greatly differ in geographic location and population characteristics), and there are strategies for dealing with this (e.g., aggregating data and then pooling them or having states run analyses themselves and then sharing results). Reasons provided by researchers for separately analyzing data from multiple states rather than pooling data included data use agreements prohibiting the pooling of data, different security requirements across state APCDs, the large

size of datasets, availability of different years of data, and different timelines for receiving data. Overall, researchers indicated that it was often unclear what APCDs allowed or did not allow researchers to do with these data, and researchers requested more transparency from states about these restrictions.

State participants indicated that states had different norms for sharing data extracts with researchers—something that researchers and state participants indicated would present a challenge to using a common data application across states. Some states have a standard limited dataset that can be released and tailored to include more or less information (e.g., full dates versus month and year only), while other states produce custom datasets for each request. Two state participants indicated that they tailor extracts based on what is requested to ensure that they release only the minimum necessary data to complete the requester's project, though one of these states is considering creating more streamlined limited extracts that could be requested and provided without tailoring. One state participant noted that they expected that state-to-state differences in output files would matter more than any differences in data request processes or layout.

Variation in File Layouts, Data Structure, and Data Cleaning

As highlighted in Goal #2 of HHS's Strategic Plan for the OS-PCORTF (ASPE, 2022), data standards and common data models facilitate data sharing and use. Stakeholders had different perspectives about how much of a barrier the different layouts of APCDs were to multi-state APCD research. One researcher who indicated that this was not much of a barrier stated that the files were "pretty similar" across states and that, while these files were "messy," that was similar to their experience with other claims datasets. Another researcher indicated that these differences were a large barrier to multi-state APCD research, saying:

> I don't know that I would want to combine state APCDs. I don't know how that would be feasible because it's not standardized at all, it's worse than how Medicaid data is not standardized state to state.

The APCD-CDL™ from the APCD Council is a potentially important tool for states to use to produce comparable APCD data (APCD Council, National Association of Health Data Organizations, and the University of New Hampshire, 2021), but barriers to adoption of the APCD-CDL™ were noted. One vendor noted that states who created APCDs before the development of the APCD-CDL™ may not have an interest in (or the funding to support) the investment needed to adopt the APCD-CDL™. This vendor noted that the transition costs associated with changing their process could be significant relative to existing APCD budgets.

A vendor also noted that, although there had been hope that the APCD-CDL™ would be adopted for voluntary submission by ERISA plans (as recommended in 2021 by the Department of Labor's State All Payer Claims Databases Advisory Committee[State All Payer Claims Databases Advisory Committee, 2021]), this has not happened. State participants generally supported using the APCD-CDL™, with the caveat that some states collect more information than is supported by it. One state participant, for example, mentioned that their APCD collects more information on primary care and behavioral health than other states do, and another state participant asserted that adopting the APCD-CDL™ would result in a loss of important information about race and ethnicity compared to the state's current data layout.

However, differences between current state data layouts and the APCD-CDL™ were framed as not insurmountable, as long as there is sufficient standardization to support useful comparisons. One state participant mentioned that they had recently changed their data submitter guide to align more closely with the APCD-CDL™ and noted that being more aligned with those standards would make it easier to participate in multi-state aggregation efforts. Another state participant said that a CDL would do the most to foster data use across states by producing directly comparable datasets.

Multiple vendors noted, however, that use of the APCD-CDL™ alone was insufficient to address differences across states that affect the validity of findings from multi-state analyses. First, one vendor raised the concern that, if the APCD-CDL™ were adopted as a format for data storage to enable the creation of more comparable analytic files for multi-state analyses, states that use different or more detailed data layouts for submission from payers might lose information when data are formatted to match the APCD-CDL™:

> The goal of CDL was if you store data in that format, it's better in a theoretical cross-state share; but if the APCD is storing all their data in the CDL [i.e., if they only keep the analytic file], then fidelity is lost.

Second, even the adoption of the APCD-CDL™ does not eliminate the possibility that heterogeneity in intake processes or other business rules will reduce the comparability of data across states: Even states that use the APCD-CDL™ might have challenges with comparability due to other steps in the data pipeline that are not specified in the APCD-CDL™. Multiple vendors voiced the concern that differences in data intake across state APCDs—especially across states that work with different vendors for data management—could make it substantially more difficult to make valid comparisons across states. The application of alternative diagnostic groupers or other proprietary algorithms for processing claims data (e.g., identifying inpatient versus outpatient hospital episodes) was identified as an example in which steps taken relatively early in the data processing pipeline can limit comparability across state APCDs in ways that are difficult or impossible to reverse. One vendor said:

> All data aggregators do it all differently, it can be like taking a copy of a copy of a copy. . . . [It is] easier to go across payers than to go across states because you don't know how data aggregators are standardizing to manipulate the data in their state.

A vendor also noted similar comparability challenges with consumer-facing price transparency tools in neighboring states: The methodologies were different, so prices reported by the two states' tools were not comparable.

One vendor felt strongly that the critical point in the data-processing pipeline for standardization across states was at the very beginning. This vendor suggested that the development and adoption of open-source algorithms and standards for intake of claims data would be a good path to address this problem. If reliable open-source software were available, then states could specify that a vendor use software that promotes standardization of the data intake process. While another vendor that works with multiple states felt that they already had good transparency about how data were handled, they indicated that there was room for greater standardization across states in other areas, such as how claims versioning was done.

In a similar vein, a different vendor called for the establishment of

> a place to share methodology for more prickly areas and to understand how data are being processed; it could be GitHub or something else for information to be shared on how analysts or data managers have constructed data variables, so that they could be used differently.

Furthermore, one vendor noted that data submission to state APCDs by CMS had longer lags than were tolerated for other payers and that the state-by-state process through which CMS provides data to state APCDs led to issues with comparability, noting:

> Every state receives Medicare data and has a unique mapping into a more standard APCD format, but that creates a new opportunity for misalignment.

Variation in Missing Data and Missing Populations

One vendor noted that missing populations in state APCDs are a greater threat to internal validity for cross-state comparisons than for single-state analyses because "issues get hidden." They noted that, due to ERISA, VHA, and other missing populations, the completeness and reliability of data for one state's price comparison tool varied widely across hospitals within the state; that vendor pointed out that these issues are even harder to track accurately and address if data are combined for multi-state comparisons. A vendor also noted that vendors are not in a good position to initiate efforts to improve data submission (for example, by encouraging voluntary submission by ERISA plans) unless state APCD leadership is interested. On this topic, one vendor said, "We are fully dependent on the state to take it as far as they want."

As noted in Chapter 2, researchers also identified variation across states in missing populations as a barrier to working with multiple state APCDs. Differences across states in the size and composition of missing populations are compounded by the scarcity of information about which populations are included in each APCD. One researcher summed this up by saying,

> I wish it were a little easier to know who's included and who is not included in the APCD. It's not always consistent across the states.

Researchers mentioned that it would be helpful if more states shared information about what proportion of (or if any) ERISA plans in a state submitted to the APCD and if the APCD includes information only on state residents seeking care in that state or anyone seeking care in that state.

One vendor echoed this criticism, saying that information about the completeness of data in an APCD they operate was not transparent; this vendor suggested that some state governments may lack the technical sophistication or resources to rigorously compare how the population in the APCD compares to a population of interest (e.g., the population of state residents with health insurance or the population with ESI coverage). This vendor also indicated that, even when a state understands the completeness of its APCD data, this information was often not reported to the public in a transparent or readily available manner.

State participants and researchers also indicated that the lack of information about race and ethnicity was a challenge, albeit a challenge that was common to all states. One vendor, however, noted that some payers had concerns about sharing demographics such as race and

ethnicity because they had not obtained consent from their policyholders to share that data for research.

Several strategies to reduce missing populations in APCDs were discussed. One state participant mentioned that a multi-state APCD could begin with a federally initiated database of currently missing data (for example, ERISA plans and FEHB program data), then add existing state APCD data, rather than attempting to harmonize existing but incomplete APCDs. It was also suggested that the federal government is better positioned than states to require participation from self-insured employers and could use specific use cases as an incentive for self-insured plans to submit data to a multi-state APCD. Another state participant mentioned that the ability to extract data from multiple states may encourage ERISA plans to submit data. Many of the self-insured employers in that state employ residents of surrounding states as well and, it was suggested, could learn from a multi-state regional APCD whether their benefits were competitive with other, similar employers.

In addition to mentioning the lack of mandatory submission by ERISA plans, one vendor suggested that state and federal policymakers are able to signal to players in the health care industry (both payers and facilities/providers) that transparency in prices is the government's expectation in hopes of promoting a culture shift that reduces barriers to data-sharing and APCD participation. Examples of such policies include funding for APCD operations and preemptively addressing arguments against sharing by payers (e.g., that the Health Insurance Portability and Accountability Act [HIPAA] prevents the sharing of claims data with APCDs or that the sharing of claims data represents a unique threat to members).

Vendors also mentioned that there is momentum toward greater data privacy, including allowing patients to opt out or censor certain parts of their records, fueled, in part, by concerns among some patients and providers about protecting reproductive health information from law enforcement.

Funding and Future of APCDs for Research

Lack of funding was mentioned as a barrier to working with multiple APCDs by state participants and researchers. Less-populous states with fewer APCD staff face a particular challenge of having less funding and less bandwidth to pursue additional funding opportunities, despite having ideas and cohesive plans for multi-state efforts. One state participant reported that they submitted a federal grant to support a three-state effort to create a public-use multi-state dataset. However, the grant was not awarded, and the work was unable to move forward without funding. Some state participants indicated that they are exploring multi-state collaborations, which may include developing common research agreements, combining data, or harmonizing outputs. However, lack of additional funding or federal support can make it difficult for states to prioritize these multi-state efforts.

Some vendors handle data processing for multiple states' APCDs, and one state participant suggested that federal monetary support could encourage such vendors to pursue harmonization efforts and facilitate multi-state comparisons. One vendor suggested that federal support for states to adopt the APCD-CDL™ would be valued because states with APCDs that predate the APCD-CDL™ may otherwise lack the resources needed to overhaul their data intake procedures and documentation.

One researcher cautioned that research proposals using APCDs may be viewed as "high risk, high reward" because of the uncertainty around what is included in APCDs and how they can and cannot be used. This researcher suggested that a grant call be issued for research using APCDs to offset data costs and encourage innovative uses of multi-state APCDs, including linkages to other datasets. This echoes what we heard from one state participant, which was that while researchers may want to access data from multiple states' APCDs in order to produce more generalizable findings, it is mainly large and well-funded research organizations or universities that are able to navigate the multiple different data request processes for each APCD.

Researchers were divided on whether they planned to continue to work with data from multiple APCDs. Some researchers viewed APCD data as much cleaner and easier to obtain than data acquired directly from private insurers. However, one researcher was not sure that the complexity of working with APCDs was worth it compared to using other data sources, such as private claims data or HCUP discharge data. Reflecting on navigating multiple application processes and trying to understand data from multiple states, another researcher lamented:

> I worry that APCDs will go to the wayside because private companies are doing a good job pulling data together at a large scale. The cost is a lot more, but there are more people and more states. There's only one vendor, it's a more streamlined process. They want you to be a happy customer.

Chapter 4. Summarizing Issues and Possible Actions

This report was produced to identify challenges and strategies for using APCDs for patient-centered outcomes research and multi-state use cases. To address these goals, we held semistructured discussions with leaders from state APCDs, data vendors that provide services to state APCDs, and researchers. Although stakeholders indicated that APCDs have the potential to be used for patient-centered outcomes research, they also noted missing data and missing populations as important limitations. Data linkages between APCDs and other administrative data were identified as an opportunity to address some of these limitations and examine research questions that cannot be addressed using data from APCDs alone.

Regarding multi-state analyses using APCDs, stakeholders saw significant potential benefits, including providing control groups to evaluate state policies and a sufficiently large sample to examine population subgroups and rare conditions. However, stakeholders mentioned barriers to using APCD data from multiple states, including differences across state APCDs in data release procedures and allowed uses, variation across states in file layout and data structures, and differences in intake and data processing procedures that could limit comparability even between states that receive similar files from payers and store data in similar layouts. Multi-state analyses have been successfully conducted, and some are currently underway, but some researchers also noted that the complexity and expense of a multi-state analysis might deter them from pursuing such work in the future, especially given uncertainty about the comparability of data across states.

Issues and Possible Actions

Stakeholders identified a number of issues, which, if addressed, would help facilitate the expanded use of APCDs for patient-centered outcomes research and multi-state analyses. These issues and possible actions suggested by stakeholders are summarized in Table 4.1.

Table 4.1. Stakeholder-Suggested Options to Facilitate Patient-Centered Outcomes Research and Multi-State Analyses Using APCDs

Issues	Possible Actions (Implementer)
Missing race and ethnicity data in APCDs	• Link APCDs and hospital admission records, vital records, or other public health data (states)[a] • Probabilistically impute this information in APCDs (states)[a]
Confusion related to 42 CFR Part 2 and the sharing of SUD data	• Clarify and improve communication to payers (federal government or states)[a]
Missing populations: Access to ERISA plan data	• Offer more outreach to ERISA plans, with customized reporting and benchmarks to encourage voluntary data submission (states)[a] • Create legislation or regulation to require ERISA submission to state APCDs (federal government)[a] • Collect ERISA plan data in a nationwide, federally supported database, which could then be shared with states and researchers (federal government)[c]
Missing populations: Access to federal health plan and health system data	• Create a dataset with these data that could be shared with states and researchers (federal government)[c]
Missing populations: Lack of data on the uninsured	• Link APCDs with hospital records, electronic health records, and other datasets to capture information on patients without continuous insurance coverage or those who self-pay for some care (states)[a]
Missing populations: Timely access to Medicare data	• Establish more-regular Medicare data submission to APCDs, potentially including using the APCD-CDL™ as a format for regular submission of Medicare data to state APCDs (states working with federal government)[b]
Uncertainty regarding what is included in APCD data across states and how these data can be used	• Gather and centrally disseminate current information about what populations and data are included in each state's APCD and what uses are allowed (states)[a] • Define metrics for APCD data completeness and support state reporting of these metrics (federal government working with states)[a]
Barriers to pursuing multi-state APCD collaborations and research	• Refine and share a draft common data application (federal government)[a] • Produce and maintain a mapping between state APCD data layouts and the APCD-CDL™ (federal government)[a] • Develop open-source software for APCD data intake and processing to help make data more comparable across states (federal government)[a] • Create a learning network for users of multiple states' APCD data (federal government or states)[a]

Issues	Possible Actions (Implementer)
	• Expand federal funding opportunities to encourage multi-state APCD work (federal government)[a] • Expand federal funding opportunities to encourage linking of APCDs to other data to conduct PCOR studies (federal government)[a] • Incentivize states to harmonize APCD datasets and data layouts, either with financial support or as a requirement to access federal health plan data (federal government)[b]
Lack of national or multi-state APCD	• Encourage states to submit their own APCD data to a federal data provider for purposes of building a multi-state resource (federal government)[c]

NOTE: The table lists issues identified by stakeholders and their suggestions of possible actions to address those issues. When possible, a party that might take a given action is identified in parentheses.
[a] These actions do not require data collection or data-sharing from the federal government.
[b] These actions potentially require data collection or data-sharing from the federal government but do not involve submission of state APCD data to a federally supported multi-state database.
[c] These actions involve establishment of a federally supported multi-state database.

Missing Race and Ethnicity Data in APCDs

Stakeholders suggested strategies that states could take to address missing data on race and ethnicity and other information relevant to analyses of health equity. Addressing missing race and ethnicity data in APCDs aligns with Goal #4 in the OS-PCORTF Strategic Plan, *Person-Centeredness, Inclusion, and Equity* (ASPE, 2022). One suggestion raised by a participating state and by data vendors is that **data linkages between APCDs and hospital admission records, vital records, or other public health data can help to fill in missing information in APCDs on demographic and socioeconomic status measures**. Another approach, which is feasible in states that collect name and street address, is **indirect estimation, in which these direct identifiers are used to probabilistically impute demographics to individuals appearing in the APCD**. This type of approach has been used to address missing and incorrect race and ethnicity information for individuals with marketplace insurance plans (Sorbero et al., 2022), Medicaid (Silva, Trivedi, and Gutman, 2019), and Medicare (Haas et al., 2019) and was mentioned by one participating state as an approach they are actively exploring for use with data from their HIE.

Potential Confusion Related to 42 CFR Part 2 and the Sharing of SUD Data

Two state participants noted continued challenges with the completeness of behavioral health and SUD data in APCDs. These state participants highlighted that payers have adopted widely varying practices for redacting these claims, attributing this variation to a perceived lack of current guidance from HHS about what claims can be shared. One state also remarked that some health care providers were uncertain about what information they could report on SUD patients, further undermining the accuracy of data submitted by payers. Although the Substance Abuse and Mental Health Services Administration has taken steps to encourage the inclusion of these data in claims databases—including the November 2022 notice of proposed rulemaking implementing changes required under the Coronavirus Aid, Relief, and Economic Security Act (Office of the Secretary, Department of Health and Human Services, 2022)—

differing interpretations of 42 CFR Part 2 continue to limit the value of APCDs for studying behavioral health disorders and SUDs. These state participants suggested that **HHS clarifications and communications around policies related to behavioral health and SUD claims could help address these challenges, along with outreach and education focused on payers or states with APCDs, who could then communicate this to payers**.

Missing Populations

Participating states and other stakeholders suggested a range of possible actions that could help to address the lack of data from ERISA plans, federal health plans, and federal health systems in APCDs.

Access to ERISA Plan Data

One approach that was described by a state participant and a vendor as successful at encouraging voluntary ERISA plan reporting was for the states to **conduct targeted outreach and offer customized reporting and benchmarks to ERISA plans**. This approach could be pursued by individual states but faces inherent limitations: Voluntarily submitted data from ERISA plans will still represent a convenience sample, and any inferences based on analyses with these data would rely on strong (and likely untestable) assumptions.

Overall, stakeholders noted that state efforts to collect ERISA plan data are inherently limited in their effectiveness under existing federal regulation due to ERISA preemption and suggested several federal actions that could better assist states in obtaining usable ERISA plan data. Multiple state participants said that federal action was necessary to address the challenges posed by ERISA plan nonparticipation in APCDs. **One approach suggested by a state participant and a vendor was for the federal government to take regulatory or legislative action to require ERISA plans to participate in state APCDs.** Another approach suggested by a state participant and a vendor was to have **the federal government mandate the collection of nationwide ERISA plan data in a new national database**. These stakeholders indicated that these data could then be shared with state APCDs.

Access to Federal Health Plan and Health System Data

Stakeholders also mentioned challenges related to missing data or untimely data submissions from federal health care payers. Participants from states with a large number of federal employees strongly emphasized the importance of obtaining data from the FEHB program. States also mentioned challenges that were due to missing data from TRICARE, the VHA, and IHS and expressed strong interest in obtaining data on enrollment and utilization for each of these programs. **A state participant and a data vendor called for the federal government to address these issues by constructing a data resource that includes data from some or all of the above federal health plans and health systems, from which data on currently missing populations could then be distributed to state APCDs.**

Lack of Data on the Uninsured

Stakeholders acknowledged how the lack of data on the uninsured in APCDs limited their value for studying populations who are persistently or even occasionally uninsured. **A state participant mentioned that this limitation can be addressed in some cases through linkage of APCD data to hospital records, electronic health records, and other datasets,**

which can capture information on patients without continuous insurance coverage or those who self-pay for some care.

Timely Access to Medicare Data

While many state APCDs already include Medicare data, a data vendor and a participating state voiced concern about the data lags involved in Medicare data submissions to the states. This data vendor also noted the potential for error introduced by the need for CMS to submit data to states with varying submission formats. As an option to address these issues related to Medicare data, this data vendor suggested that **the federal government could provide more-regular submissions of Medicare data to state APCDs using the APCD-CDL™ rather than tailoring the data submission to each state's data layout**.

Uncertainty Regarding What Is Included in APCD Data Across States and How These Data Can Be Used

Researchers and participating states also highlighted a need for more information on the similarities and differences of APCDs across states:

- How can fields in each state's data layout be mapped to a common format for use in analysis, and how do data definitions differ across states?
- What percentage of the state population is included in the APCD, and how complete are data submissions on each included population segment?
- Which payers and populations are included or excluded in each state?
- What uses are allowed with each state's data?

Participants indicated that improving the collection of rigorous and comparable information to address these and other questions would be valuable to data users and would facilitate both single-state and multi-state research with APCDs. Researchers noted that better information about the populations included in each APCD and allowed uses would reduce barriers to initiating projects using multi-state APCDs. We note that some of this information is available on the APCD Council's website, including, for each state, links to claims data collection rules, data release rules, and the data request process (APCD Council, undated-a). To enhance the visibility and usability of this information for researchers, **states might consider summarizing this information and sharing it in a centralized location**.

Alternatively, one vendor suggested that the federal government could support efforts at making information about what populations are included in APCDs by **providing states with clearly defined specifications for a set of metrics reflecting APCD data completeness**. This vendor suggested that **such metrics might be reported in a similar format by the states and that the federal government might provide funding for states to establish comparable reporting on data completeness**.

Barriers to Pursuing Multi-State APCD Collaborations and Research

Several barriers to multi-state APCD collaborations and research were identified by stakeholders, and a number of possible actions were suggested to address these barriers.

To reduce researcher burden associated with filling out multiple different applications to obtain APCD data from several states and encourage greater use of multiple APCDs,

researchers voiced support for the states adopting a common data application. The draft common data application included in Appendix D can be used as a model.

However, state participants indicated that the comparability of APCD content, differences in the data release processes, and differences in the types of files available for independent data users were challenges that would still need to be addressed to meaningfully reduce this burden, and that the effort involved in completing separate state applications was not the most significant barrier to multi-state analyses. One challenge mentioned by state participants and vendors is that substantial effort is needed to understand how states' data layouts compare to one another. One suggestion made by a state participant was that a **publicly available and regularly maintained mapping of states' data layouts to the APCD-CDL™**, which might be carried out or supported by the federal government, would facilitate multi-state use cases and the development of a multi-state APCD.

A similar concern raised by multiple vendors was that even states with identical data layouts may not have truly comparable data due to differences in data submission quality checks, thresholds for acceptance of submissions, business rules for data processing, episode groupers, or other technical differences in data intake and processing. One vendor suggested that **the development and adoption of open-source software that could be used for data intake and processing would help ensure that information collected by different state APCDs was comparable**.

Several state APCDs have established user groups that meet regularly for information-sharing between APCD staff, data vendors, and data users, such as independent researchers. **Researchers suggested that the creation of a similar learning network or user group focused on multi-state APCD analyses could provide a forum for sharing lessons about multi-state analyses.** An example of a similar network is the Medicaid Data Learning Network, managed by AcademyHealth with support from the Commonwealth Fund and the Robert Wood Johnson Foundation, which provides a forum for researchers to share their experiences and best practices for working with new Medicaid data (AcademyHealth, 2023).

Another barrier to multi-state APCD research identified by state participants and researchers is funding. Researchers observed that, at present, funders may view multi-state APCD research as high risk (i.e., unacceptably likely to fail), albeit with the potential for high reward. One researcher suggested that the federal government and other research funders (such as foundations) might help to **promote multi-state APCD research by issuing requests for applications or other targeted calls for proposals specifically for projects that use multi-state APCD data**. Such research funding or grants to states could also help promote linkages of APCD data to other datasets. One current example of this is the Robert Wood Johnson Foundation's Health Data for Action (Data Access Award), which provides access to certain datasets for no charge—including APCD data from two states (Robert Wood Johnson Foundation, 2023). Similarly, state participants indicated that **making resources available to expand and harmonize state-level data linkage efforts could allow for important multi-state use cases involving linked data**.

State participants and vendors suggested that **the federal government could also facilitate multi-state APCD use cases by taking steps to incentivize the collection of more comparable data in different state APCDs**. While **financial support** could be offered to

create such incentives, state participants and vendors noted that **the creation of a national database containing data on populations missing from state APCDs could also be leveraged to incentivize standardization of state APCD data**. If the federal government could collect federal and ERISA claims data from missing populations, state participants and data vendors suggested that state APCDs would have considerable interest in accessing these data. The federal government could then incentivize standardization of technical specifications across states where such variation hampers multi-state use cases. Specifically, data vendors suggested that the state receipt of federal payer and health system data could be conditioned on

1. standardization of data submission formats where these overlap with the APCD-CDL™ (without limiting the ability of states to continue using their own state-specific fields in addition to APCD-CDL™ fields)
2. standardization of data quality checks and thresholds for data submission acceptance
3. standardization of intake processes or business rules for processing data.

Lack of National or Multi-State APCD

Researchers also voiced interest in obtaining multi-state APCD data from a national or multi-state APCD data resource. One model discussed by a state participant might resemble HCUP, in that **states might be encouraged to submit their own APCD data to a national database, which might then disseminate multi-state data (including missing populations from ERISA and federal health plans and federal health systems)**. To construct such a federal claims database, states would need to be encouraged to submit their own APCD data to the federal data provider for purposes of building a multi-state resource. A more detailed proposal along these lines was laid out in McAvey (2022).

Some researchers did question the value of APCDs for multi-state use cases, noting that, at present, existing commercially owned databases offered a more usable resource for comparing commercially insured individuals across states than individual state APCDs. A researcher noted that these commercial databases often include Medicare and Medicaid populations covered through managed care arrangements. However, these existing datasets are often unaffordable for researchers at institutions with fewer resources, and the voluntary nature of these databases means that the continued participation of payers (or their continued availability to researchers) is subject to uncertainty. Creating a lower-cost alternative option could enable more-equitable access to multi-state claims data and enable patient-centered outcomes research on questions that might be challenging to address with existing claims databases.

Conclusions and Next Steps

APCDs can be used to conduct patient-centered outcomes research that addresses a range of questions important to patients, caregivers, clinicians, and policymakers. The use of data from multiple state APCDs may further facilitate these studies by providing control groups to evaluate state policies and allowing the examination of population subgroups and rare conditions. Though current challenges related to missing data, missing populations, and differing APCD data limit broad multi-state uses of APCDs, stakeholders have identified a number of opportunities to address these challenges at the state and federal levels.

Appendix A. Semistructured Discussion Guides

ASPE contracted RAND to conduct a series of discussions with stakeholders in the spring and summer of 2023 covering challenges and strategies for using APCDs for patient-centered outcomes research and multi-state research—with a focus on missing data, missing populations, and data linkages.

One-on-one discussions were held with state APCD leaders in six states, four data vendors that provide services to state APCDs, and four independent researchers from academic and nonprofit organizations who have conducted research using multiple state APCDs. In addition, two one-hour group discussions with participants from five states were held to discuss linking APCD data to additional data sources, such as vital records, hospital discharge data, and cancer registries.

Full details about the methods used for this project are described in Chapter 1. The discussion guides used in this project are included below.

Initial State Discussion Guide

1. Is your APCD data currently being used for any multi-state research or use cases, either by your agency or outside researchers? Has there been any multi-state work in the past?
2. In thinking about analyses that may use APCD data from multiple states, what types of research questions or "use cases" might be most interesting to you and your state?
    a. Are there issues of mutual policy interest to people served by federal and state health programs that lend themselves to research using APCDs?
    b. In your opinion, what issues would be of most interest to inform patient or provider decisionmaking?
    c. [If time allows] In your opinion, how can APCD data be used to generate new evidence on the safety, effectiveness, and patient outcomes associated with interventions used in health care?
    d. [If time allows] In your opinion, how can APCDs be used to improve knowledge about new treatment and technologies used in health care?
3. Are there challenges in using your state's APCD for a multi-state analysis today? We're curious about data gaps—for example, are there data that are missing or not comparable to other states that may limit a multi-state analysis?
    a. [If yes] What would be needed to make cross-state comparisons more possible?
    b. [If yes] How can gaps be filled? Can existing data be standardized and major gaps filled to allow cross-state comparisons?
    c. [If time allows] Any ideas about what methods would be useful for studying patient outcomes over time, across settings, and across states given these data challenges?

4. To what extent do missing data raise concerns about APCD research findings or the value of APCDs? For example, limited data from ERISA plans and Medicare.
    a. Has your state taken steps to address missing data challenges in the APCD? If yes, which problems have you tried to address, and what steps have been implemented? What steps have been planned or discussed, but not yet implemented?
    b. What can be done to facilitate data submission by health plans covered under ERISA?
5. What are strategies to improve data around equity? This may include race/ethnicity, rural residence, gender identity, preferred language, disability status, and income.
6. To what extent can you follow individuals within your state longitudinally across payers? To what extent can you track individuals over settings of care?
    a. Do limitations in your ability to observe individuals longitudinally across payers restrict the value of the APCD for answering certain research questions?
    b. [If limitations acknowledged] What has been done or could be done in your state to address this limitation?
7. [If time allows] Are there specific policies (in your state or at the federal level) that present barriers to improving data quality and completeness?
8. Do you think a common data application could facilitate research using APCD from multiple states? By a common data application, we mean a single data request form that could be submitted by external researchers to request data from multiple states simultaneously.
    a. [If yes] In what ways?
    b. [If no] Why not?
    c. What features would make a common data application more valuable for your state?
9. What else could be done to facilitate research using APCD data from multiple states?

Additional Items Included in Researcher Discussion Guide

Discussions with researchers utilized all the questions from the Initial State Discussion Guide, rephrased to be relevant to researchers, plus the following questions:

1. What challenges have you encountered in conducting multi-state analyses using APCDs?
    a. Potential probes:
        i. Obtaining data?
        ii. Cleaning data?
        iii. Analyzing data?
        iv. Missing populations? Examples include lack of data from Federal Employees Health Benefits (FEHB) Program and ERISA plans.
        v. Missing data? Examples include race/ethnicity, preferred language, gender identity.
        vi. Following individuals over time and across settings?

vii. [if time allows/topic arises] Varying data quality across states?

b. Probe: Were challenges consistent or different across states?

2. In thinking about analyses that may use APCD data from multiple states, what types of research questions or "use cases" might be most interesting to you?
    a. In your opinion, what issues would be of most interest to inform patient or provider decisionmaking?
    b. In your opinion, how can APCD data be used to generate new evidence on the safety, effectiveness, and patient outcomes associated with interventions used in health care?
    c. In your opinion, how can APCDs be used to improve knowledge about new treatment and technologies used in health care?

Additional Items Included in Vendor Discussion Guide

Discussions with data vendors utilized all the questions from the Initial State Discussion Guide, rephrased to be relevant to data vendors, plus the following questions:

1. Which state APCDs are you working with and in what capacity?
    a. [Probe as needed] What kind of services do you provide?
    b. [Probe as needed] Approximately how many/which states do you work with?

Data Linkages Discussion Guide

1. What types of data do you currently link to APCD data?
2. Can you tell me a little bit about the data linkage process?
3. Why did you move forward with this linkage/these linkages?
4. What were the key challenges in linking these data?
5. What are the key challenges in using these linked data?
6. What do you see as the "value-added" for linking these data? In other words, what can you do with the linked data that you cannot do with APCD data alone?
7. How have these linked data been used either by your agency, other agencies in your state, or by outside researchers? Have the linked data been used in any multi-state research studies?
8. In thinking about analyses that may use linked APCD data from multiple states, what types of research questions or "use cases" might be most interesting to you and your state?
9. How might data linkage be used to address health inequities in your state? This may include inequities around race/ethnicity, rural residence, gender identity, preferred language, disability status, and income.
10. Are there any resources or activities that you think could help states harmonize their efforts to link APCD data with other sources?
11. What future efforts to link APCDs to other data sources are planned by your state?

Appendix B. Use Case 1: Out-of-Pocket Spending on Insulin and Total Costs of Care for Insulin Users

Purpose

The purpose of this use case is to provide guidance for examining OOP spending on insulin, as well as overall OOP spending and the total cost of care (including medical spending), for diabetes patients who use insulin ("insulin users"), using one or more state APCDs. This use case is not intended to provide step-by-step directions for someone new to working with APCD data, and some necessary details, such as code lists, are omitted. Rather, this use case is intended to be used as a starting point for individuals interested in understanding how APCDs may be used.

Research Questions

The specifications presented in this document can be used to explore several questions, including the following:

1. What is the average OOP price paid for a 30-day supply of insulin?
2. What is the yearly OOP cost for insulin users associated with utilization of
    a. insulins?
    b. all diabetes drugs?
    c. all prescription drugs?
    d. total OOP costs (medical + prescription)?
3. What is the yearly total spending (OOP + plan paid amounts) for
    e. all prescription drugs?
    f. total cost of care (medical + pharmacy)?

Background and Motivation

Insulin prices have risen in recent decades, and affordability challenges are now widespread even among patients with health insurance. Besides patient financial burden, high OOP costs for insulin are a concern because they are associated with worse adherence and higher rates of preventable hospitalizations and emergency department (ED) visits. OOP spending on insulin is therefore a timely example of economic burden as a key patient outcome, as defined in the 2019 reauthorization of OS-PCORTF (ASPE, 2022).

Not all insured patients are exposed to high OOP costs for insulin, and average OOP spending on insulin for commercially insured patients has risen much less than list prices and total payments (Laxy et al., 2021). However, high OOP costs are widespread among commercially insured patients with high-deductible health plans (HDHPs) or coinsurance-based cost-sharing, as well as among Medicare beneficiaries with Part D coverage (Cefalu et al., 2018). Paying for insulin can also be a major financial burden for lower-income patients across many insurance types.

Value of Using APCDs to Study Insulin OOP Spending

Much of the existing evidence on insulin affordability challenges and their consequences either uses nationwide data that do not readily allow state-level analysis or is limited to the Medicare beneficiary population.[2] State APCDs could be used to provide state-specific and substate estimates of the extent of insulin affordability challenges and how their prevalence varies across and within payers. The inclusion in APCDs of medical and pharmacy claims also makes them well suited for analyses describing insulin users' OOP spending and total spending (including amounts paid by health plans) on prescription drugs and medical care more broadly.

In comparison to other data sources (such as household surveys and other public or private claims datasets), state APCD data offer several advantages for studying insulin OOP costs. These include

- inclusion of multiple payers and coverage types from both public and private insurance sources, allowing comparison of OOP spending across coverage types
- inclusion of medical and pharmacy claims for the same individuals, allowing measurement of total costs of care (in contrast to datasets that contain only pharmacy claims)
- the possibility of state-specific estimates with APCD data (with limitations noted below), which are not feasible with all health care datasets
- earlier availability of APCD data with a shorter data lag than household survey data or some commercial claims datasets (although data lags vary widely between states, and not all APCDs have shorter data lags than other data sources).

Insulin affordability has also been a priority of state and federal policymakers in recent years, leading to a number of policy changes and voluntary efforts by the private sector to decrease costs. Examples in Medicare include the Part D Senior Savings Model Test (which began in 2021) and selected provisions of the Inflation Reduction Act (effective in 2023). Voluntary actions by Eli Lilly and other insulin manufacturers are also intended to reduce list prices and limit OOP spending for privately insured patients. State initiatives, including in some states with well-developed APCDs (e.g., Colorado, Utah, Minnesota, and Washington), have also required benefit design changes that limit OOP costs for insulin.[3]

While the specifications described here do not propose a rigorous evaluation of these policy changes, evidence on insulin OOP spending over time may be of value to policymakers interested in whether the problems that have motivated these policy responses remain widespread or have been begun to show signs of improvement. Methods proposed here for using APCD data to measure insulin OOP spending could be used to evaluate some of these interventions in future work. Furthermore, pooling data from multiple states could allow some innovative extensions to the proposed study design, including the definition of a comparison group for statewide policy interventions.

Data and Sample Construction

This analysis will focus on the population of insulin users observed in an APCD, defined as enrollees who fill at least one insulin prescription during a calendar year in the study period. We propose defining two different study populations.

An initial set of analyses (focused on the OOP price for a 30-day supply) will include all insulin prescriptions observed in the APCD, including those filled by patients who have insurance transitions or who are not observed in all months. Estimates of the price per prescription for this sample will address the question "What is the average OOP price paid for insulin among insulin users observed in the APCD?"

However, analysis focused on annual medical and prescription drug spending for insulin users may be difficult to interpret if individuals with part-year coverage or who have insurance transitions during the year are included in the sample. Annual spending measures for those with part-year coverage (i.e., who are unobserved in some months of the year) might be difficult to interpret due to deductibles and other nonlinear features of benefit design, while spending measures for those who are covered for the whole year but switch coverage types or payers will reflect variation in prices and benefit design across insurance plans. We note that OOP costs among those experiencing insurance transitions may also be of interest, but we do not propose an approach to examine this group here.

Analyses addressing our second and third research questions, which focus on annual OOP spending and total costs, will therefore restrict attention to insulin users who have a full year of continuous coverage with the same insurance plan.

Key Measures and Data Elements Needed

This analysis will require numerous variables submitted in the eligibility, medical claims, and pharmacy claims files. Table B.1 provides a list of APCD data elements needed for the proposed analysis, drawing from data elements listed in the APCD-CDL™ produced by the APCD Council, the National Association of Health Data Organizations, and the University of New Hampshire (APCD Council, National Association of Health Data Organizations, and the University of New Hampshire, 2021).

Table B.1. Common Data Layout Data Elements Required for Insulin Use Case

APCD-CDL™ Data Element #	Data Element Name	Needed to Construct:	File
CDLMC002	Payer Code	Linking Files Within Payers and Plans	Medical
CDLMC003	Plan ID	Linking Files Within Payers and Plans	Medical
CDLMC005	Payer Claim Control Number	Identifying/Processing Final Medical Claims	Medical
CDLMC006	Line Counter	Identifying/Processing Final Medical Claims	Medical
CDLMC007	Version Number	Identifying/Processing Final Medical Claims	Medical
CDLMC008	Cross Reference Claims ID	Identifying/Processing Final Medical Claims	Medical
CDLMC032	Type of Bill – Institutional	Medical OOP, Total Cost of Care	Medical

APCD-CDL™ Data Element #	Data Element Name	Needed to Construct:	File
CDLMC087	Revenue Code	Medical OOP, Total Cost of Care	Medical
CDLMC119	Date of Service – From	Medical OOP, Total Cost of Care	Medical
CDLMC120	Date of Service – Thru	Medical OOP, Total Cost of Care	Medical
CDLMC121	Service Units – Quantity	Medical OOP, Total Cost of Care	Medical
CDLMC122	Unit of Measure	Medical OOP, Total Cost of Care	Medical
CDLMC125	Plan Paid Amount	Total Cost of Care	Medical
CDLMC126	Co-Pay Amount	Medical OOP, Total Cost of Care	Medical
CDLMC127	Coinsurance Amount	Medical OOP, Total Cost of Care	Medical
CDLMC128	Deductible Amount	Medical OOP, Total Cost of Care	Medical
CDLMC129	Other Insurance Paid Amount	Medical OOP, Total Cost of Care	Medical
CDLMC130	COB/TPL Amount	Medical OOP, Total Cost of Care	Medical
CDLMC131	Allowed Amount	Medical OOP, Total Cost of Care	Medical
CDLMC133	Drug Code	Insulin OOP (for Insulin Pumps)	Medical
CDLMC156	Type of Claim	Hospital Utilization	Medical
CDLMC157	Claim Status	Medical OOP, Total Cost of Care	Medical
CDLMC158	Denied Claim Line Indicator	Medical OOP, Total Cost of Care	Medical
CDLMC159	Claim Adjustment Reason Code	Medical OOP, Total Cost of Care	Medical
CDLMC160	Claim Line Type	Medical OOP, Total Cost of Care	Medical
CDLME002	Payer Code	Plan Type	Eligibility
CDLME002	Payer Code	Linking Files Within Payers and Plans	Eligibility
CDLME003	Plan ID	Linking Files Within Payers and Plans	Eligibility
CDLME004	Member Insurance Product/Category Code	Plan Type	Eligibility

APCD-CDL™ Data Element #	Data Element Name	Needed to Construct:	File
CDLME005	Start Year of Submission	Coverage Indicator (Identify Population of Interest)	Eligibility
CDLME006	Start Month of Submission	Coverage Indicator (Identify Population of Interest)	Eligibility
CDLME007	Insured Group or Policy Number	Linking Files Within Payers and Plans	Eligibility
CDLME018	Member Gender	Demographics (Gender)	Eligibility
CDLME019	Member Date of Birth	Demographics (Age)	Eligibility
CDLME026	Member ZIP Code	Substate Geography	Eligibility
CDLME027	Member FIPS County Code	Substate Geography	Eligibility
CDLME029	Race 1	Demographics (Race)	Eligibility
CDLME030	Race 2	Demographics (Race)	Eligibility
CDLME032	Hispanic Indicator	Demographics (Ethnicity)	Eligibility
CDLME033	Ethnicity 1	Demographics (Ethnicity)	Eligibility
CDLME036	Medical Coverage Under This Plan	Plan Type	Eligibility
CDLME037	Pharmacy Coverage Under This Plan	Plan Type	Eligibility
CDLME040	Primary Insurance Indicator	Plan Type	Eligibility
CDLME041	Coverage Type	Plan Type	Eligibility
CDLME043	Market Category Code	Plan Type	Eligibility
CDLME064	HDHP Indicator	Plan Type	Eligibility
CDLPC002	Payer Code	Linking Files Within Payers and Plans	Pharmacy
CDLPC003	Plan ID	Linking Files Within Payers and Plans	Pharmacy
CDLPC005	Payer Claim Control Number	Identifying/Processing Final Pharmacy Claims	Pharmacy
CDLPC006	Line Counter	Identifying/Processing Final Pharmacy Claims	Pharmacy
CDLPC007	Version Number	Identifying/Processing Final Pharmacy Claims	Pharmacy
CDLPC008	Cross Reference Claims ID	Identifying/Processing Final Pharmacy Claims	Pharmacy
CDLPC009	Insured Group or Policy Number	Linking Files Within Payers and Plans	Pharmacy
CDLPC023	Date Prescription Filled	Identifying/Processing Final Pharmacy Claims	Pharmacy
CDLPC025	Drug Code	Insulin Type	Pharmacy

APCD-CDL™ Data Element #	Data Element Name	Needed to Construct:	File
CDLPC032	Quantity Dispensed	Pharmacy OOP	Pharmacy
CDLPC033	Days' Supply	Pharmacy OOP	Pharmacy
CDLPC034	Drug Unit of Measure	Pharmacy OOP	Pharmacy
CDLPC035	Prescription Number	Pharmacy OOP	Pharmacy
CDLPC037	Plan Paid Amount	Pharmacy OOP, Total Cost of Care	Pharmacy
CDLPC038	Allowed Amount	Pharmacy OOP	Pharmacy
CDLPC039	Sales Tax Amount	Pharmacy OOP	Pharmacy
CDLPC040	Ingredient Cost/List Price	Pharmacy OOP, Total Cost of Care	Pharmacy
CDLPC041	Postage Amount Claimed	Pharmacy OOP, Total Cost of Care	Pharmacy
CDLPC042	Dispensing Fee	Pharmacy OOP, Total Cost of Care	Pharmacy
CDLPC043	Co-Pay Amount	Pharmacy OOP, Total Cost of Care	Pharmacy
CDLPC044	Coinsurance Amount	Pharmacy OOP, Total Cost of Care	Pharmacy
CDLPC045	Deductible Amount	Pharmacy OOP, Total Cost of Care	Pharmacy
CDLPC046	COB/TPL Amount	Pharmacy OOP, Total Cost of Care	Pharmacy
CDLPC047	Other Insurance Paid Amount	Pharmacy OOP, Total Cost of Care	Pharmacy
CDLPC048	Member Self-Pay Amount	Pharmacy OOP, Total Cost of Care	Pharmacy
CDLPC049	Payment Arrangement Type Flag	Pharmacy OOP, Total Cost of Care	Pharmacy
CDLPC065	Record Status Code	Identifying/Processing Final Pharmacy Claims	Pharmacy
CDLPC066	Claim Line Type	Identifying/Processing Final Pharmacy Claims	Pharmacy
N.A. (not in CDL)	Longitudinal ID	Linking Files Across Payers and Over Time	Eligibility, Medical, Pharmacy

NOTE: COB/TPL = coordination of benefits and third-party liability; FIPS = Federal Information Processing Standards.

We also assume the availability of a longitudinal person ID that can be used to link multiple public and private eligibility and claims records belonging to the same individual across

payers (e.g., pharmacy and medical coverage, Medigap and Medicare fee-for-service [FFS] coverage).

Measuring Out-of-Pocket Spending and Total Spending on Prescription Drugs

A patient's OOP liability (patient payment amount) on a prescription can be measured as the sum of deductible, copay, and coinsurance amounts, which are reported in APCD **pharmacy claims files**.

Variables needed:

- deductible amount
- copay amount
- coinsurance amount
- plan paid amount.

Additional pharmacy file variables may also be needed to de-duplicate claims, to select final status claims (in some APCDs), and to link pharmacy files to member eligibility records and medical claims (see Table B.1).

Classifying Prescriptions, Measuring Insulin Supply, and Identifying Insulin Users

The OOP cost measure of primary interest will be the normalized OOP price per 30 days' insulin supply paid by insulin users in a calendar year, defined as in Laxy et al. (2021). The annual OOP measures that sum OOP costs over all prescriptions filled in a year for different sets of prescriptions, including insulins, all diabetes drugs, and all drugs, can also be calculated.

In addition to the cost-related variables listed above, construction of these measures will require variables appearing in APCD **pharmacy claims files**.

Variables needed:

- drug code (NDC)
- days' supply
- date prescription filled.

Days' supply as reported on the pharmacy claim can be used to construct a measure of OOP prices that is standardized to reflect the average price for a 30-day supply, following methods used in Laxy et al. (2021). This measure allows calculation of estimates of average OOP prices that include individuals with part-year coverage or who initiate (or cease) insulin use during a coverage year.

For a subset of patients with full-year coverage, annual prescription drug OOP cost measures can be calculated for different sets of prescription claims:

1. insulins (all insulins)
2. insulins by type (e.g., rapid-acting, short-acting, intermediate-acting, long-acting, mixed, combination, concentrate)
3. diabetes drugs
4. all prescriptions.

Crosswalks from Medispan can be used to assign insulin type based on the NDC, following the taxonomy of insulin types used in RAND's evaluation of the Part D Senior Savings Model Test (Taylor et al., 2022). Identification of additional diabetes drugs can be done using RxNorm. (Medispan is a commercial product that must be licensed, while RxNorm is a publicly available database maintained by the National Library of Medicine.)

Measuring Out-of-Pocket Spending and Total Spending on Medical Care

In addition to pharmacy spending, a measure of total OOP spending that includes both pharmacy and medical care can be constructed. OOP spending (patient payment amounts) associated with medical claims can be constructed using several variables from APCD **medical claims files**.

Variables needed:

- deductible amount
- copay amount
- coinsurance amount
- plan paid amount.

Additional medical file variables may also be needed to de-duplicate claims, to select final status claims (in some APCDs), and to link medical files to member eligibility records and pharmacy claims (see Table B.1).

Measuring Patient Characteristics and Insurance Coverage

Information from APCD **member files** will be used to observe demographics and the substate area of residence for insulin users in our study population.

Variables needed:

- year and month of birth
- gender[4]
- race
- ethnicity
- geography (ZIP code or FIPS county code, depending on availability).

Age, gender, and race and ethnicity may be of interest for describing the state's insulin user population and for conducting within-state subgroup analyses describing variation across groups in OOP spending and total costs of care. Geographic variables may similarly be used for subgroup analyses comparing insulin OOP spending and total costs of care between urban and rural areas within a state (defined by Rural-Urban Continuum codes if ZIP code is available or by metropolitan/nonmetropolitan status if county codes are available).

Assigning Insurance Coverage Type

Information from APCD **eligibility files** can be used to classify the type of insurance coverage held by insulin users and to define the period when insulin users are under observation in the APCD.

Variables needed:

- payer code
- member insurance product/category code
- start year of submission
- start month of submission
- medical coverage under this plan
- pharmacy coverage under this plan
- primary insurance indicator
- coverage type
- market category code
- HDHP indicator.

Using guidance from RAND's prior work for ASPE (Carman and Dworsky, 2020), eligibility file data can be used to assign individuals to one of five broad coverage types:

1. Medicare
2. Medicaid
3. ESI
4. Marketplace coverage
5. other nongroup coverage.

Suggested Approach for Reporting Findings

The following tables could be constructed to illustrate variation in insulin costs from a single state for a single year.

Tables Describing OOP Price per 30 Days' Insulin Supply

Table B.2 below provides the structure of an example table designed to address the first research question, describing the OOP price per insulin prescription for all prescription fills observed in the APCD. Example Table 1 would present averages for the APCD and stratified by type of coverage, while Example Table Shell 2 (which would have a structure parallel to Example Table 1, but with different stratification) presents results for subgroups of insulin users.

- Example Table 1: Average OOP prices for 30-day insulin supply, by type of coverage
  - Sample: all insulin prescription fills in APCD
  - Variables used for stratification (subsamples):
    - type of coverage
  - Statistics to report:
    - mean, standard deviation, and quantiles of OOP price per 30-day supply, averaged over all insulin prescriptions observed in APCD
    - proportion of fills with OOP price per 30-day supply falling into intervals.
- Example Table 2: Average OOP price for 30-day insulin supply: by demographic subgroup
  - Sample: all insulin prescription fills in APCD
  - Variables used for stratification (subsamples):

        - gender and age
        - race and ethnicity (if data sufficiently complete)
        - urban/rural geography
    - Statistics to report: same as Example Table 1.

Tables Describing Annual OOP Costs and Total Cost of Care for Insulin Users

To examine OOP spending and the total cost of care for insulin users, we propose restricting the population of interest to individuals with full-year continuous coverage in the same health plan (i.e., same coverage type and same health plan) with no insurance transitions or gaps in coverage. Example Table 3 (illustrated as Table B.3 below) would report the average, standard deviation, and quantiles of OOP and total spending for insulin users, stratified by type of coverage (as above) to allow for comparisons across payers and facilitate comparisons across states with different sources of health coverage.

- Example Table 3: Annual OOP spending and total cost of care for insulin users, by type of coverage
    - Sample: insulin users with full-year (12 months) continuous coverage
    - Variables used for stratification (subsamples):
        - type of coverage
    - Statistics to report:
        - mean of costs by type:
            - OOP costs
                - insulin OOP costs
                - diabetes drug OOP costs
                - all prescription OOP costs
            - total OOP costs (pharmacy + medical)
                - total costs of care
                - prescription drugs
                - medical care
            - total cost of care (prescription + medical)
        - mean, standard deviation, and quantiles.

[1] In this report, we use the acronym *PCOR* only when "patient-centered outcomes research" is used as a modifier, as in "PCOR studies."

[2] Examples of datasets commonly used in this literature include the Medical Expenditure Panel Survey—Household Component (MEPS-HC) and commercial claims databases that may have limited geographic detail and are unlikely to be representative at the state level, such as Marketscan or Optum.

[3] CO ST § 10-16-151; UT ST § 31A-22-626; WA ST 48.43.780, made permanent by 2023 Wash. Legis. Serv. Ch. 16 (S.S.B. 5729); and MN ST § 62Q.48.

[4] States vary in whether sex or gender is collected; the CDL uses a field for sex.

Table B.2. Example Table Reporting Average OOP Prices for 30-Day Insulin Supply, by Type of Coverage

	Coverage Source					
Statistic	**Medicare**	**Medicaid**	**ESI**	**Marketplace**	**Other Nongroup**	**Total**
Mean OOP price						
(Standard deviation)						
Minimum						
25th percentile						
50th percentile						
75th percentile						
90th percentile						
95th percentile						
Maximum						
Proportion of fills with OOP cost per 30-day supply in range:						
<= $35						
$35.01 to $50						
$50.01 to $100						
> $100						
N prescriptions						
N unique persons						

Table B.3. Example Table Reporting Average Annual OOP Spending and Total Cost of Care for Insulin Users, by Type of Coverage

	Coverage Source					
Statistic	**Medicare**	**Medicaid**	**ESI**	**Market-place**	**Other Nongroup**	**Total**
OOP costs						
Insulin OOP costs						
Diabetes drug OOP costs						
All prescription OOP costs						
Total OOP costs (pharmacy + medical)						
Total costs of care						
Prescription drugs						
Medical care						
Total cost of care (prescription + medical)						
N person-year observations						
N unique persons						

Multi-State Comparisons

We have laid out the tables described above on the assumption that they could be produced using identical specifications in multiple state APCDs. Additional comparisons of mean OOP prices or annual spending across states might be of interest and could be conducted using quantities estimated for the tables listed above.

Anticipated Challenges and Strategies for Addressing Them

We anticipate several challenges that will arise in this use case. In some instances, these are inherent limitations of using claims data to study insulin pricing, while in other instances, the challenges arise specifically in the context of APCDs.

Limitations in Measuring OOP and Total Costs Using Pharmacy Claims Data

Despite the advantages of APCD data for accurately measuring patient cost-sharing on pharmacy claims, APCDs have important limitations for measuring OOP costs paid by insulin users. Key missing information about patient OOP costs includes the following:

1. Insurance premiums—a component of OOP spending that can contribute to patient financial burden and that may be affected by policies that make insurance coverage more generous—are not observed in the CDL, although they are collected by at least one participant state.
2. Coupons and OOP spending not associated with insurance claims are not observed.
3. Patient Affordability Program participation and generosity are not observed.

Because Medicare patients are known to be a population that has faced high OOP costs for insulin, we also note a major limitation of APCD data for studying insulin OOP spending for Medicare beneficiaries in particular:

4. Part D Low-Income Subsidy (LIS) status is not directly observed in APCD submissions.

The Part D LIS, which reduces patient cost-sharing and premium payments in Medicare Part D, is not typically collected in APCD data, meaning that more investigation may be needed to verify that insurers that submit Part D plan data to APCDs distinguish between patient cost-sharing and cost-sharing amounts covered by CMS through LIS. To the extent that insurers report the total of patient cost-sharing and LIS payments in the APCD field for patient cost-sharing, OOP spending on prescription drugs for LIS-eligible Medicare beneficiaries may be overstated in APCD data.

We also note that, like other pharmacy claims data, measures of total drug spending based on the APCD are likely to overstate total prescription drug spending because they do not contain information about manufacturer rebates provided to payers.

APCD Is Not Representative of State Population (Missing Populations)

APCDs typically contain data on the entire population with certain types of public and nongroup coverage but do not capture the entire universe of individuals with health insurance in a state for several reasons, most notably the omission of several federal programs and incomplete data submission from ERISA-regulated ESI plans due to the *Gobeille*

v. Liberty Mutual Insurance Company decision. These limitations and their implications for the use case described here are discussed more fully below.

We also acknowledge that APCD data do not include the uninsured, who are the group most exposed to high OOP spending due to high insulin list prices. Even statewide estimates from an APCD (which may require very strong assumptions) would at best reflect OOP spending for the insured population, not the uninsured.

The population of interest for any analysis using APCD data must be defined to be clear about the applicability of the estimates to insured populations with certain types of insurance. These issues, discussed below, are especially relevant for a multi-state analysis.

Limitations Related to Federal Health Plans and Health Systems

For those with comprehensive health insurance coverage through federal insurance sources (e.g., FEHB and TRICARE), it may be reasonable to redefine the population of interest for the study to exclude individuals with these types of coverage. However, federal health systems (VHA and IHS) that are characterized by high rates of concurrent federal and nonfederal coverage pose a greater challenge for interpreting APCD-based analyses because APCDs do not typically have information about individuals' eligibility for care from these systems.

In the context of insulin pricing, the threat to external validity may be limited if patients with concurrent federal and nonfederal coverage can be assumed to obtain insulin either exclusively through the federal system (in which case they would not appear as insulin users in the APCD) or exclusively through their APCD-covered insurance (in which case the APCD would accurately reflect their OOP spending on insulin). We think it is unlikely that this assumption would hold in practice, but auxiliary analyses using the MEPS-HC or similar household survey data could be used to evaluate the validity of this assumption.

Limitations Related to ERISA Plans

The challenge posed by ERISA plans is different in nature because it affects the completeness of APCD data on the state's population with ESI coverage (which is the most prevalent source of insurance coverage in the United States). ERISA plans may differ systematically from fully insured plans (which are generally mandated to submit data to APCDs), in which case it may be invalid to extrapolate from ESI plans observed in APCD data to the rest of the ESI sector.

Potential Approaches to Addressing Limitations

Despite these limitations, it may still be of interest to report a single number summarizing the OOP price of insulin for the state's insured population. Under the very strong assumption that the prevalence of insulin use and the OOP price is identically distributed between individuals with ERISA and non-ERISA ESI plans within a state after conditioning on age, gender, and metropolitan status, it would be possible to extrapolate OOP prices from APCD data to the ERISA-regulated segment of the ESI market in order to produce an estimate of the average OOP price in a state for ESI-covered individuals.

This extrapolation would require outside data on the joint distribution of type of coverage, age, gender, and metropolitan status within the state (e.g., using the American Community Survey or Current Population Survey Annual Socioeconomic Supplement) to derive analysis weights that could be used to reweight the APCD population to match the demographic and

geographic distribution of individuals with ESI coverage within the state. Similar extrapolations could be done to produce statewide estimates for the population of individuals with APCD-covered insurance sources (Medicaid, Medicare, ESI, Marketplace, and other nongroup) based on outside estimates of the covered population.

The potential advantage of developing such weights is that they could also be used to facilitate multi-state comparisons—e.g., comparing OOP spending on insulin between states with different distributions of insurance sources, age, gender, and metropolitan residence by reweighting both states' data to match a fixed population structure and distribution of coverage types. Such a comparison may be of interest for states interested in benchmarking or for policy evaluation in future work.

A potential challenge in developing such weights for the study of insulin OOP costs is that accurate state-level estimates of diabetes prevalence or insulin utilization rates (defined as the proportion of covered individuals who are insulin users) by type of insurance may not be readily available. More-accurate estimates may be possible in states with state-specific health interview surveys (such as the California Health interview Survey or Colorado Health Access Survey). However, the validity of the strong assumptions needed for this approach would need to be evaluated carefully.

Demographic Variables, Including Race and Ethnicity, Are Often Incomplete

APCD data on race and ethnicity is often incomplete, especially for those with private insurance coverage. This will limit our ability to conduct within-state subgroup analyses and may reduce the comparability of subgroup estimates across states. In states where name and street address are collected, this limitation could potentially be addressed using a validated imputation algorithm, such as the RAND-developed Bayesian Improved Surname Geocoding method (RAND Corporation, undated).

Limited Information About Income and Economic Status

High OOP spending on insulin is especially burdensome for lower-income patients. However, APCD data typically lack information about family income or other individual-level measures of economic status. In the absence of linkages to data that more directly measure economic status (such as earnings records from the state unemployment system or tax return data), APCD data are not ideal for analyses focused on insulin affordability or cost burden (sometimes defined by comparing OOP spending to various measures of monthly income).

Limitations of APCD Specific to a Multi-State Approach

Comparability of Data Elements and Derived Variables Is Uncertain Without Careful Investigation

In general, analysts considering a multi-state comparison of OOP and total costs for insulin users would need to carefully evaluate the comparability across states of data elements used in the analysis to ensure that the same variable measures the same construct in different states. While code sets for diagnoses, procedures, and drug codes are uniform across states, differences in data intake, quality control thresholds applied to submissions, business rules, de-duplication of claims, or other steps in the data pipeline could lead to differences across

states in the relationship between values measured in the APCD and actual costs borne by patients or payers. These issues are discussed further in Chapter 3.

One possible strategy for addressing these challenges might be to reframe the research questions considered in multi-state analyses to focus on changes over time within states, rather than comparing the level of spending in cross-section at a given point in time. Such analyses could be more robust to certain forms of measurement error caused by differences in states' data, such as differences in de-duplication that result in overcounting of costs by a proportional constant amount. Analytic techniques such as linear or generalized linear regression models that include payer fixed effects could then be used to obtain consistent estimates of changes in spending under different and arguably weaker assumptions than might be necessary to draw valid conclusions about differences in the level of OOP costs or total spending.

The value of multi-state analyses that focus on within-state changes over time will ultimately depend on the objectives of the states or research funders. Some questions of interest—e.g., "How much more or less do insulin patients pay in Colorado than in Utah?"—would not be well served by such methods. However, analytic methods that focus on within-state changes may be very well suited to questions about the effects of statewide policies, such as "Did implementation of insulin pricing reforms in Colorado reduce OOP spending by commercially insured insulin users relative to a comparison group of insulin users in Utah?" (In fact, the Agency for Healthcare Research and Quality funded a study in 2022 that will use the Colorado and Utah APCDs to address this question.[1])

Additional Challenges with Demographic Variables

In addition to incomplete race/ethnicity data in many state APCDs, cross-state comparisons that incorporate demographic information may be hampered by differing rates of data completeness. States also vary in the code sets or accepted values used for race and ethnicity, which may further complicate comparisons across states.

As noted above, imputation could be used to assign comparable codes for cross-state comparisons involving subgroup estimates based on race or ethnicity. However, not all states collect or release name and street address data needed to perform reliable imputations of race and ethnicity, so the feasibility of this approach would depend on which states are participating in the analysis.

Differences in APCD Completeness and Missing Populations May Limit Comparability of Outcomes and Estimated Effects Across States

Averages calculated for the APCD population as a whole are likely not comparable across states due to differences in APCD coverage and data submission from different sources, and they cannot be assumed to be representative of the state population.

Comparisons within certain coverage sources may be more appropriate, but the value of such comparisons will depend on the completeness of APCD data, which is likely to vary across types of payers and may differ across states for specific payer types:

- Within-insurance-type comparisons for government payers (Medicare and Medicaid) may be more feasible than comparison within commercial coverage types, but

inquiries with state APCD administrators about the completeness and availability of these data (especially Medicare data) would be important for guiding the interpretation of any such comparisons.

- Within-insurance-type comparisons for Marketplace and other nongroup insurance types are likely possible, but interpretation of such comparisons would need to account carefully for other potential differences in the populations with each coverage type (e.g., state decisions around Medicaid expansion).
- Within-insurance-type comparisons for ESI are likely to be the most problematic due to differences across states in the proportion of ESI enrollees who are covered by ERISA plans, in the extent of voluntary submission by ERISA plans, and in the composition of enrollees in ERISA plans that voluntarily submit.

As discussed above, analysis weights that reweight APCD data to match external coverage estimates could potentially be used to facilitate comparisons across states with differing levels of data completeness for specific types of coverage.

Appendix C. Use Case 2: Total Costs for Individuals with Long COVID

Purpose

The purpose of this use case is to provide guidance for estimating the average total cost of care for a person who has long COVID in 2022 using one or more state APCDs. We specify how to identify a cohort of individuals with long COVID during 2022 and, for more advanced users, offer considerations for how to identify a comparison group of similar individuals without long COVID. We then provide guidance on calculating and comparing health care costs for both groups. This use case is not intended to provide step-by-step directions for someone new to working with APCD data, but rather to be used as a starting point for individuals interested in understanding how APCDs may be used to study long COVID costs.

Background and Motivation

It is estimated that 10 to 40 percent of people infected with COVID-19 develop long COVID (Chen et al., 2022; Cutler, 2022; Logue et al., 2021; Thompson et al., 2022). Long COVID is typically diagnosed 30 to 60 days following a COVID-19 infection and may include respiratory and heart symptoms (e.g., cough and chest pain), neurological symptoms (e.g., headache and brain fog), and fatigue, among other symptoms, that may last weeks, months, or years (U.S. Centers for Disease Control and Prevention, 2022). Understanding the health care costs associated with long COVID is important because of the significant implications of long COVID for health care systems, insurers, and individuals. Furthermore, promoting the use of APCDs for studying long COVID is important for advancing patient-centered outcomes research. Due to the ability to conduct longitudinal analyses using state APCDs, these datasets may offer opportunities for comparing the effectiveness of different interventions for treating long COVID and calculating OOP spending for individuals with long COVID.

Value of Using APCDs to Examine This Topic

Because many APCDs allow researchers to follow unique individuals over time and across settings of care, APCDs offer an opportunity to study health care costs associated with long COVID. APCDs are well suited for this type of study because many allow for the study of individuals as they churn across different types of insurance over time, which enables a lengthy period of follow-up post-diagnosis. For example, a report from Colorado released in December 2022 used the Colorado APCD to examine the number of Coloradans diagnosed with long COVID, finding that women were more likely to experience long COVID than men (Colorado Office of Saving People Money on Health Care, 2022).

Furthermore, APCDs provide detailed information about prices paid for health care services by insurers and patients. Detailed price information in APCDs has been used to examine variation in prices for COVID-19 testing (Colorado Center for Improving Value in Health Care, 2023), variation in primary care across types of health insurance (Oregon Health Authority,

2022), and variation in prices across health plans (Whaley et al., 2020). Thus, APCDs can offer important insights into patient and insurer payments related to long COVID.

Data and Sample Construction

This analysis requires information from eligibility files (e.g., dates of enrollment in medical insurance and the type of insurance [e.g., Medicaid, Medicare, private insurance]), medical claims files (e.g., diagnosis codes, payment amounts), and pharmacy claims files (e.g., payment amounts). It assumes the availability of a longitudinal person identifier that can be used to link multiple eligibility and claims records belonging to the same individual over time and across settings of care. Table C.1 provides information about the files and data elements used in this analysis, based on the APCD-CDL™ from the APCD Council, the National Association of Health Data Organizations, and the University of New Hampshire (APCD Council, National Association of Health Data Organizations, and the University of New Hampshire, 2021).

Costs will be generated using 2022 data. Because the International Classification of Diseases, Tenth Revision (ICD-10) code for post-acute sequelae of COVID-19 (i.e., long COVID) was introduced in October 2021, a study period of October 2021 through the end of December 2022 is recommended to ensure that diagnoses of long COVID are not missed during 2022.

Using eligibility files, a sample limited to individuals continuously enrolled in medical insurance during this period should be identified. To ensure that costs can be calculated consistently, the sample should be limited to individuals with the same type of insurance coverage during this period. The span of insurance enrollment can be identified based on the plan effective date and plan term date, with the insurance type identified using the member insurance/product category code. We suggest focusing on individuals aged 18 years and older because long COVID may present differently or be diagnosed differently in children. The resulting eligibility file should be constructed at the person-month level, so that everyone has 15 rows spanning October 2021 through December 2023. As described in Table C.1, the eligibility file will be used to obtain information on gender,[2] race, ethnicity, and insurance type. In addition, a unique member identifier will be used to link records in the eligibility file to the medical claims and pharmacy claims.

Table C.1. List of Variables Needed to Construct Analytic File

APCD-CDL™ Data Element #	Data Element Name	Description	Derived From
CDLME016	SSID or Member ID	Unique identifier used to follow individuals over time across care settings and across insurers	Eligibility file
CDLME040	Primary Insurance Indicator	Insurance Type/Product Code	Eligibility file
CDLME004	Member Insurance/ Product Category code	Use categories from eligibility file	Eligibility file
CDLME050	Plan Effective Date	Date eligibility started for this member under this plan type	Eligibility file

APCD-CDL™ Data Element #	Data Element Name	Description	Derived From
CDLME051	Plan Term Date	Last continuous day of coverage (date eligibility ended) for this member under this plan	Eligibility file
CDLME018	Sex or Gender	Use categories from eligibility file	Eligibility file
CDLME019	Date of Birth	N/A	Eligibility file
CDLME029	Race	Use categories from eligibility file	Eligibility file
CDLME032	Hispanic Ethnicity	Use categories from eligibility file	Eligibility file
CDLMC007	Version Number	Begins with 0 and is incremented by 1 for each subsequent version of the claim	Medical claims
CDLMC160	Claim Line Type	Report the code that defines the claim line status in terms of adjudication. Valid codes are: O=Original; V=Void; R=Replacement; B=Back Out; A=Amendment; D=Denial.	Medical claims
CDLMC119	Date of Service – From	CCYYMMDD. First date of service for this service line.	Medical claims
CDLMC120	Date of Service – Thru	CCYYMMDD Last date of service for this service line.	Medical claims
CDLMC033	Place of Service	N/A	Medical claims
CDLMC037	Principal Diagnosis	N/A	Medical claims
CDLMC038–CDLMC062	Other Diagnosis, 1–24	N/A	Medical claims
CDLMC125	Plan Paid Amount	Includes all health plan payments and excludes all member payments	Medical claims
CDLMC131	Allowed Amount	Maximum amount contractually allowed	Medical claims
CDLMC126	Co-Pay Amount	Co-payment dollar amount paid for which the individual is responsible	Medical claims
CDLMC127	Coinsurance Amount	The dollar amount of coinsurance for this claim line paid.	Medical claims
CDLMC128	Deductible Amount	The dollar amount for this claim line applied to the deductible.	Medical claims

APCD-CDL™ Data Element #	Data Element Name	Description	Derived From
CDLPC007	Version number	begins with 0 and is incremented by 1 for each subsequent version of the claim	Pharmacy claims
CDLPC066	Claim Line Type	Report the code that defines the claim line status in terms of adjudication. Valid codes are: O=Original; V=Void; R=Replacement; B=Back Out; A=Amendment; D=Denial.	Pharmacy claims
CDLPC023	Date Prescription Filled	N/A	Pharmacy claims
CDLPC037	Plan Paid Amount	Includes all health plan payments and excludes all member payments	Pharmacy claims
CDLPC038	Allowed Amount	Maximum amount contractually allowed	Pharmacy claims
CDLPC043	Co-Pay Amount	Co-payment dollar amount paid for which the individual is responsible	Pharmacy claims
CDLPC044	Coinsurance Amount	The dollar amount of coinsurance for this claim line paid.	Pharmacy claims
CDLPC045	Deductible Amount	The dollar amount for this claim line applied to the deductible.	Pharmacy claims
CDLPC048	Member Self-Pay Amount	Amount paid by member in addition to those listed in CDLPC043, CDLPC044, CDLPC045	Pharmacy claim

NOTE: SSID = Social Security ID.

Considerations for When Additional Data Are Available

Ideally data through the end of 2023 would be extracted to allow a 12-month follow-up period for all individuals with a diagnosis date in 2022. At the time of writing this report, 2023 had not yet ended; thus, we have proposed an approach using the most recent data available. We encourage using data through the end of 2023 and beyond when available. This would allow for a full year of follow-up on all individuals diagnosed with long COVID, as individuals with later diagnosis dates may have different treatment patterns in the early months of their diagnoses.

Identifying Individuals with Long COVID

Individuals with long COVID will be identified using ICD-10 code U09.9. This diagnosis code does not have to be observed as a primary diagnosis for an individual to be considered as having long COVID. Individuals will be identified as having long COVID when ICD-10 code U09.9 is observed in their medical claims during the study period. The definition of long COVID is likely to evolve over time. These specifications are written assuming that long COVID

is an absorbing state (i.e., once an individual is diagnosed with long COVID, they always have long COVID). Users of these specifications should consider whether this definition remains appropriate at the time of their analysis.

To ensure that we are identifying an individual's initial long COVID diagnosis, we will look back to October 2021 when the long COVID diagnosis was first introduced. Table C.2 shows examples of how observation periods are calculated for each individual.

- For individual #1, their first long COVID diagnosis is observed in February 2022. Thus, the observation period for individual #1 spans February 2022 through the end of December 2022 (11 months).
- For individual #2, their initial long COVID diagnosis is observed in November 2021. Because costs will be calculated in 2022, the observation period for individual #2 spans January 2022 through the end of December 2022 (12 months).
- For individual #3, their initial long COVID diagnosis is observed in March 2022. Thus, the observation period for individual #3 spans March 2022 through the end of December 2022 (10 months).

Table C.2. Illustration of Look-Back and Follow-Up Period for Identifying a Long COVID Case

Year	2021			2022											
Month	O	N	D	J	F	M	A	M	J	J	A	S	O	N	D
Person 1					L*	L	L				L	L	L		
Person 2		L*					L	L	L				L	L	L
Person 3						L*	L	L		L	L			L	

NOTE: Gray shading indicates the months included in the observation period for each person. L* indicates when the first long COVID diagnosis was observed. L indicates when long COVID diagnoses were observed in an individual's medical claims.

Additional Considerations and Future Directions for Identifying Long COVID

The World Health Organization defines an individual as having long COVID if they have at least two new-onset persistent symptoms lasting for 60 days after a COVID-19 infection. The U.S. Centers for Disease Control and Prevention defines an individual as having long COVID if they have at least one new-onset persistent symptom lasting for 30 days after a COVID-19 infection. In states with a linkage to an HIE or state laboratory testing data, it may be possible to identify total costs of long COVID dating back to the initial positive COVID-19 lab test result. This could improve the interpretability of the study by potentially isolating new-onset long COVID cases as well as allow a look at the initial COVID costs prior to a long COVID diagnosis. Implementing these specifications using linked APCD and HIE data, however, is out of scope of the analysis described in this use case.

Identifying a Comparison Group

In addition to calculating health care costs for individuals with long COVID, it may also be interesting to compare total annual health care costs for individuals with and without long COVID in order to understand the magnitude of costs that may be associated with long COVID. Long COVID incidence may be associated with factors (e.g., age and gender) that may also predict utilization and spending. We caution that the best approach for identifying an appropriate comparison group will likely vary based on the dataset and as the definition of long COVID evolves.

To provide a benchmark for spending that is relevant for long COVID patients, we will calculate a weighted average total cost of care for a sample of patients that match the long COVID patients on

- age
- gender
- insurance coverage type (defined using guidance from Carman and Dworsky [2020]):
  - Medicare
  - Medicaid
  - ESI
  - Marketplace coverage
  - other nongroup coverage.

The use of a within-payer and within-state comparison group will allow the calculation of total costs of care for long COVID patients as dollar and percentage differences from similar patients without long COVID, facilitating comparison of long COVID cost impacts across states and payers with differing demographics and health care prices.

Considerations for a Comparison Group

Depending on the goals of the analysis and data availability, different comparison groups could be considered.

- **Health care utilizers:** By definition, the group of individuals with long COVID will have health care utilization, but the comparison group may not. Thus, the comparison group would mechanically have lower costs than the long COVID group. Because emerging research suggests that individuals with chronic health conditions may be more likely to develop long COVID (Subramanian et al., 2022), it may be of interest to construct a comparison group of "health care utilizers" to compare how costs differ for individuals with long COVID and individuals who have used health care in the study period. To do so, individuals in the comparison group could be required to have at least one bill for medical care during the study period and could potentially be matched to long COVID patients on the basis of an index date or initial utilization event of a similar type.
- **Individuals hospitalized with COVID:** In addition, individuals who were hospitalized with COVID but did not progress to long COVID may also provide an alternative comparison group. Moreover, because individuals with chronic conditions may be at higher risk of long COVID or experience more complications due to long COVID, it may also be of interest to make comparisons among individuals with chronic conditions with and without long COVID.

- **Individuals with and without COVID:** In states with a linkage to an HIE or state laboratory testing data, it may be possible to construct comparison groups of individuals with (1) long COVID, (2) a prior COVID diagnosis and no long COVID diagnosis, and (3) no long COVID or COVID diagnosis. However, because individuals with mild cases of COVID may not interact with the health systems, these data may not be an appropriate way to identify cases of COVID.

Because data-use agreements typically require that small cell sizes (i.e., <11) are not publicly reported, consideration should be paid to the sample sizes of the comparison group and the group with long COVID, overall and among subgroups.

Estimating Health Care Costs

Costs will be estimated overall and specific to long COVID in each month of 2022. Costs will be estimated using final-action, paid claims. Costs of interest include the following:

- Paid amount: the amount paid to a health care provider per episode by the health plan or employer
- Patient payment: the OOP amount paid by a patient, which can be calculated by summing the deductible, copayment, and coinsurance payment
- Allowed amount: the amount paid to a health care provider per episode, including amounts paid by the health plan and any amounts due from the patient, such as deductibles, copayments, and coinsurance. The allowed amount should be the sum of patient payment and the paid amount.

The specific variable names for these cost variables in the medical claims and pharmacy claims files are listed in Table C.1. To estimate costs specific to long COVID using medical claims, we can sum costs over two sets of claims:

- Sum payments on claim lines that have a primary diagnosis of long COVID in each month (narrower measure of long-COVID-related costs).
- Sum payments on claim lines that have any diagnosis of long COVID in each month (broader measure of long-COVID-related costs).

If a claim spans multiple months (e.g., enters hospital on 01/29/2022 and leaves hospital on 02/05/2022), the costs should be assigned to the month based on when the service was completed (i.e., using the end-of-service date).

Pharmacy claims do not have diagnosis codes, limiting our ability to identify prescriptions written specifically for long COVID. Thus, pharmacy claims will be excluded in calculations of long COVID-specific costs. However, it may be possible to estimate drug spending associated with long COVID by linking medical claims with long COVID diagnoses to pharmacy claims by member ID and date. This linkage would identify prescriptions occurring on the same date as a visit with a long COVID diagnosis. While this may lead to misclassifying some prescriptions as related to long COVID, this could be considered an upper bound of an estimate of drug spending associated with long COVID.

To estimate total costs for individuals in the long COVID and comparison groups, using medical and pharmacy claims:

- Sum payments on all medical claim lines for individuals.

- Sum payments on all pharmacy claim lines for individuals.

When summing these payments for individuals with long COVID, all costs should be included in all months—regardless of whether a diagnosis for long COVID is observed at that time.

It may be of interest to compare costs by place of service (e.g., office versus hospital). This could be accomplished by using the place-of-service variable as follows:

- Office = 11, 17, 50, 71, 72
- Inpatient hospital = 21
- ED = 23
- Urgent care facility = 20.

Because the place-of-service variable may be missing for some claims, additional information may be used to identify the place of service. For example, ED visits can also be identified using revenue center codes (0450–0459 or 0981) and Healthcare Common Procedure Coding System codes (99281, 99282, 99283, 99284, 99285).

Costs will be generated at the level of the person-month. These costs will be averaged for each group (those with and without long COVID) and multiplied by 12 to generate estimated annual costs. In addition to estimating average costs, we also suggest examining costs at the 25th, 50th, and 75th percentiles to illustrate a range of costs incurred. Some individuals with long COVID may be observed in 2022 prior to their diagnosis of long COVID. These months will not be used to calculate average costs for patients with or without long COVID.

Examining Costs Across Subgroups

As illustrated in Table C.3, we suggest examining costs separately by insurance type and within insurance type reporting costs by age group, gender, and race/ethnicity.

Table C.3. Example Table Illustrating Total Costs by Subgroups Among Individuals With and Without Long COVID

Subgroup	Among People With Long COVID	Among People Without Long COVID
Commercial insurance (can report by type: ESI, Marketplace, other nongroup)		
By gender		
By age group		
By race/ethnicity		
Medicare Advantage		
By gender		
By age group		
By race/ethnicity		
Medicare FFS		
By gender		
By age group		
By race/ethnicity		
Medicaid		
By gender		
By age group		
By race/ethnicity		

Multi-State Applications

These specifications can be applied to the analysis of one or more state APCDs. Using data from multiple state APCDs has the potential to improve this analysis. First, these specifications could be used to compare average costs across states. Second, if analyzing data from a single state, there may not be sufficient numbers to generate estimates of small subgroups (e.g., within age, gender, and race). Pooling data across multiple states may allow for the exploration of costs in these small subgroups. Third, this analysis would be strengthened if the analysis were able to follow people who move across states or track care obtained in states different from an individual's home state. While this is not yet the norm, a multi-state APCD that links individual records across states could make this possible. Fourth,

using data from multiple states could allow for the examination of the impact of different state policies. These specifications could be extended to examine the impact of different state policies on coverage of COVID treatment and examine whether those are related to rates of long COVID diagnoses and costs. That said, analyses using more than one state APCD should be approached with care, as they require understanding the similarities and differences of APCDs across states, including the populations included (e.g., varying shares of ERISA plan reporting) and the file layouts (e.g., using the APCD-CDL™).

Anticipated Challenges and Strategies for Addressing Them

Long COVID is defined by a time to presentation and therefore requires an index date. Ideally the index date would be an initial infection date, but that is likely to be challenging to determine using claims data due to the use of at-home testing or no testing at all. Therefore, we use the first appearance of U09.9 to identify individuals with long COVID. Moreover, in some cases, long COVID may be viewed as a "diagnosis of exclusion," where the diagnosis is only given if other diagnoses are first ruled out. Thus, observing an initial diagnosis date is unlikely to be synonymous with the start date of the condition.

Because long COVID is a relatively new condition, some long COVID diagnoses may be missed by clinicians, and thus misdiagnosed or undiagnosed individuals would not be accurately identified. This potential challenge may lessen over time as consensus grows regarding the diagnosis of long COVID. In addition, given the myriad of symptoms ascribed to long COVID, it may be challenging to examine costs specific to long COVID, which relies on clinicians to include a long COVID diagnosis when billing. Thus, comparing total costs for individuals with and without long COVID may be the preferred comparison for now.

One concern with APCD data is the potential lack of external validity because not all insurers submit data to APCDs. Another threat to external validity is that these specifications are focused on individuals who are continuously enrolled in the same type of insurance (e.g., commercial insurance, Medicaid, Medicare FFS, Medicare Advantage) for the duration of the study period, and thus costs may be different for individuals who churn across types of insurance during this time period. Moreover, the populations included in APCD data may vary across states (e.g., the share of ERISA plans submitting to APCDs varies).

Considerations for Interpretation of Results

The definition of long COVID is likely to evolve over time. These specifications are written assuming that long COVID is an absorbing state (i.e., once an individual is diagnosed with long COVID, they always have long COVID). Users of these specifications should consider whether this definition remains appropriate at the time of their analysis.

Finally, these specifications are not intended to be used to provide estimates of the incidence or prevalence of long COVID. APCDs are unlikely to provide an accurate estimate of the incidence or prevalence of long COVID because these datasets do not include information on uninsured populations and because not all insured populations are included in APCDs. For example, APCDs have limited information from ERISA plans and from the FEHB program.

Appendix D. Common Data Application

I. Requester Information

1. Name:
2. Title:
3. Email address:
4. Phone number:
5. Shipping address:
6. Organization name and address:
7. Organization type (select one):
    a. Academic institution
    b. Research organization (not academic institution)
    c. Other nonprofit organization
    d. Health insurer
    e. Health care provider
    f. State agency
    g. Federal agency
    h. Trade association, lobbying group, consortium
    i. Independent consultant
    j. Pharmaceutical, biotechnology, medical product firm
    k. Other:

II. Project Information

8. Primary investigator (PI) name and title:
    a. PI qualifications: Describe previous experience using claims data. This question should be answered by the PI or project manager and should encompass the experience of the entire project team who will be using the data.
    b. Other users' names and titles:
9. Title of project:
10. Project start and end dates:
11. Describe your study background, objectives, aims, and purpose.
12. What is the purpose of your project (select one)?
    a. Research
    b. Health care operations
    c. Public health activities
    d. Other:
13. Provide an overview of your intended use of the data.
    a. Describe how the data will be beneficial to your project. Use quantitative indicators of public health importance where possible. For example:

variation in costs of care; rates of underutilization or overutilization of service; health system performance measures; the effect of public health initiatives, health insurance, etc.

b. Explain why the planned project could not be practicably conducted without access to and use of protected health information (PHI).
c. Please describe your cohort and how it is the minimum necessary to achieve your research objectives. Include estimated cohort size.

14. Describe the intended product or report that will be derived from the requested data.
15. Do you anticipate that the results of your analysis will be published or made publicly available? If yes, how do you intend to disseminate the results of the study (e.g., publication in a professional journal, poster presentation, newsletter, webpage, seminar, conference, statistical tabulation, etc.)?
16. Describe your plans to use or otherwise disclose all payer claims database (APCD) data, or any data derived or extracted from such data, in any paper, report, website, statistical tabulation, seminar, or other setting that is **not** disseminated to the public.
17. Briefly explain why completing this project is in the public interest. Use quantitative indicators of public health importance where possible—for example, numbers of deaths or incident cases; age-adjusted, age-specific, or crude rates; or years of potential life lost. Uses that serve the public interest include but are not limited to health cost and utilization analysis to formulate public policy; studies that promote improvement in population health, health care quality, or access; and health planning tied to evaluation or improvement of state or federal government initiatives.
18. Select the level of detail you are requesting:
    a. Standard de-identified dataset
    b. Limited dataset
    c. Identified dataset
19. Complete the Data Element Dictionary to identify the specific data elements that are required for this project, and attach it to this application. In keeping with the minimum necessary standard established under the Health Insurance Portability and Accountability Act (HIPAA), APCD policy is to release only those data elements that are required to complete your project. Please include a rationale for each data element requested.

III. Data Privacy

20. Explain how your use of data will involve no more than a minimal risk to the privacy of individuals. As part of your response, please address how you will protect the data from improper use or disclosure and ensure that the data will not be reused or disclosed to any other person or entity, except as required by law, for authorized oversight of the research for which the data was requested, or for other research for which the use or disclosure of PHI would be permitted under 45 Code of Federal Regulations 164.512(i)(2)(ii). The response to this element should be

brief and should not represent a comprehensive data security plan. Data security is addressed later in the application.

21. Please describe the techniques you will use to prevent re-identification when findings or outputs result in cell sizes less than 11 (e.g., aggregation, cell suppression, generalization, or perturbation).
22. Do you intend to link or merge APCD data to other data?
    a. If yes, please indicate below the types of data to which APCD data will be linked:
        i. Individual patient-level data (e.g., disease registries, death data)
        ii. Individual provider-level data (e.g., American Medical Association Physician Masterfile)
        iii. Individual facility-level data (e.g., American Hospital Association data)
        iv. Aggregate data (e.g., Census data)
        v. Other (please describe):
    b. If yes, describe the dataset(s) to which the APCD data will be linked, indicate which APCD data elements will be linked, and specify the purpose for each linkage.
    c. For each proposed linkage above, please describe your method or selected algorithm (e.g., deterministic or probabilistic) for linking each dataset. If you intend to develop a unique algorithm, please describe how it will link each dataset.
    d. Attach or provide below a complete listing of the variables from all sources to be included in the final linked analytic file.
    e. Please identify the specific steps you will take to prevent the identification of individual patients in the linked dataset.
23. Has an institutional review board (IRB) reviewed your project?
    a. If yes, a copy of the approval letter and protocol must be included with the application package.
    b. No, this project is not human subject research and does not require IRB review.

IV. Data Security

24. Record-level or derivative data that can be re-identified must be destroyed within 30 days of the end of the data use agreement, in a manner that renders it unusable, unreadable, or indecipherable. What are your plans for destruction of the dataset and any potentially identifiable elements of the data once the data use agreement has expired?
25. Please complete and attach the Data Management Plan Self-Attestation Questionnaire (DMP SAQ) from the Research Data Assistance Center: https://resdac.org/request-form/dmp-saq

Appendix E. Annotated Common Data Application

Introduction

In pursuit of the Office of the Assistant Secretary for Planning and Evaluation's goal of supporting efforts to develop better data resources and inform ongoing work within the U.S. Department of Health and Human Services, the states, and the health research community to build national all payer claims database (APCD) capacity to conduct patient-centered outcomes research, we have developed a prototype common application that could be used to apply for access to data from multiple state APCDs at once, addressing one barrier to research using multiple state APCDs under the status quo. To expedite researcher access to APCD data from multiple states, we developed this draft prototype by reviewing elements that appear in six existing state APCD data applications, the Healthcare Cost and Utilization Project purchase application for state databases, the Research Data Assistance Center's (ResDAC's) data application, and conversations with key stakeholders.

A remaining question to be discussed with states is how and where to submit such an application. One option is for researchers to submit one application to a central processor to obtain data from all participating states; another option is for all states to use the same application but not a central processor, necessitating that researchers complete one application but submit it to as many states' APCDs as needed.

Annotations describing the context or rationale for items appear below relevant items in bold, red text.

I. Requester Information

1. Name:
2. Title:
3. Email address:
4. Phone number:
5. Shipping address:
6. Organization name and address:
7. Organization type (select one):
    a. Academic institution
    b. Research organization (not academic institution)
    c. Other nonprofit organization
    d. Health insurer
    e. Health care provider
    f. State agency
    g. Federal agency
    h. Trade association, lobbying group, consortium
    i. Independent consultant
    j. Pharmaceutical, biotechnology, medical product firm
    k. Other:

The contact information requested here was fairly uniform across the core states' existing APCD data applications. Item 7 (Organization type) represents an amalgam of several items from different states' applications. Some states may have different expectations or scrutiny of projected products based on the type of organization requesting data. Items 6 and 7 could potentially be used to identify conflicts of interest or data requests for commercial uses.

II. Project Information

8. Primary investigator (PI) name and title:
    a. PI qualifications: Describe previous experience using claims data. This question should be answered by the PI or project manager and should encompass the experience of the entire project team who will be using the data.
    Several states ask for PI qualifications, which may help administrators know whether data requestors may need technical assistance.

    b. Other users' names and titles:
9. Title of project:
10. Project start and end dates:
11. Describe your study background, objectives, aims, and purpose.
    Different states asked for a range of information related to the study. This may include a brief 150-word abstract or may include more-extensive explanation of project objectives, brief summary of the literature, specific research question(s), individual specific aims, project methodology, and description of intended products or reports to be derived from the requested data.

12. What is the purpose of your project (select one)?
    a. Research
    b. Health care operations
    c. Public health activities
    d. Other:

    Potential subcategories of this item that were used in some state applications include the following: assess utilization of health care services, observe cost trends, compare providers/health plans, create or enhance a commercial product or service, assess population health.

13. Provide an overview of your intended use of the data.
    a. Describe how the data will be beneficial to your project. Use quantitative indicators of public health importance where possible. For example: variation in costs of care; rates of underutilization or overutilization of service; health system performance measures; the effect of public health initiatives, health insurance, etc.
    b. Explain why the planned project could not be practicably conducted without access to and use of protected health information (PHI).
    c. Please describe your cohort and how it is the minimum necessary to achieve your research objectives. Include estimated cohort size.
14. Describe the intended product or report that will be derived from the requested data.

15. Do you anticipate that the results of your analysis will be published or made publicly available? If yes, how do you intend to disseminate the results of the study (e.g., publication in a professional journal, poster presentation, newsletter, webpage, seminar, conference, statistical tabulation, etc.)?
16. Describe your plans to use or otherwise disclose APCD data, or any data derived or extracted from such data, in any paper, report, website, statistical tabulation, seminar, or other setting that is **not** disseminated to the public.

    The information requested in items 13–16 was commonly covered across the core states' APCD data applications.
17. Briefly explain why completing this project is in the public interest. Use quantitative indicators of public health importance where possible—for example, numbers of deaths or incident cases; age-adjusted, age-specific, or crude rates; or years of potential life lost. Uses that serve the public interest include but are not limited to health cost and utilization analysis to formulate public policy; studies that promote improvement in population health, health care quality or access; and health planning tied to evaluation or improvement of state or federal government initiatives.

    Several states ask some variation of this item. It is not clear how many (if any) use this question to screen out potential commercial uses of their data. If the purpose is to identify requests for data for commercial uses, a more direct question may be preferred.

18. Select the level of detail you are requesting:
    a. Standard de-identified dataset
    b. Limited dataset
    c. Identified dataset

    Not all states provide all three options, or anything besides a fully customized dataset. Some states do not have a standard limited dataset. Some states are limited legislatively in what level of detail they can release. Similarly, there is not a common definition of a standard de-identified dataset or a limited dataset. Given the purpose of this project, the items in this application generally assume that the requester is asking for potentially identifiable information.

19. Complete the Data Element Dictionary to identify the specific data elements that are required for this project, and attach it to this application. In keeping with the minimum necessary standard established under the Health Insurance Portability and Accountability Act (HIPAA), APCD policy is to release only those data elements that are required to complete your project. Please include a rationale for each data element requested.

    While many states do offer a limited dataset option, these generally do not include any potentially identifiable information and may be of limited use for researchers. For custom or research-level datasets, most states ask for individual data element selection, with justification for additional elements beyond what is included in any available limited dataset.

III. Data Privacy

20. Explain how your use of data will involve no more than a minimal risk to the privacy of individuals. As part of your response, please address how you will protect the data from improper use or disclosure and ensure that the data will not be reused or disclosed to any other person or entity, except as required by law, for authorized oversight of the research for which the data was requested, or for other research for which the use or disclosure of PHI would be permitted under 45 Code of Federal Regulations 164.512(i)(2)(ii). The response to this element should be brief and should not represent a comprehensive data security plan. Data security is addressed later in the application.
21. Please describe the techniques you will use to prevent re-identification when findings or outputs result in cell sizes less than 11 (e.g., aggregation, cell suppression, generalization, or perturbation).
    Most states did not include a distinct cell size in their re-identification prevention questions. Two states limited reporting of cell sizes less than 11, and one state limited reporting of cell sizes less than 30.

22. Do you intend to link or merge APCD data to other data?
    a. If yes, please indicate below the types of data to which APCD data will be linked:
        i. Individual patient-level data (e.g., disease registries, death data)
        ii. Individual provider-level data (e.g., American Medical Association Physician Masterfile)
        iii. Individual facility-level data (e.g., American Hospital Association data)
        iv. Aggregate data (e.g., Census data)
        v. Other (please describe):
    b. If yes, describe the dataset(s) to which the APCD data will be linked, indicate which APCD data elements will be linked, and specify the purpose for each linkage.
    c. For each proposed linkage above, please describe your method or selected algorithm (e.g., deterministic or probabilistic) for linking each dataset. If you intend to develop a unique algorithm, please describe how it will link each dataset.
    d. Attach or provide below a complete listing of the variables from all sources to be included in the final linked analytic file.
    e. Please identify the specific steps you will take to prevent the identification of individual patients in the linked dataset.

    About half of the applications reviewed asked about data linkages. This item is comprehensive to include the linkage information requested by those states. One state strongly prefers to conduct any data linkages in house prior to providing the dataset in order to prevent potential re-identification by linkage.

23. Has an institutional review board (IRB) reviewed your project?
    a. If yes, a copy of the approval letter and protocol must be included with the application package.

b. No, this project is not human subject research and does not require IRB review.

IV. Data Security

24. Record-level or derivative data that can be re-identified must be destroyed within 30 days of the end of the data use agreement in a manner that renders it unusable, unreadable, or indecipherable. What are your plans for destruction of the dataset and any potentially identifiable elements of the data once the data use agreement has expired?
Every state asked for some confirmation of data destruction plans.

25. Please complete and attach the Data Management Plan Self-Attestation Questionnaire (DMP SAQ) from ResDAC: https://resdac.org/request-form/dmp-saq
The amount of information about data security plans requested by states varied widely. Some states were less strict, and some states were stricter. The DMP SAQ is extensive; it is 24 pages long and covers access controls, auditing and accountability, security assessments, incident response, media protection and security, physical and environmental controls, risk assessments, and privacy controls, among others. The ResDAC questionnaire is offered as a potential example because it is comprehensive and is already used to access state Medicaid data.

Abbreviations

APCD	all payer claims database
APCD-CDL™	All Payer Claims Database Common Data Layout
ASPE	Office of the Assistant Secretary for Planning and Evaluation
CDL	common data layout
CFR	Code of Federal Regulations
CMS	Centers for Medicare & Medicaid Services
COB/TPL	coordination of benefits and third-party liability
COVID	coronavirus disease
COVID-19	coronavirus disease 2019
DMP SAQ	Data Management Plan Self-Attestation Questionnaire
ED	emergency department
ERISA	Employee Retirement Income Security Act
ESI	employer-sponsored insurance
FEHB	Federal Employees Health Benefits
FFS	fee-for-service
FIPS	Federal Information Processing Standards
HCUP	Healthcare Cost and Utilization Project
HDHP	high-deductible health plan
HHS	U.S. Department of Health and Human Services
HIE	health information exchange
HIPAA	Health Insurance Portability and Accountability Act
ICD-10	International Classification of Diseases, Tenth Revision
IHS	Indian Health Service
IRB	institutional review board
LIS	Low-Income Subsidy
MEPS-HC	Medical Expenditure Panel Survey—Household Component

NDC	National Drug Code
OOP	out-of-pocket
OS-PCORTF	Office of the Secretary Patient-Centered Outcomes Research Trust Fund
PBM	pharmacy benefit manager
PCOR	patient-centered outcomes research
PHI	protected health information
PI	primary investigator
ResDAC	Research Data Assistance Center
SSN	Social Security number
SUD	substance use disorder
VHA	Veterans Health Administration

References

AcademyHealth, "Medicaid Data Learning Network," webpage, July 2023. As of August 7, 2023: https://academyhealth.org/about/programs/medicaid-data-learning-network

APCD Council, "Claims Data Release Rules," webpage, undated-a. As of August 7, 2023: https://www.apcdcouncil.org/claims-data-release-rules

APCD Council, "Interactive State Report Map," webpage, undated-b. As of August 7, 2023: https://www.apcdcouncil.org/state/map

APCD Council, National Association of Health Data Organizations, and the University of New Hampshire, "APCD Common Data Layout," webpage, February 2021. As of August 7, 2023: https://www.apcdcouncil.org/apcd-common-data-layout-apcd-cdl

ASPE—*See* Office of the Assistant Secretary for Planning and Evaluation.

Burke, Laura G., Xiner Zhou, Katherine L. Boyle, E. John Orav, Dana Bernson, Maria-Elena Hood, Thomas Land, Monica Bharel, and Austin B. Frakt, "Trends in Opioid Use Disorder and Overdose Among Opioid-Naive Individuals Receiving an Opioid Prescription in Massachusetts from 2011 to 2014," *Addiction*, Vol. 115, No. 3, 2020.

Carman, Katherine Grace, and Michael Dworsky, *Transitions in Coverage: Colorado 2011–2017 (HP-HAC-3-3a)*, RAND Corporation, PR-A1552-2, 2020.

Carman, Katherine Grace, Michael Dworsky, Sara E. Heins, Nabeel Qureshi, Daniel Schwam, Shoshana R. Shelton, and Christopher M. Whaley, *State All Payer Claims Databases: Understanding the Current Landscape and Challenges to Use*, RAND Corporation, PR-A1857-1, May 2022.As of August 7, 2023: https://aspe.hhs.gov/reports/all-payer-claims-data-bases-2022

Cefalu, William T., Daniel E. Dawes, Gina Gavlak, Dana Goldman, William H. Herman, Karen Van Nuys, Alvin C. Powers, Simeon I. Taylor, and Alan L. Yatvin, "Insulin Access and Affordability Working Group: Conclusions and Recommendations," *Diabetes Care*, Vol. 41, No. 6, 2018.

Chen, Chen, Spencer R. Haupert, Lauren Zimmermann, Xu Shi, Lars G. Fritsche, and Bhramar Mukherjee, "Global Prevalence of Post COVID-19 Condition or Long COVID: A Meta-Analysis and Systematic Review," *Journal of Infectious Diseases*, Vol. 226, No. 9, 2022.

Colorado Center for Improving Value in Health Care, "COVID-19 Testing Price Variation," webpage, 2023. As of May 25, 2023: https://civhc.org/covid-19/covid-19-testing-price-variation/

Colorado Office of Saving People Money on Health Care, *Report on Long COVID in Colorado*, December 2022.

Cutler, David M., "The Costs of Long COVID," *JAMA Health Forum*, Vol. 3, No. 5, 2022.

Haakenstad, Annie, Summer Sherburne Hawkins, Lydia E. Pace, and Jessica Cohen, "Rural-Urban Disparities in Colonoscopies After the Elimination of Patient Cost-Sharing by the Affordable Care Act," *Preventive Medicine*, Vol. 129, 2019.

Haas, Ann, Marc C. Elliott, Jacob W. Dembosky, John L. Adams, Shondelle M. Wilson-Frederick, Joshua S. Mallett, Sarah Gaillot, Samuel C. Haffer, and Amelia M. Haviland, "Imputation of Race/Ethnicity to Enable Measurement of HEDIS Performance by Race/Ethnicity," *Health Services Research*, Vol. 54, No. 1, 2019.

Laxy, Michael, Ping Zhang, Stephen R. Benoit, Giuseppina Imperatore, Yiling J. Cheng, Edward W. Gregg, Shuang Yang, and Hui Shao, "Trends in Total and Out-of-Pocket Payments for Insulin Among Privately Insured U.S. Adults with Diabetes from 2005 to 2018," *Diabetes Care*, Vol. 44, No. 10, 2021.

Little, Roderick J. A., and Donald B. Rubin, *Statistical Analysis with Missing Data*, 3rd ed., Vol. 793, Wiley, 2019.

Logue, Jennifer K., Nicholas M. Franko, Denise J. McCulloch, Dylan McDonald, Ariana Magedson, Caitlin R. Wolf, and Helen Y. Chu, "Sequelae in Adults at 6 Months After COVID-19 Infection," *JAMA Network Open*, Vol. 4, No. 2, 2021.

McAvey, K., "Realizing the Promise of All Payer Claims Databases: A Federal and State Action Plan," Manatt, December 2022.

McCarthy, Douglas, *State All-Payer Claims Databases: Tools for Improving Health Care Value—Part 1: How States Establish an APCD and Make It Functional*, the Commonwealth Fund, December 2020.

Miller, Patrick B., Denise Love, Emily Sullivan, Jo Porter, and Amy Costello, "All-Payer Claims Databases: An Overview for Policymakers," University of New Hampshire, 2010. As of August 10, 2023:
https://scholars.unh.edu/cgi/viewcontent.cgi?article=1127&context=ihpp

NIH RePORTER, "Effect of a State-Level Co-Payment Cap for Insulin on Utilization and Glycemic Control," Project Number 1R03HS029202-01, webpage, undated. As of September 21, 2023:
https://reporter.nih.gov/search/rlURWlyhoUeIHARd0BQYLw/project-details/10571244

Office of the Assistant Secretary for Planning and Evaluation, *Building Data Capacity for Patient-Centered Outcomes Research, Office of the Secretary Patient-Centered Outcomes Research Trust Fund Strategic Plan: 2020–2029*, U.S. Department of Health and Human Services, September 2022. As of August 7, 2023:
https://aspe.hhs.gov/os-pcortf-strategic-plan-2020-2029

Office of the Secretary, Department of Health and Human Services, "Confidentiality of Substance Use Disorder (SUD) Patient Records," *Federal Register*, Vol. 87, No. 231, December 2, 2022.

Oregon Health Authority, *Primary Care Spending in Oregon 2020*, May 2022.

RAND Corporation, "RAND Bayesian Improved Surname Geocoding," webpage, undated. As of August 10, 2023:
https://www.rand.org/health-care/tools-methods/bisg.html

Robert Wood Johnson Foundation, "Health Data for Action (Data Access Award): Leveraging Health Data for Actionable Insight," June 2023. As of August 10, 2023: https://anr.rwjf.org/viewCfp.do?cfpId=1696&cfpOverviewId=

Silva, Gabriella C., Amal N. Trivedi, and Roee Gutman, "Developing and Evaluating Methods to Impute Race/Ethnicity in an Incomplete Dataset," *Health Services and Outcomes Research Methodology*, Vol. 19, No. 2–3, 2019.

Sorbero, Melony E., Roald Euller, Aaron Kofner, and Marc N. Elliott, *Imputation of Race and Ethnicity in Health Insurance Marketplace Enrollment Data, 2015–2022 Open Enrollment Periods*, RAND Corporation, RR-A1853-1, 2022.As of August 10, 2023: https://www.rand.org/pubs/research_reports/RRA1853-1.html

State All Payer Claims Databases Advisory Committee, *State All Payer Claims Databases Advisory Committee Report with Recommendations Under Section 735 of the Employee Retirement Income Security Act of 1974*, 2021. As of August 10, 2023: https://www.dol.gov/sites/dolgov/files/ebsa/about-ebsa/about-us/state-all-payer-claims-databases-advisory-committee/final-report-and-recommendations-2021.pdf

Steenland, Maria, Anna Sinaiko, Amy Glynn, Therese Fitzgerald, and Jessica Cohene, "The Effect of the Affordable Care Act on Patient Out-of- Pocket Cost and Use of Preventive Cancer Screenings in Massachusetts," *Preventive Medicine Reports*, Vol. 15, 2019.

Subramanian, Anuradhaa, Krishnarajah Nirantharakumar, Sarah Hughes, Puja Myles, Tim Williams, Krishna M. Gokhale, Tom Taverner, Joht Singh Chandan, Kirsty Brown, Nikita Simms-Williams, et al., "Symptoms and Risk Factors for Long COVID in Non-Hospitalized Adults," *Nature Medicine*, Vol. 28, No. 8, 2022.

Taylor, Erin Audrey, Dmitry Khodyakov, Christine Buttorff, Stacie B. Dusetzina, Preethi Rao, Zachary Predmore, Matthew Cefalu, Catherine E. Cooke, Asa Wilks, Shiyuan Zhang, Sarah Dalton, and Monique Martineau, *Part D Senior Savings Model: Model Reach and Scope*, Centers for Medicare & Medicaid Services, RAND Corporation, EP-69078, 2022. As of July 25, 2023: https://www.rand.org/pubs/external_publications/EP69078.html

Thompson, Ellen J., Dylan M. Williams, Alex J. Walker, Ruth E. Mitchell, Claire L. Niedzwiedz, Tiffany C. Yang, Charlotte F. Huggins, Alex S. F. Kwong, Richard J. Silverwood,et al., "Long COVID Burden and Risk Factors in 10 UK Longitudinal Studies and Electronic Health Records," *Nature Communications*, Vol. 13, No. 3528, 2022.

U.S. Centers for Disease Control and Prevention, "Long COVID or Post-COVID Conditions," webpage, updated December 16, 2022. As of May 8, 2023: https://www.cdc.gov/coronavirus/2019-ncov/long-term-effects/index.html

Whaley, Christopher M., Brian Briscombe, Rose Kerber, Brenna O'Neill, and Aaron Kofner, *Nationwide Evaluation of Health Care Prices Paid by Private Health Plans: Findings from Round 3 of an Employer-Led Transparency Initiative*, RAND Corporation, RR-4394-RWJ, 2020.As of July 25, 2023: https://www.rand.org/pubs/research_reports/RR4394.html

[1] See NIH RePORTER (undated) for additional details on this work in progress.
[2] States vary in whether sex or gender is collected; the CDL uses a field for sex.