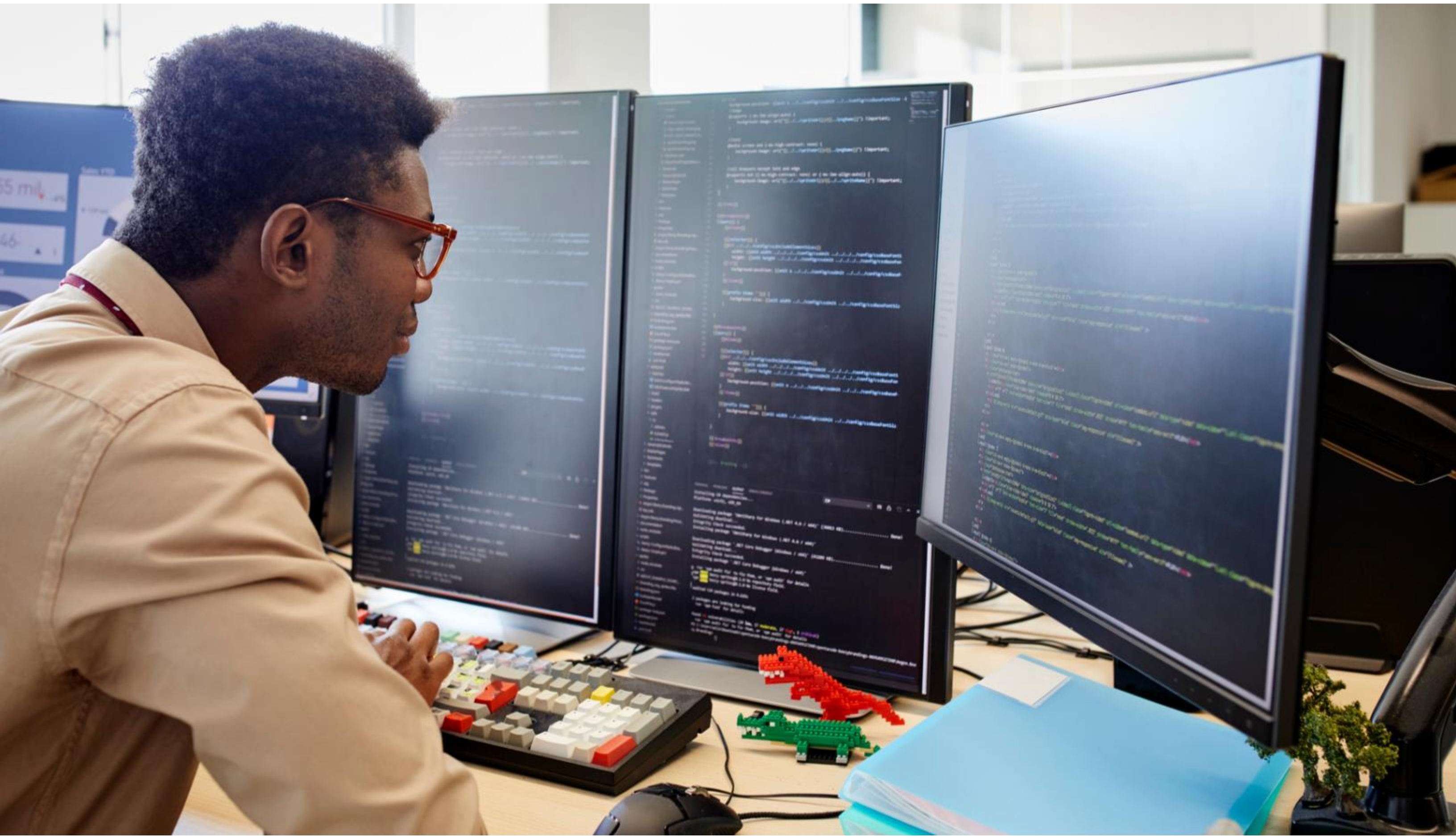
WHITE PAPER

JULY 2022

Expanding Access to Medicare and Medicaid Claims Data Across PCORnet® Clinical Research Networks

Presented by:	Presented to:
NORC at the University of Chicago	Patient-Centered Outcomes Research Institute (PCORI)



Authors

Prashila Dullabh, MD
Krysta Heaney-Huls, MPH
Desirae Leaphart, MPH
Courtney Zott, MPH
Andrew B. Chiao, MPH

NORC at the University of Chicago
4350 East - West Highway, 8th Floor
Bethesda, MD 20814

Acknowledgments

NORC at the University of Chicago would like to express its gratitude to the Clinical Research Networks participating in PCORnet® and the Coordinating Center for PCORnet and subject matter experts that generously offered their expertise and time. We would like to thank Christine Cox for her expert input. We would also like to acknowledge the guidance provided by Claudia Grossmann, Nik Koscielniak, Julianna Paris, Penny Mohr, and Erin Holve on this project.

Suggested Citation

Dullabh, P., Heaney-Huls, K., Leaphart, D., Zott, C., & Chiao, A. B. Expanding Access to Medicare and Medicaid Claims Data Across PCORnet® Clinical Research Networks. Patient Centered Outcomes Research Institute; July 2022. Prepared by NORC at the University of Chicago under Contract No. IDIQ-TO#27-NORC-SCI-AOSEPP-05-24-2021

Disclaimer

All statements, findings, and conclusions in this publication are solely those of the authors and do not necessarily represent the views of the Patient-Centered Outcomes Research Institute (PCORI) or its Board of Governors. This publication was developed through a contract to support PCORI's work. Questions or comments may be sent to PCORI at info@pcori.org or by mail to Suite 900, 1828 L Street, NW, Washington, DC 20036.

Table of Contents

Executive Summary

The Patient-Centered Outcomes Research Institute (PCORI®) is undertaking a multi-step assessment process as it considers data infrastructure enhancements to PCORnet®, the National Patient-Centered Clinical Research Network. PCORnet® ("the Network") is a "network of networks" funded by PCORI® to improve the nation's capacity to conduct definitive health research, particularly comparative clinical effectiveness research (CER).

PCORI plans to prioritize the funding related to PCORnet® in infrastructure enhancements, including the PCORnet® Common Data Model (CDM), that build on the unique capabilities of the PCORnet data structures and align with PCORI's National Priorities for Health[a] adopted in October 2021. The National Priorities for Health include:

1. Increase Evidence for Existing Interventions and Emerging Innovations in Health
2. Enhance Infrastructure to Accelerate Patient-Centered Outcomes Research (PCOR)
3. Advance the Science of Dissemination, Implementation, and Health Communication
4. Achieve Health Equity
5. Accelerate Progress Toward an Integrated Learning Health System.

In alignment with these priorities, PCORI identified three areas of interest for data infrastructure enhancements: 1) social determinants of health (SDOH); 2) patient-reported outcomes (PROs) and other patient-generated health data (PGHD); and 3) Centers for Medicare & Medicaid Services (CMS) claims data. All three opportunities for data infrastructure enhancement have potential to increase data capture, availability, and use across PCORnet® Clinical Research Networks (CRNs).[b] This can facilitate the conduct of definitive national studies that advance evidence-based approaches to improve health outcomes and health equity by informing healthcare decisions and care delivery.

This White Paper is the third in the series developed by NORC at the University of Chicago (NORC) and summarizes NORC's findings on the current state of ***CMS claims data access across PCORnet® CRNs***. Challenges and opportunities to expand such access to claims for PCORnet are addressed, particularly for multi-network studies (i.e., studies involving two or more CRNs). Claims data

---

[a] In 2021, the PCORI Board of Governors also approved a set of Prioritizing Principles for Infrastructure Funding Relating to PCORnet to guide decision-making about PCORI infrastructure funding for the next stages of PCORnet. These Priorities include: "IV. Build on the unique capabilities of the PCORnet data structures, prioritizing investments that will align with the PCORI Strategic Research Priorities" and "Recognize, enable, and promote the value of PCORnet to contribute to a learning health care system through effective partnerships with all stakeholders."

[b] Clinical Research Networks (CRNs) participating in PCORnet consist of two or more partner institutions (e.g., hospitals, integrated delivery systems, federally qualified health centers) who have formed a network. The CRN facilitates data capture, standardization, and data sharing within its own network and has also chosen to join the "network of networks" known as PCORnet.

complements the clinical data held by the CRNs; linking the two can help provide a more complete and longitudinal picture of the patient experience. Improving access to claims data can help enable

PCORI's long-term goals for the Network are to explore the range of multi-network PCOR questions for which claims and clinical data are needed to conduct definitive national studies. More specifically, CMS claims data are highly curated and represent a significant portion of the US population utilizing healthcare services. However, the process for requesting Medicare or Medicaid data from CMS can take up to 3-5 months to complete,[1] and depending on the type and amount of data requested, the costs of the data can be tens of thousands of dollars.[2] Additionally, exchange of CMS data between participating CRNs is not easily supported. Particularly for complex, multi-network studies, claims data access can be challenging to the detriment of conducting research efficiently. Such PCORnet studies have been accomplished in the past (e.g., ADAPTABLE, PCORnet Bariatric Study).

These findings are based on key informant interviews (KIIs) with PCORnet® CRN Principal Investigators, PCORI Board of Governors members, CMS, and a Roundtable on March 30, 2022, which provided PCORnet investigators an opportunity to engage with the Office of Enterprise Data & Analytics (OEDA) at CMS on ongoing and future efforts to improve access to CMS Medicare and Medicaid claims data.

Current State

The KIIs provided crucial insights into the current landscape of both Medicare and Medicaid claims data access across PCORnet. Most CRNs have access to Medicare and Medicaid claims data via a patchwork of approaches and use cases. The approaches CRNs use to facilitate research access to Medicare and Medicaid claims data are summarized below.

Most CRN prime and partner institutions access Medicare data through the Research Data Assistance Center (ResDAC). For CRNs with Medicaid data, all but one CRN requests these data directly from state Medicaid agencies.

Medicare Data Access and Analysis. CRN prime and partner institutions acquire Medicare data using four different approaches which reflect a range of different considerations including research needs and goals:

For the purposes of this White Paper, we define the following terms related to CRNs and health systems participating in PCORnet® as follows:

A **CRN prime institution** is the lead health system of the CRN. The prime institution provides oversight of the CRN in terms of PCORnet activities and is the home institution of the CRN's principal investigator.

A **CRN partner institution** refers to one of multiple health systems, integrated delivery systems, or federally qualified health centers that are a part of the CRN.

A **CRN coordinating site** refers to either a prime or partner institution serving in the role of study coordinator.

PCORnet participant refers to one or more of the health systems participating in PCORnet.

- *CMS claims data linked and stored centrally by CRN prime institution*- A CRN partner institution sends finder files—which identify all the individuals for which the requestor wants Medicare claims data—to either CMS or to their CRN prime institution. The prime institution receives and stores the combined data centrally under a Data Use Agreement (DUA) between the prime institution and CMS. The CRN prime institution makes the combined data accessible to approved researchers.
- *CMS claims data linked and stored locally by CRN partner institutions*- CRN partner institutions send finder files directly to CMS, and the partner institutions receive and store data locally under individual DUAs between the partner institutions and CMS. Data are analyzed locally for approved research studies.
- *CMS claims data received by both CRN prime and partner institutions*- CRN partner institutions send finder files directly to CMS, and both the CRN prime and partner institution receive and analyze data locally under a CMS DUA.
- *CMS claims data accessed through care-focused program*- A CRN partner institution participates in specific CMS programs (i.e., Qualified Entity (QE) Program[c] or Accountable Care Organization (ACO)[d]), which provides access to Medicare claims data that can be used for CMS-approved research purposes.

Medicaid Data Access and Analysis. CRN prime and partner institutions employ three different approaches for acquiring Medicaid data and providing access to researchers. CRNs either:

1. Conduct linkages to national or state Medicaid data centrally at the CRN prime institution which then makes data available to partner institution researchers upon request.
2. Conduct linkages to state Medicaid data centrally at the CRN prime institution and distribute partner institution specific extracts to each partner institution for local storage and analysis.
3. Support both options described above; the CRN prime institution distributes specific extracts to each CRN partner institution for local storage and analysis and maintains aggregate-level data for linkage centrally.

Approaches to Multi-Network Studies. One of the defining features of PCORnet is its national scale, including over 40 health systems across the CRNs participating in PCORnet. Multi-network studies, which include two or more CRNs, leverage this scale and are a priority for PCORI's funding of PCORnet. Three different approaches are used to conduct multi-network studies using claims data within PCORnet:

---

[c] Through the QE program, CRN partner institutions can obtain Medicare Parts A, B, and D data from CMS at a reduced cost. Data is obtained primarily to evaluate provider performance but can be reused for research.

[d] CRN partner institutions can work with ACOs affiliated with their health systems to obtain data, but they must follow each ACO's data access requirements, which can vary by ACO, and the data must be used to support quality assurance.

1. PCORnet participants across multiple CRNs perform distributed analytics in which analyses using claims data are conducted separately at each site. In this approach, each PCORnet partner participating in the study separately requests data for their own sites from CMS. While this is an effective approach to link claims and clinical data, it is inefficient, resulting in duplication of effort and time lag.
2. One PCORnet partner serves as the primary coordinating site and performs data linkage and analyses centrally. The coordinating site refers to a CRN prime or partner institution that has coordinated a single data request to CMS with a finder file representing patient data from all health systems participating in the study. CMS sends the PCORnet participants' claims data (with patient identifiers removed) to the coordinating site for analysis limiting the access of partner institutions to their linked data.
3. The Coordinating Center for PCORnet acts as the primary research site; PCORnet participants may access the data through a secure, virtual connection to conduct the analytics. As reported in option two, this limits the access of partner institutions to their linked data.

Challenges to Expanding CRN Access to Medicare and Medicaid Data from CMS Within PCORnet

Despite many CRNs having experience accessing CMS data, several challenges were reported particularly as they relate to using CMS data for efficient multi-network studies that include a large number of participants and participation by several CRNs. Based on the current state of Medicare and Medicaid claims use across PCORnet® CRNs, NORC identified six salient challenges for expansion of access to CMS data:

Challenge 1: Cost. All key informants from each CRN prime institution reported the cost of Medicare and Medicaid data from CMS as a barrier to data acquisition**.** Many CRNs request data on a study-by-study basis to reduce costs, but this limits the utility of CMS data acquisitions (i.e., reuse of the data) for broader research goals of PCORnet.

Challenge 2: Timeliness. Key informants reported challenges with receiving timely claims data, with some reporting lags of a year or more for Medicare and Medicaid data. Informants from one prime institution also reported timeliness as a challenge to using T-MSIS data, given that these data are not available on a quarterly basis.

Challenge 3: DUA restrictions. CMS DUAs generally allow for only one data recipient (custodian) for each data request and once the data are received, the recipient cannot move the data to other locations.[3] This data use requirement limits CRN prime and partner institutions from distributing CMS claims data to partner institutions within their network or other CRN partner institutions from other networks if conducting a multi-network study.

Challenge 4: DUA processes. Key informants reported that CMS' process for completing DUAs for both initial use and reuse approval is difficult to navigate in a timely manner, introducing delays in CMS data acquisition. Additionally, the process for accessing Medicaid data from state Medicaid agencies differs by state. Navigating individual state DUA processes can create additional burden for CRN partner institutions, particularly those that cover multiple states or do not have established relationships with state Medicaid agencies.

Challenge 5: Privacy. Some PCORnet participants are hesitant to share identifiable data to acquire CMS data. Each coordinating site must navigate their individual partner institutions' preferences on sharing identifiable data with other institutions. This can limit the approaches PCORnet participants use to acquire and analyze CMS claims data. Of note, a key security feature of the PCORnet infrastructure is that the data stay with each Network partner behind its firewall, protected under the Health Insurance Portability and Accountability Act of 1996 (HIPAA) and are not amassed into a single data pool or data warehouse.

Challenge 6: Medicaid data heterogeneity. Medicaid data quality and completeness differ across states. Key informants from CRN prime institutions reported that the heterogeneity of Medicaid claims data across state Medicaid agencies presents data harmonization challenges to using state-based Medicaid data for research.

Opportunities for Expanding CRN Access to Medicare and Medicaid Data from CMS Within PCORnet

During the Roundtable, CMS discussed several ways in which CRN prime and partner institutions can access Medicare and Medicaid data. These opportunities are grouped into two categories: 1) improvements in research data access; and 2) new programs for Medicare and Medicaid data access.

1. Leverage Improvements in CMS Data Access

CMS is engaged in several efforts to improve access to Medicare and Medicaid claims data for research use. CRN prime and partner institutions can maximize the access and use of CMS claims data for research by leveraging the following opportunities:

Opportunity 1: Research Data Request Process Improvements. CMS is working to redesign and improve current processes to access CMS research data, including revamping the existing DUA forms and streamlining the processes for describing data security plans. These improvements can help reduce some of the administrative burdens associated with acquiring and re-using CMS research data.

Opportunity 2: Transformed Medicaid Statistical Information System (T-MSIS) Medicaid Data. CMS has undertaken an effort to transform the collection of Medicaid data from states and improve the analytic utility of national Medicaid data. Obtaining Medicaid data for research through T-MSIS addresses many of the challenges expressed by key informants from CRNs, including reducing the cost, non-standardized state data models, and administrative burden in developing separate agreements with states to acquire Medicaid data.

For CRNs that do not have established DUAs with state Medicaid agencies, acquiring T-MSIS data provides a single path to access standardized Medicaid data for research use. To further support the use of these data for multi-network studies, the Coordinating Center for PCORnet could explore mapping T-MSIS data to the PCORnet® CDM.

Opportunity 3: Explore New Approaches to CMS Data Access Between CRNs and Partner Institutions. CMS offers a number of data access programs that could present new options for PCORnet® CRNs to access Medicare and Medicaid data via CMS. Based on feedback during the Roundtable, CRN attendees most frequently expressed the following “CRN and Partner Institutions Data Access Options” as the most promising opportunity for their CRN to explore.

- *Option 1: Data Access Requests are Coordinated by a Coordinating Site.* Option 1 enables a coordinating site and additional institutions participating in a research study to coordinate submission of a single finder file, including patient identifiers, to CMS. CMS conducts the linkage and returns the combined extract to the coordinating site and partner-specific extracts to each partner institution at a reduced data fee.
- *Option 2: Data Access Requests are Submitted by Individual Institutions Participating in PCORnet.* In the second option, a coordinating site and additional institutions participating in a research study can opt to each send their finder file directly to CMS. CMS compiles the received files, conducts the linkage, then returns the full extract to the coordinating site and partner-specific extracts to each partner institution at a reduced data fee, provided the partner institutions coordinate a timely submission of their finder files to CMS.

Opportunity 4: Leverage the CMS Virtual Research Data Center (VRDC) for Multi-network Studies. The VRDC offers an alternative option to acquiring and analyzing claims data. Utilizing the CMS VRDC environment would allow both single CRN and multi-network studies to conduct consolidated, centralized analyses, which can improve the statistical power of these types of studies. To use the VRDC, CRN partner and/or prime institutions would enter a joint DUA and obtain the desired number of analyst seats in the VRDC. Analysts would securely upload patient-level clinical data. Then CMS would link to the requested claims data and make the full extract of data available in the secure VRDC user space. The overall cost to use the VRDC can be significantly lower than the cost of purchasing physical data, depending on the amount of data requested. In addition, Medicare fee-for-service claims data are available in quarterly updates during the calendar year in which they are submitted for payment. While the VRDC presents a cost-effective and timelier option for accessing claims data for multi-network studies, its use poses other challenges. Health systems participating in PCORnet have expressed hesitancy to share protected health information with external entities that host the data in an environment outside of the health systems’ control.

2. Explore Use of New CMS Programs for Medicare and Medicaid Data Access

CMS has developed a suite of Fast Healthcare Interoperability Resources (FHIR) application programming interfaces (APIs) to access Medicare fee-for-service claims data. FHIR-based APIs allow for data to be accessed more quickly than traditional research data access pathways. These APIs use the Health Level Seven International (HL7) FHIR standard[4] for Medicare claims data rather than the traditional research identifiable file (RIF) file format, which health systems would need to map to the PCORnet® CDM.

Opportunity 5: Data at the Point of Care API Pilot. CMS is piloting the Data at the Point of Care API program, which allows providers to obtain Medicare fee-for-service claims data for their patient population to improve patient health and care delivery. As HIPAA-covered entities, providers participating in the CMS Data at the Point of Care pilot program can incorporate Medicare beneficiary data into their designated record set, which enables them to use the data for other purposes as permitted under the law (e.g., for research with the individual's authorization or an IRB-approved waiver of authorization). This program does not require a CMS research DUA, removing the burden of maintaining these agreements with CMS, and there are no fees paid to CMS to acquire these data. However, because these data sharing programs are focused on providers and efforts to improve quality of care rather than support for research networks, approaches to partnerships and agreements would need to be refocused to work with the care delivery component of CRN institutions.

Opportunity 6: Blue Button 2.0. CMS's Blue Button 2.0 program, which enables beneficiaries to access their own Medicare fee-for-service data and connect it to applications, services, and research programs they trust. Beneficiaries can grant permission for applications, services, and research programs to access their data and can revoke permission at any time. Patient directed data may create opportunities for health systems participating in PCORnet to incorporate Medicare data into patient-enrolled research studies. Beneficiaries can grant permission to applications, services, and research programs for access to their four most current years of data and can revoke permission at any time. While there are no data acquisition fees paid to CMS for data obtained via the Blue Button program, there would be costs associated with the development of technology to deliver and aggregate Blue Button data and the utility of this approach is limited to studies that can elicit consent from participants.

Opportunity 7: Privacy-Preserving Record Linkage. CMS is considering implementing Privacy-Preserving Record Linkage (PPRL) as part of a larger initiative within the U.S. Department of Health and Human Services. The Agency is currently investigating which of the numerous PPRL tools would be the best path forward, though the timeline for selection is unclear.

As most PCORnet participants are familiar with the current CMS research data request process and are currently willing to exchange unencrypted patient identifiers to facilitate linkages, the use of PPRL by CMS may not materially improve access to Medicare or Medicaid claims.

In Figure 1 below, we map the challenges to the opportunities presented in this White Paper.

Figure 1. CMS Claims Data Opportunities Mapped to Challenges

Opportunity	Challenges Identified
Research Data Request Process	Navigating complex CMS DUA process; reducing administrative burden in data acquisition
T-MSIS Medicaid Data	Potential reduction in costs of Medicaid claims data Navigating individual state DUA processes Potential reduction in the variation of quality, timeliness, and administrative burden in obtaining Medicaid data from multiple states
Explore New Approaches to CMS Data Access Between CRNs and Partner Institutions	Potential reduction in costs of Medicare and Medicaid claims data DUA restrictions on distributing claims data
Leverage the CMS VRDC for Multi-network Studies	DUA restrictions on distributing claims data Potential reduction in costs of Medicare and Medicaid data Improvement in latency of Medicare and Medicaid claims data
Data at the Point of Care API Pilot **Blue Button 2.0**	Potential reduction in costs of Medicare claims data Improvement in latency of Medicare claims data
Privacy-Preserving Record Linkage	Data privacy concerns with sharing identifiable data

Conclusions

This White Paper articulates findings on the current state of CMS claims data access across PCORnet® CRNs, along with challenges and opportunities to expand such access, particularly for multi-network studies. Of the opportunities CMS presented to Roundtable attendees, the CRN prime institutions found the following most promising:

1. Leverage new distributed data access options for CRN prime institutions
2. Utilize T-MSIS to access Medicaid data
3. Explore Data at the Point of Care pilot
4. Explore Blue Button 2.0 API

These opportunities for data infrastructure enhancement have potential to increase Medicare and Medicaid data availability and use for national studies that advance evidence-based approaches to improve health outcomes and health equity. The information presented in this White Paper should inform PCORI of the CMS data access opportunities and facilitate discussion with health systems participating in PCORnet and relevant Network governance as they work to expand access to CMS claims data for multi-network studies.

1. Introduction

The Patient-Centered Outcomes Research Institute (PCORI®) is undertaking a series of data convenings for PCORnet®, the National Patient-Centered Clinical Research Network. The convenings comprise a multi-step assessment process to identify potential strategic investment opportunities for enhancements to the PCORnet data infrastructure, including the PCORnet® Common Data Model (CDM). As part of this process, PCORI® contracted with NORC at the University of Chicago (NORC) to explore opportunities to advance the PCORnet data infrastructure in the three areas: 1) social determinants of health (SDOH); 2) patient-reported outcomes (PROs) and other patient generated health data (PGHD); and 3) access to Centers for Medicare & Medicaid Services (CMS) claims data. This White Paper is the third in this series and focuses on opportunities related to CMS claims data access.

Background

PCORnet is a large, nationally representative “network of networks” funded by PCORI to improve the nation’s capacity to efficiently conduct definitive health research, particularly comparative effectiveness research (CER). PCORnet is a national resource that supports efficient, effective, and impactful CER. In addition to identifying strategic investment opportunities for PCORnet, PCORI’s Board of Governors adopted PCORI’s five National Priorities for Health[5] in October 2021 to set new strategic priorities. These National Priorities for Health serve as ambitious long-term goals to guide PCORI’s funding of patient-centered comparative clinical effectiveness research (CER) and other engagement, dissemination and implementation, and research infrastructure initiatives, and include:

1. Increase Evidence for Existing Interventions and Emerging Innovations in Health
2. Enhance Infrastructure to Accelerate Patient-Centered Outcomes Research (PCOR)
3. Advance the Science of Dissemination, Implementation, and Health Communication
4. Achieve Health Equity
5. Accelerate Progress Toward an Integrated Learning Health System.

In early 2021, the PCORI Board of Governors also approved a set of Prioritizing Principles for Infrastructure Funding Relating to PCORnet to guide decision making about PCORI infrastructure funding for the next stages of PCORnet. These Priorities include: “IV. Build on the unique capabilities of the PCORnet data structures, prioritizing investments that will align with the PCORI Strategic Research Priorities,” and, “VII. Recognize, enable, and promote the value of PCORnet to contribute to a learning health care system through effective partnerships with all stakeholders.” PCORI plans to prioritize investments for enhancements to the PCORnet infrastructure that build on the unique capabilities of the PCORnet data structures and align with PCORI’s National Priorities for Health.

The work described in this White Paper aims to inform and advance the National Priorities for Health, with a focus on one in particular: *Enhance Infrastructure to Accelerate PCOR.* Enhancing PCORnet data infrastructure by increasing access to CMS data is critical to improve availability and use of these data across health systems participating in PCORnet and will support PCORI's Phase 3 goal of using PCORnet to conduct definitive national studies that advance PCORI's National Priorities for Health, including efforts to accelerate progress towards a learning health system. Additionally, PCORI's authorizing law outlines direction for CMS to make available claims data while maintaining appropriate privacy safeguards.

> "The Secretary shall, with appropriate safeguards for privacy, make available to the Institute such data collected by the Centers for Medicare & Medicaid Services under the programs under titles XVIII, XIX, and XXI, as well as provide access to the data networks developed under section 937(f) of the Public Health Service Act, as the Institute and its contractors may require to carry out this section."[6]

Phase 3 of PCORnet currently includes eight Clinical Research Networks (CRNs) (ADVANCE Network, Greater Plains Collaborative (GPC), INSIGHT, OneFlorida, PaTH, PEDSnet, REACHnet, and STAR, see Figure 2 and Appendix A) and a Coordinating Center that work together to improve data access, patient engagement, and partnerships for PCOR. Each CRN encompasses multiple health systems, such as hospitals, integrated delivery systems, and federally qualified health centers (FQHCs) (see Appendix A).

Figure 2. Clinical Research Networks (CRNs) participating in PCORnet, December 2021

Source: Developed by the Duke Clinical Research Institute (DCRI) with funding through a PCORI Award (RI-DCRI-01-PS3). Used with permission from PCORI.

Each CRN collectively holds electronic health record (EHR) data for millions of patients. These data are harmonized to the PCORnet® CDM and accessed either through the Coordinating Center for

PCORnet® Front Door or CRNs' front door processes. PCORnet also includes patient partners who participate in all Network governance. Patient partners provide input and perspectives on priority research questions, study design, experiences as participants in research studies, and adoption of research findings in patient care.[7]

The Coordinating Center supports the maintenance of the PCORnet® CDM and an infrastructure to facilitate research. The PCORnet® CDM is one of the resources PCORnet participants use to enhance the usability and "research readiness" of data available via the CRN partner institutions. The PCORnet® CDM standardizes data from the health systems participating in PCORnet® into a consistent format. Data collected by health systems participating in PCORnet through a variety of processes are transformed to conform with the standards defined in the PCORnet® CDM. Once in the PCORnet® CDM, de-identified data are sent to the Coordinating Center for PCORnet and undergo a data curation process and quarterly quality control checks. If the data are approved, they are considered research ready. A key security feature of the PCORnet infrastructure is that the data stay with each network partner behind its firewall, protected under the Health Insurance Portability and Accountability Act of 1996 (HIPAA) and are not amassed into a single data pool or data warehouse.

While nearly all the CRNs have established approaches to link clinical and claims data, all have expressed a desire to improve the efficiency and scalability of their approach. Moreover, PCORnet participating CRNs uniformly see the value of bringing a spectrum of clinical and claims data to key questions prioritized by PCORI's National Priorities for Health.

The Network has demonstrated previous capabilities to access, map, and leverage CMS claims data for research. For example, the Coordinating Center for PCORnet recently mapped CMS claims data concepts to the PCORnet® CDM, describing more than 500 field and value-set relationships.[8] The mapping procedures are available on the Coordinating Center's GitHub page. The main use of Medicare claims data within the Coordinating Center has been to augment data captured by CRN partner institutions as part of pragmatic clinical trials (e.g., EHRs, PROs). One example of this is the ADAPTABLE pragmatic trial, in which claims data was crucial to confirm clinical end-points that were not necessarily captured in the EHR. Another example is the PCORnet Bariatric Study, which linked EHR data to insurance claims

Aspirin Dosing: A Patient-Centric Trial Assessing Benefits and Long-term Effectiveness (ADAPTABLE) Study Use of Medicare Claims Data

For the ADAPTABLE study, researchers within PCORnet used Medicare claims data that were mapped to the PCORnet® CDM to identify and confirm potential end points not captured in the local health systems' HER data warehouses (termed out-of-network events) using the same programming algorithm, such as examining high-dose versus low-dose aspirin for heart attack and stroke prevention among patients with cardiovascular disease.[e]

---

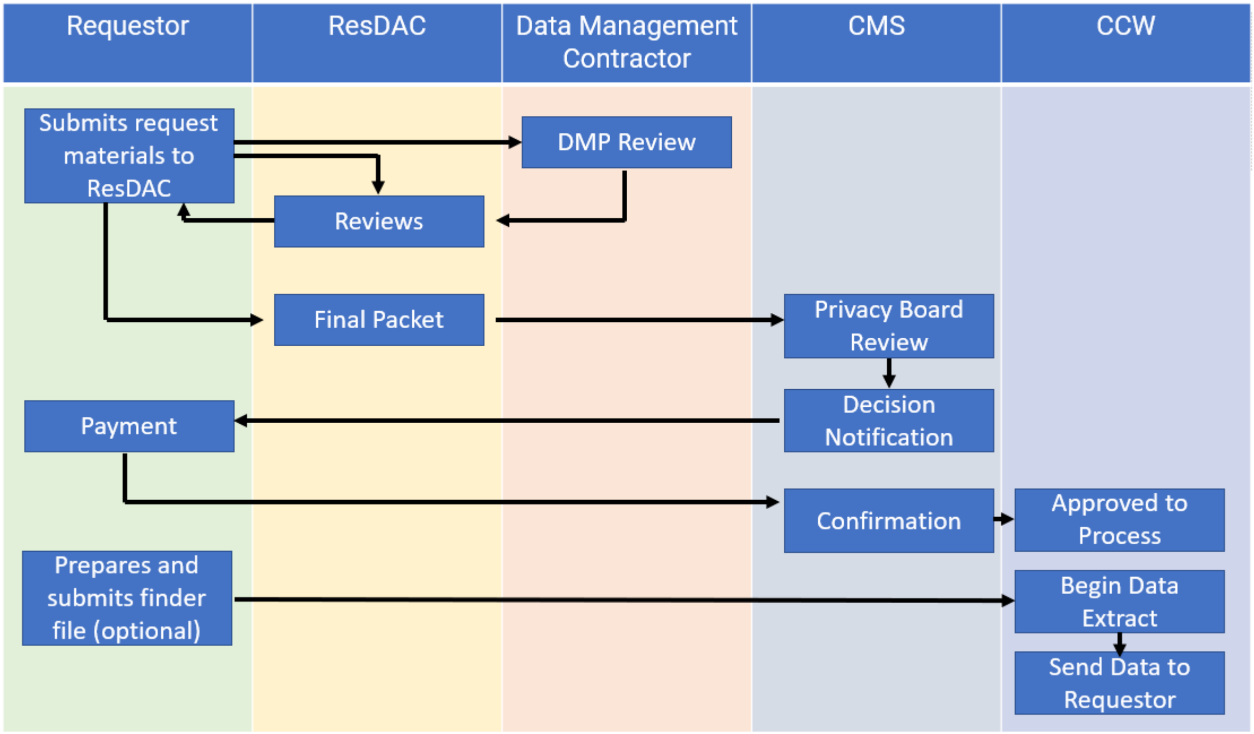
[e] PCORnet-CMS Pilot Linkage Project Process & Results. Duke Clinical Research Institute. Accessed June 22, 2022. https://pcornet.org/wp-content/uploads/2020/06/4275_PCORnet-CMS-Data-Linkage-Team_PCORnet-White-Paper2.pdf

data (including CMS data) and state or national death indices to support more complete capture of major adverse events following bariatric surgery.[9]

Background on Current Mechanisms for CMS Data Access

This section provides a brief background on the current mechanisms for acquiring CMS Medicare and Medicaid claims data which we further detail in Appendix C. To obtain Medicare and Medicaid claims data from CMS, data requestors (i.e., researchers) must complete a DUA that establishes the terms of the use of data for the study and submit this to the Research Data Assistance Center (ResDAC).[10] ResDAC serves as the CMS research contractor to provide academic, non-profit, and government researchers assistance with requesting Medicare and Medicaid claims data, responding to questions or issues with data, and providing data training (see Figure 3).

Figure 3. Overview of Current Process for Requesting Claims Data from CMS



Alternative options for reusing Medicare data obtained through other CMS programs include:

Qualified Entity (QE) program. Through the QE program, participants can obtain Medicare Parts A, B, and D data from CMS at a reduced cost for the primary purpose of evaluating provider performance. Participants can reuse the Medicare data for research. As a QE, a participant must meet the annual requirements of the CMS QE program, including public provider performance reporting.

Accountable Care Organization (ACO) data. Participants of this program can work with ACOs affiliated with their health systems to obtain data, but they must follow each ACO's data access requirements, which can vary by ACO. CMS data obtained through the ACO program include data use restrictions and must be used to support quality assurance.

The process for requesting Medicare or Medicaid data from CMS can take up to 3-5 months to complete,[11] and depending on the type and amount of data requested, the costs of the data can be tens of thousands of dollars.[12] This White Paper investigates some of these challenges as well as others identified through key informant interviews with Principal Investigators from PCORnet® CRNs. Previous efforts within the Network to understand how CRNs participating in PCORnet access CMS data are described in Appendix D.

Roadmap of Paper and Intended Audience

This current White Paper, the third in the PCORnet Data Convening series 1) reviewed the current landscape of how CRNs participating in PCORnet access and use Medicare and Medicaid claims data, and 2) explored potential opportunities for expanding access to such data. We first present the current state of Medicare and Medicaid claims data access and use across CRNs participating in PCORnet, based on a series of interviews with key informants. We then discuss current challenges to accessing and using Medicare and Medicaid data for research. Finally, we present opportunities to address these challenges and improve the accessibility of Medicare and Medicaid claims data across PCORnet. The intended audience for this paper includes PCORI staff, the PCORI Board of Governors, the PCORnet® Steering Committee, the Coordinating Center for PCORnet, and PCORnet® CRNs. NORC's qualitative assessment of Medicare and Medicaid claims data access within PCORnet contributes to PCORI's interest in funding the conduct of PCOR and CER that accurately reflects the outcomes and needs of patients.

2. Methods

Two data collection activities informed the findings synthesized in this White Paper:

1. Key informant interviews with Principal Investigators from PCORnet® CRNs, members of PCORI's Board of Governors, the Assistant Secretary for Planning and Evaluation (ASPE), and CMS between August 2021 and March 2022.
2. A two-hour Roundtable on March 30, 2022, with Principal Investigators from PCORnet® CRNs to discuss findings from the informant interviews and discuss opportunities and initiatives for enhancing CMS claims data access across PCORnet.

Analytic Approach and Research Questions

We developed guiding research questions focused on three domains to inform key informant interview protocol development to guide our analysis. Figure 4 summarizes the domains and key questions.

Figure 4. Domains and Key Questions

Domains	Key Questions
Current Landscape of Medicare and Medicaid Data within PCORnet	Which patient populations are typically represented in the CMS data CRNs request to access? What governance processes do CRNs have in place for researchers to request access to Medicare and Medicaid data for research? What data infrastructures have CRNs implemented within their networks for researchers to use Medicare and Medicaid data for specific research studies?
Challenges to Medicare and Medicaid Data Access for Research Purposes	What challenges and facilitators do CRNs experience accessing Medicare and Medicaid data from CMS? What challenges and facilitators do CRNs experience accessing Medicaid data from state Medicaid agencies?
Opportunities to Enhance Access to Medicare and Medicaid for Multi-Network CER Studies	What resources or processes are currently most helpful when navigating the CMS data request process? What CRN practices or policies around Medicare and Medicaid data access could be scaled to facilitate data access for multi-network research studies? What PCORnet shared infrastructure could facilitate access to cross-network claims data to support multi-network research studies? What activities could PCORI undertake to support access to, and/or enable broader research use of, Medicare and Medicaid data for multi-network research studies?

Key Informant Interviews

We conducted interviews with 13 key informants to understand the current state of Medicare and Medicaid claims data access for research within PCORnet, as well as challenges and opportunities confronting data requestors. Key informants included perspectives both internal and external to PCORnet (see Appendix E, Figure E-1). Figure 5 shows the distribution of interviews conducted with key informants by stakeholder type.

Figure 5. Key Informant Interviews by Stakeholder Type

Stakeholder Type	Number of Interviews
CRN Principal Investigators	8
PCORI Board of Governors	2
Federal Informants	3

Roundtable Discussion

NORC convened a Roundtable discussion on March 30, 2022. The Roundtable focused on two main objectives: 1) discuss opportunities to improve access to Medicare and Medicaid claims data across

PCORnet, and 2) explore strategies for supporting multi-network studies with linked Medicare and Medicaid claims data in alignment with PCORI's Phase 3 strategic goals for PCORnet. In total, 51 attendees were present at the meeting. Meeting attendees represented PCORnet® CRNs (GPC, OneFlorida, ADVANCE, REACHnet, and STAR), the Coordinating Center for PCORnet, CMS, and ASPE.

3. Key Findings

Key findings are organized into three sections: 1) current landscape of Medicare and Medicaid claims data within PCORnet; 2) challenges to enhancing access to Medicare and Medicaid claims data; and 3) opportunities to enhance access to CMS claims data for multi-network CER.

3.1. Current Landscape of Medicare and Medicaid Claims Data within PCORnet

This section presents main themes from key informant interviews, beginning with a description of the types of Medicare and Medicaid claims data available across CRNs, data governance, the approaches used to request and store CMS data by CRN prime and partner institutions, followed by a description of approaches for conducting multi-network studies using claims data.

3.1.1. Types of Medicare and Medicaid Data Available Among CRNs

Based on our key informant interviews, access to and mechanisms for obtaining CMS Medicare and Medicaid data files vary based on the specific research needs of researchers at CRN prime and partner institutions (see Figure 6). Both Medicare and Medicaid data are relevant for most researcher studies; however, PEDSnet and ADVANCE generally limit CMS data acquisition to Medicaid data due to their specific patient populations (i.e., pediatrics and FQHCs, respectively).

Figure 6. Mechanisms for Medicare and Medicaid Claims Data Access within CRNs

	ADVANCE	GPC	INSIGHT	OneFlorida	PaTH	PEDSnet	REACHnet	STAR
Medicare - ResDAC		●	●	●	●		●	●
Medicare - QE	●							
Medicare - ACO			●					
Medicare - United States Renal Data Renal Data System (USRDS)						●		
Medicaid - T-MSIS		●						
Medicaid - State	●		●	●		●	●	●

	ADVANCE	GPC	INSIGHT	OneFlorida	PaTH	PEDSnet	REACHnet	STAR
Medicaid - All-payer claims databases (APCDs)						●		●

Most PCORnet® CRNs currently use ResDAC to obtain Medicare data, although two CRNs also leverage their participation in CMS' QE or ACO programs for research use of Medicare data. The scope of Medicare data available among CRNs varies in population coverage, latency, and timeframe. Depending on the CRN, Medicare claims data may cover all beneficiaries in a CRN's coverage area, which can be tens of millions of patients, or be limited to smaller, specific populations on a study-by-study basis. Key informants representing some PCORnet® CRNs indicate their prime institutions have acquired multiple years of Medicare claims data, generally spanning five to seven years since 2011. Figure 7 below provides a description of the Medicare data available within PCORnet® CRNs.

Figure 7. Description of Available Medicare Data within Each CRN

CRN	Description
ADVANCE	OCHIN, the CRN prime institution, has access to Medicare data as a QE and has Medicare claims data, limited to QE-related research efforts, for certain ADVANCE partner institutions' patients in California and Massachusetts.
GPC	The University of Kansas Medical Center (KUMC) has claims data for all covered Medicare fee-for-service beneficiaries in nine states (Illinois, Iowa, Kansas, Minnesota, Missouri, Nebraska, Texas, Utah, and Wisconsin) where GPC partner institutions are located. The beneficiary data cover approximately 20 million GPC patients linked to the claims data.
INSIGHT	INSIGHT projects that are eligible as quality improvement projects use claims data from the CRN prime institution's (Weill Cornell Medicine) ACO. The ACO data include Medicare claims for 1 million of the 14 million INSIGHT patients.
OneFlorida+	The University of Florida utilizes ResDAC to access Medicare claims data. Partner institutions send finder files to ResDAC and the University of Florida sends ResDAC patient IDs so ResDAC can link claims data to the patient ID.
PaTH	The University of Pittsburgh has five years of Medicare claims data from 2011–2015 for all patients served by PaTH partner institutions in Maryland, Pennsylvania, and Washington, DC.
PEDSnet	PEDSnet's prime institution (Children's Hospital of Philadelphia (CHOP)) and partner institutions generally do not attempt to collect Medicare data because they have a primarily pediatric patient population. However, CHOP is working to acquire Medicare data for children with end-stage renal disease (ESRD) through the USRDS for the PRESERVE (Predicting the Safety and Effectiveness of Inferior Vena Cava Filters) study.
REACHnet	The Louisiana Public Health Institute (LPHI) acquires Medicare data on a study-by-study basis from CMS, with scope varying by study.
STAR	STAR's prime institution (Vanderbilt University Medical Center (VUMC)) and a partner institution (the University of North Carolina at Chapel Hill (UNC-Chapel Hill)) acquired and maintained separate repositories that together represent approximately six years of Medicare claims data on more than 800,000 patients through 2018.

There are three principal approaches PCORnet participants currently use to access Medicaid data:

1. Enter into agreements directly with individual state Medicaid Agencies either to obtain data annually or on a study-by-study basis;
2. Obtain national Medicaid data via CMS; or
3. Use other approaches including APCDs, commercial payers, and past PCORnet® Health Plan Research Networks to acquire Medicaid data.

Of those CRN prime and/or partner institutions that use Medicaid data, all but one uses the first approach and enter into agreements directly with individual state Medicaid Agencies. For example, OneFlorida's prime institution, the University of Florida, has negotiated agreements with the states of Florida and Texas to obtain claims data for all Medicaid and dually eligible patients in those states. GPC is the only CRN whose prime institution, the University of Missouri, uses the second approach, obtaining Medicaid data via CMS through the Medicaid Analytic eXtract (MAX) program and T-MSIS for all GPC partner institutions. Finally, in the third approach, PEDSnet has used a variety of methods to acquire Medicaid data, including facilitating linkages to APCDs and entering into agreements with commercial payers that offer Medicaid products.

As described earlier, not all CRN partner institutions request access to Medicaid claims data—PaTH's prime institution, the University of Pittsburgh, for example, has not received requests from partner institution investigators for Medicaid claims data, but indicated interest in obtaining Medicaid data in the future.

Figure 8 below provides further detail on the Medicaid claims data to which CRN partner institutions have access for research purposes.

Figure 8. Description of Available Medicaid Data Access within CRNs

CRN	Description
ADVANCE	OCHIN, the CRN prime institution, currently establishes agreements on a state-by-state basis to obtain Medicaid claims data but does not have Medicaid data for all states covered by the ADVANCE network. Starting in 2012, OCHIN received Medicaid data annually from the state of Oregon to support certain research studies. Use of Oregon's Medicaid data must be approved by a state institutional review board (IRB) on a protocol-by-protocol basis.
GPC	GPC's prime institution is the only CRN network partner that obtains data via CMS. KUMC acquired three years of Medicaid data (2011, 2012, and 2017) through the MAX program and T-MSIS. The institution obtains T-MSIS data for its consistency, ease of acquisition, and national coverage.
OneFlorida+	The University of Florida possesses claims data for all Medicaid and dually eligible patients in Florida and Texas through agreements with those states' Medicaid agencies. Early on, the University of Florida conducted stakeholder engagement with the state of Florida to determine how access to Medicaid data would be mutually beneficial to the state and the CRN. The University of Florida's existing relationship with the state allowed it to conduct program evaluation and data quality work and align research interests, resulting in the state giving the University of Florida fully identified Medicaid data. Researchers may use the University of Florida's Medicaid data for observational studies after meeting certain institutional requirements; however, state permission is required for interventional studies.

CRN	Description
PEDSnet	CHOP and PEDSnet partner institutions have used a variety of methods to acquire Medicaid data, including facilitating linkages to APCDs, entering into agreements with commercial payers that offer Medicaid products, and entering into agreements with past Health Plan Research Networks in PCORnet that offered Medicaid products.
REACHnet	LPHI acquires Medicaid data from Louisiana, with data requested on a study-by-study basis.
STAR	VUMC, STAR's prime institution, has an agreement with Tennessee's Medicaid program. UNC-Chapel Hill, a partner institution, also has an agreement with the Medicaid programs in North and South Carolina to obtain Medicaid data on a study-by-study basis.

3.1.2. Approaches to Requesting and Storing Medicare and Medicaid Claims Data for Research Use

PCORnet® CRNs employ different approaches to requesting access to CMS data and storage and governance of these data for research use.

Approaches to Requesting Medicare Data. Four approaches to requesting Medicare data reflect the range of different research needs and goals, as well as CMS data use rules across the CRNs. Figure 10 describes each approach and presents a real-world example of each approach.

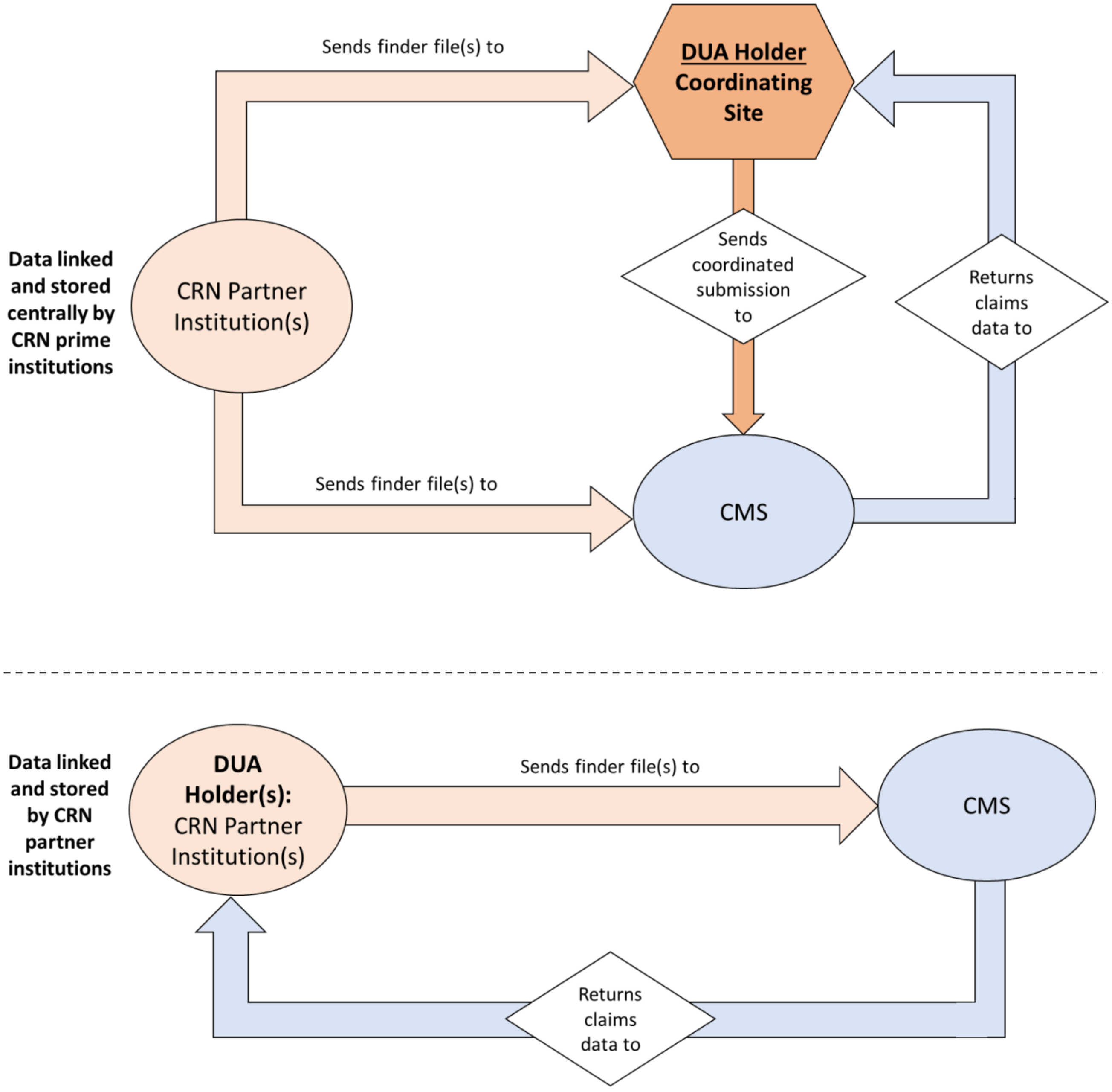
CMS claims data linked and stored centrally by CRN prime institution- A CRN partner institution sends finder files—which identify all the individuals for which the requestor wants Medicare claims data—to either CMS or to their CRN prime institution. The prime institution receives and stores the combined data centrally under a Data Use Agreement (DUA) between the prime institution and CMS. The CRN prime institution makes the combined data accessible to approved researchers.

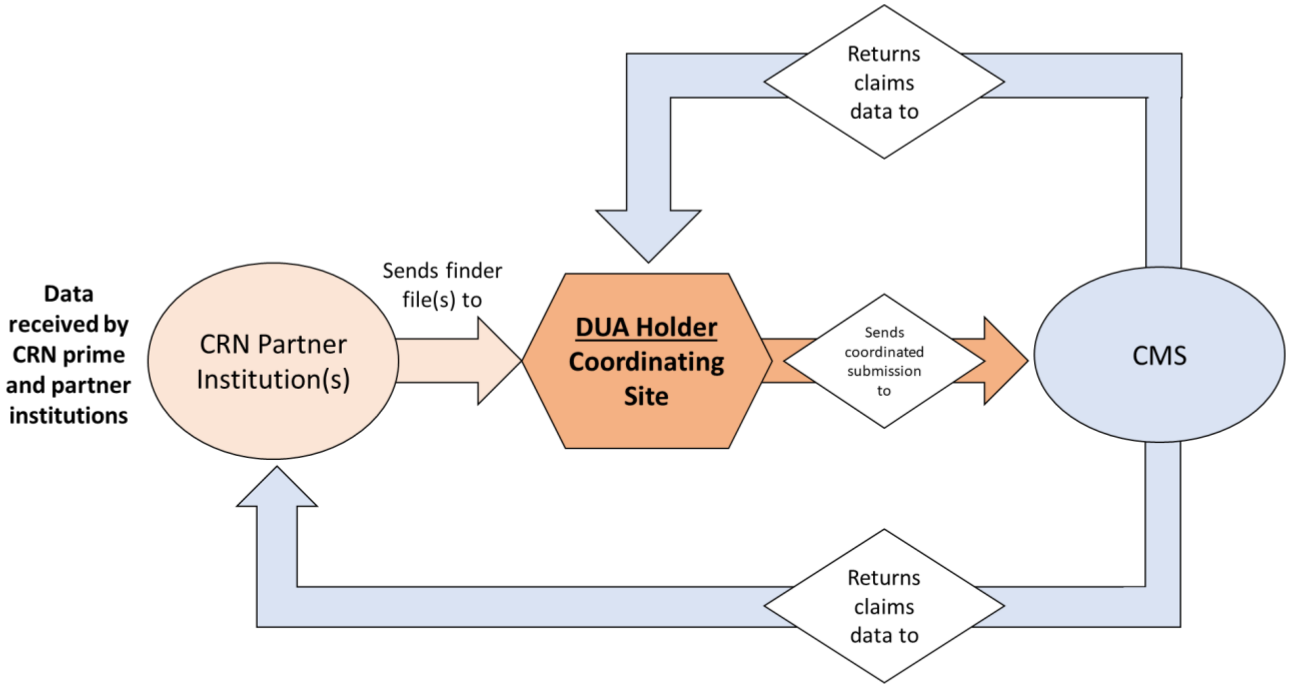
CMS claims data linked and stored locally by CRN partner institutions- CRN partner institutions send finder files directly to CMS, and the partner institutions receive and store data locally under individual DUAs between the partner institutions and CMS. Data are analyzed locally for approved research studies.

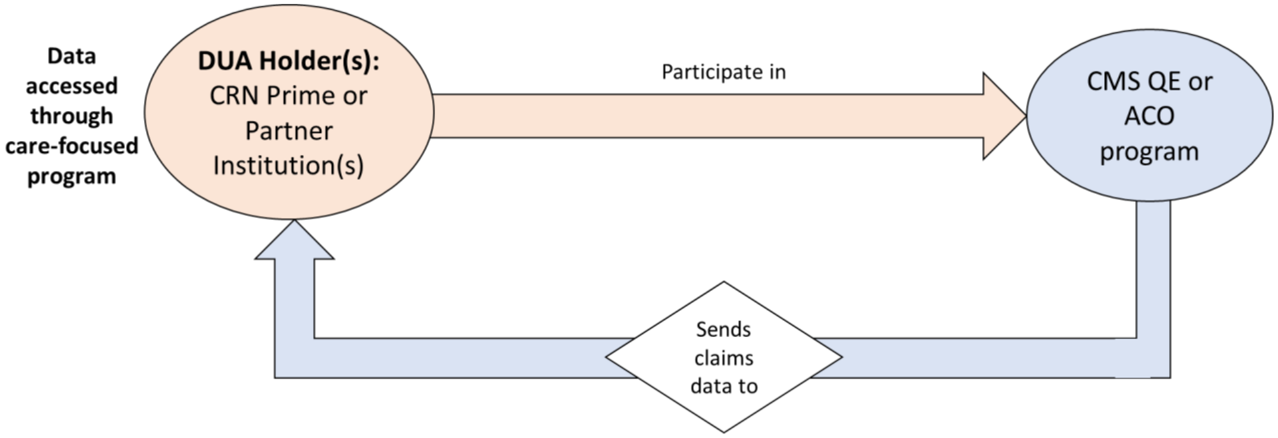
CMS claims data received by both CRN prime and partner institutions- CRN partner institutions send finder files directly to CMS, and both the CRN prime and partner institution receive and analyze data locally under a CMS DUA.

CMS claims data accessed through care-focused program- In a fourth approach, a CRN prime or partner institution participates in a CMS program (e.g., QE, ACO) which provides access to Medicare claims data that can be used for CMS-approved research purposes.

Figure 9. Approaches for Requesting Medicare Data Within a Single CRN







According to information gathered during interviews with key informants, three CRNs use the first approach. For example, under the GPC prime institution's DUA with CMS, GPC partner institutions send finder files directly to CMS, and GPC's prime institution receives and stores claims data centrally via a platform called GROUSE (Greater Plains Collaborative Reusable Observable Unified Study Environment). The GPC prime institution's agreement with CMS does not allow partner institutions to receive a copy of their linked claims data extract (i.e., the data for their specific patient populations). However, researchers, including both those from partner institutions and external institutions, can access the combined claims data via GROUSE for approved research studies.

Under the second approach, STAR's prime institution, Vanderbilt University Medical Center, assists research teams at STAR partner institutions to obtain a DUA with CMS. Each partner institution sends its finder file directly to CMS for the study cohort. The research teams at each institution then receive and store the data locally.

In the third approach, PaTH's prime institution has established a unique agreement with CMS allowing both the prime institution and its partner institutions to receive and store Medicare data, which is accomplished under a single DUA at the prime institution-level. Each CRN partner institution sends its own finder file to CMS. PaTH researchers can reuse the data stores at the prime institution via a virtual desktop if they obtain the necessary approvals.

Finally, two CRNs take the fourth approach. For example, ADVANCE's prime institution, OCHIN, is also an approved CMS QE and can reuse the Medicare data it receives on a quarterly or annual basis through CMS' QE program for CMS-approved research projects.

A table of approaches to requesting CMS data used across CRNs is provided in Appendix F, Figure F-2.

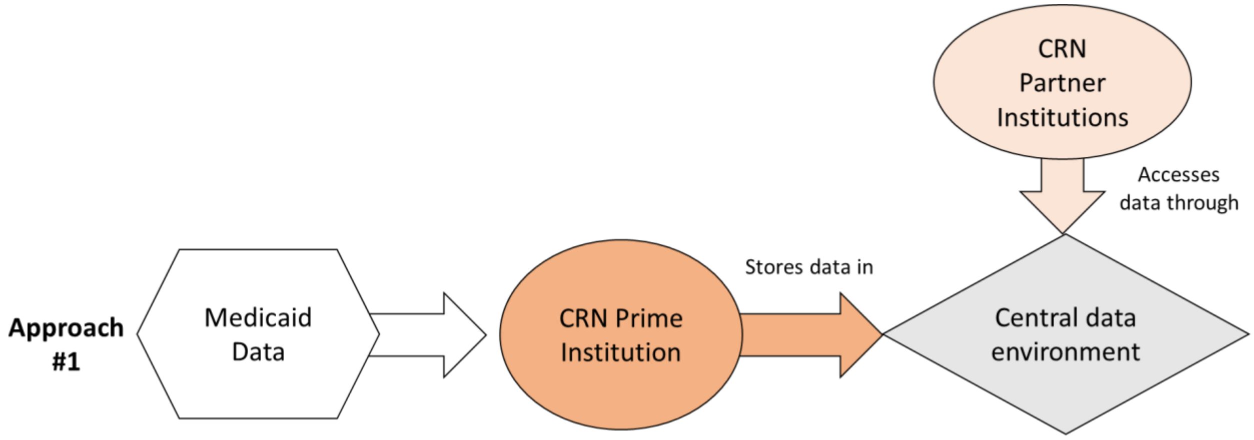
Approaches to Requesting Medicaid Data. PCORnet® CRNs employ three different approaches to requesting Medicaid data. Figure 10 describes each approach and presents a real-world example of each approach.

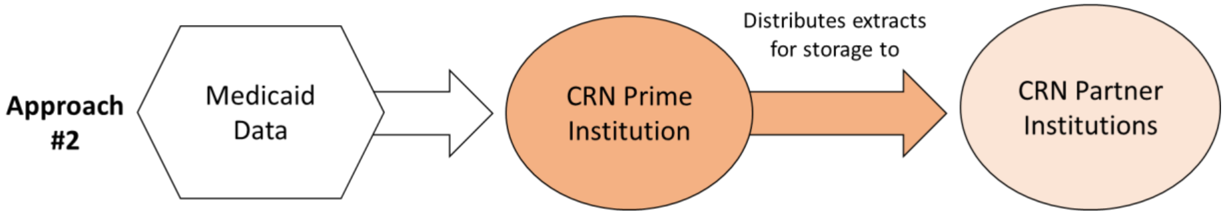
Approach 1. A CRN prime institution conducts linkages to national or state Medicaid data centrally and then makes data available to researchers upon request.

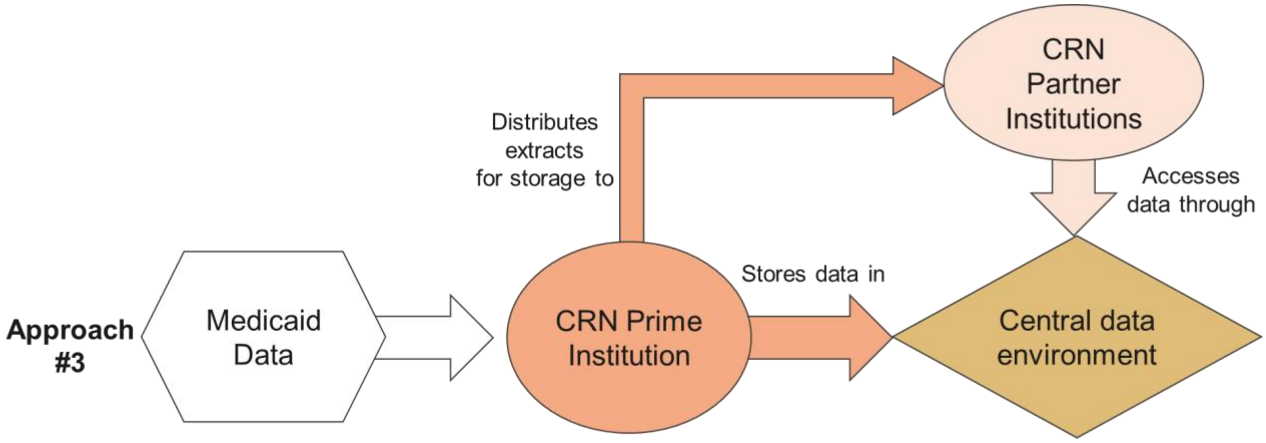
Approach 2. A CRN prime institution conducts linkages to state Medicaid data centrally and distributes partner institution specific extracts to each partner institution for local storage and analysis.

Approach 3. Both options are supported; the CRN prime institution distributes specific extracts to each CRN partner institution for local storage and analysis and maintains aggregate level data for linkage centrally.

Figure 10. Approaches for Requesting Medicaid Data Within a Single CRN







Two CRNs take the first approach. For example, STAR's prime institution facilitates linkages to Medicaid data and distributes files for local storage to research teams on a study-by-study basis. Another three CRNs make Medicaid data linkages available centrally that researchers can request to use. For example, GPC's prime institution makes Medicaid data available through GROUSE. Finally, two CRNs use the third approach. The prime institution for PEDSnet, the Children's Hospital of Philadelphia (CHOP), supports partner institutions in obtaining Medicaid data for specific studies. CHOP also stores the data that has been obtained for studies, and internal researchers with certain permissions can access it to explore feasibility of future research questions. Further examples of CRN data access requests models are provided in Appendix F, Figure F-3.

Data Governance for Research Use. PCORnet® CRNs use two broad approaches to facilitate access to Medicare and Medicaid data: 1) centralized, and 2) federated. In the centralized model, the CRN prime institution stores claims data in either its own data enclave or hosts that data on a third-party platform using a commercial vendor. In the federated model, CRN partner institutions store the claims data in their own environments. Figure 11 shows the approaches used within each PCORnet® CRN.

Figure 11. Approaches for Facilitating Research Data Access to Medicare and/or Medicaid Data Among CRNs

	ADVANCE	GPC	INSIGHT	OneFlorida	PaTH	PEDSnet	REACHnet	STAR
Centralized: Stored centrally by the CRN prime institution	●		●	●		●	●	
Centralized: Stored centrally, but hosted on a third-party platform on behalf of the CNR prime institutions		●			●			
Federated: Stored locally by CRN partner institutions								●

The most common approach (employed by four CRNs) is to store data centrally at the prime institution. For example, OneFlorida's prime institution operates a single, secure data warehouse called the OneFlorida Data Trust. Two CRNs use a third-party platform to host data centrally. For example, the GROUSE program managed by GPC's prime institution is a central, cloud-based data repository built with the Snowflake Data Warehouse on Amazon Web Services' (AWS) cloud computing platform. While centralized approaches can be easier from a CMS governance perspective, they can also be less flexible and useful analytically to partners.

Within PEDSnet, partner institutions store Medicaid data locally, but the PEDSnet prime institution operates and maintains a centralized secure data warehouse where study-specific data are stored. Additional details of the approaches used within each CRN to store Medicare and Medicaid data are provided in Appendix F, Figures F-4 and F-5, respectively.

3.1.3. Approaches to Multi-Network Studies

One of the defining features of PCORnet is its national scale, including over 40 health systems across the CRNs participating in PCORnet. Multi-network studies, which include two or more CRNs, leverage this scale and are a priority for PCORI's funding of PCORnet. PCORnet participants currently use three different approaches to conduct multi-network studies using linked claims data within PCORnet.

The first involves distributed analytics, in which PCORnet participants across multiple CRNs perform distributed analytics in which analyses using claims data are conducted separately at each site. In this approach, each PCORnet partner participating in the study separately requests data for their own sites from CMS. While this is an effective approach to link claims and clinical data, it is inefficient resulting in duplication of effort and time lag. Generally, DUAs prevent CRNs from transferring data which reduces the statistical power of these multi-network studies.

The second approach involves centralized analytics, in which One PCORnet partner serves as the primary coordinating site and performs data linkage and analyses centrally. The coordinating site refers to a CRN prime or partner institution that has coordinated a single data request to CMS with a finder file representing patient data from all health systems participating in the study. CMS sends the PCORnet participants' claims data (with patient identifiers removed) to the coordinating site for analysis limiting the access of partner institutions to their linked data.

The third approach also involves centralized analytics, but with the Coordinating Center for PCORnet as the coordinating site. The Coordinating Center for PCORnet acts as the primary research site; PCORnet participants may access the data through a secure, virtual connection to conduct the analytics. As reported in option two, this limits the access of partner institutions to their linked data. Figure 12 below provides further information on each approach and outlines examples employed by different CRNs.

Figure 12. Detailed CRN Approaches to Multi-Network Studies[13, 14]

Approach 1: The PCORnet partner institutions participating in the study agree on a study population and analytic plan, and each institution conducts analyses separately.	• **Example:** This method was implemented in a retrospective cohort study on bariatric surgery in which organizations from 11 CRNs participated (including CRNs not participating in PCORnet).
Approach 2: One PCORnet partner institution acts as the research coordinating site and uses patient IDs shared by institutions participating in the study to aggregate all data for analysis	• **Example:** Partner organizations within STAR and PaTH participated in a study led by the University of Alabama at Birmingham (UAB) in which all participating organizations sent de-identified clinical data to UAB. UAB then created and sent one finder file to ResDAC. CMS then returned a limited data file to UAB for linkage to the clinical data.
Approach 3: The Coordinating Center for PCORnet serves as the coordinating site and links participating institution's claims and clinical data	• **Example:** The ADAPTABLE study and the Pragmatic Evaluation of Events and Benefits of Lipid-Lowering in Older Adults (PREVENTABLE) study.

3.2. Challenges to Enhancing Access to Medicare and Medicaid Data for Multi-Network CER Studies

This section describes key challenges for expansion of access to CMS data across PCORnet® CRNs through data infrastructure enhancements. Despite many CRNs having experience accessing CMS data, several challenges were reported particularly as they relate to using CMS data for efficient multi-network studies that include a large number of participants and participation by several CRNs. Six challenges were identified during the key informant interviews: 1) cost; 2) timeliness; 3) DUA restrictions; 4) DUA processes; 5) privacy; and 6) Medicaid data heterogeneity. Three challenges were common across Medicare and Medicaid data types and other challenges were related to Medicare or Medicaid data type.

> **“Most studies are hardly using claims data, and our Medicare data are old now. Most researchers are not so interested in looking at data from [seven or more years ago]. It would be great to even add a year of data every five years to have longitudinal data, but the cost has just been a barrier.”**
>
> -CRN Key Informant

3.2.1. Common Challenges

This section includes the three challenges that were common across PCORnet® CRNs regarding access to Medicare and Medicaid claims data.

Challenge 1: Cost.

All CRN Principal Investigators reported the cost of Medicare and Medicaid data from CMS as a significant barrier to purchasing claims or keeping claims acquisitions updated**.** The total cost to obtain a research identifiable file (RIF) depends on the files requested, cohort size, frequency of data needed, whether a finder file is needed, and whether preliminary or final data are requested.[15]

In some cases, data cost has ultimately affected CRNs’ approaches to making claims data available to researchers. For example, cost was a key factor in STAR’s decision to move from requesting access to Medicare and Medicaid data for its entire patient population to requesting these data on a study-by-study basis. Previously, STAR would purchase or refresh both types of claims data annually, but the cost to maintain this process and navigate reuse could not be justified given low researcher demand. INSIGHT placed significant effort on limiting Medicare claims data requests to one million patients to remain cost-effective, despite the desire for more data.

GPC specifically reported cost as a barrier to adding more recent years of Medicare fee-for-service and Medicaid claims data to its GROUSE environment. GPC has Medicare fee-for-service claims data up to 2017 and national Medicaid data for 2011, 2012 and 2017. To add more recent claims data to the GROUSE environment, GPC is considering adding claims purchases into the budget of a National Institutes of Health (NIH) grant.

Challenge 2: Timeliness.

Medicare and Medicaid annual claims files lag by a year or more,[16] which reduces demand among researchers who typically need current data, particularly for clinical trials. As of April 2022, the latest year available for final Medicare fee-for-service files is 2021, Medicare Advantage is 2019, and the latest year available for T-MSIS Medicaid final data is 2019, with preliminary 2020 T-MSIS and Medicare Advantage data available.

Latency has also influenced CRNs' approaches to making claims data available for researchers, rendering the development and maintenance of repositories a less attractive option. For example, PEDSnet refrained from purchasing national Medicaid data—instead, helping researchers obtain relevant claims data on a study-specific basis—due to data latency. Another major factor in STAR's shift to a study-by-study approach was lag time of Medicare and Medicaid annual claims files that reduced demand for the data. STAR reported that researchers typically want near real-time data (with a three-month lag at most). Furthermore, the COVID-19 pandemic has accentuated researcher demand for more real-time data.

> **"The data is 18-24 months old, so demand has been low. So, two years ago we stopped linking it every year and we said we're going to wait until someone actually asks for it, then we'll do it on a project basis."**
>
> -CRN Key Informant

As a partial solution to the latency issue, INSIGHT uses Medicare claims obtained through the CRN partner institutions that serve as an ACO, as those data are made available at least quarterly. However, this limits studies conducted with claims data to those that qualify as quality improvement projects.

Informants from one prime institution also reported timeliness as a challenge to using T-MSIS data.

Challenge 3: DUA restrictions.

CRNs that aim to conduct multi-network studies or studies with more than one partner institution face additional challenges regarding the receipt of CMS claims data. CMS DUAs typically allow for only one data recipient for each data request, and once the data are received, the requestor cannot move the data to other locations. This requirement limits a CRN prime institution or coordinating site from distributing claims data to partner institutions or other CRN participants if conducting a multi-network study.

Because of the CMS restrictions on distributing claims data within PCORnet, data that are acquired by individual partner institutions must be analyzed using a distributed model in which each site conducts analytics of its own data. Researchers can then choose to conduct meta-analyses of results obtained from multiple sites. In instances in which a PCORnet partner serves as the coordinating site and obtains CMS data for multiple partners, it must have the technological infrastructure to allow researchers to access the data securely.

3.2.2. Medicare-Specific Challenges

Challenge 4: DUA processes.

PCORnet® CRNs are generally limited to using CMS claims data to conduct research that falls within the scope of the originally approved research use. The CRN prime and/or partner institutions must seek a new DUA to cover new research uses and may request to reuse data originally purchased for a different approved research use. The CMS process for data reuse approval can be difficult to navigate in a timely manner, introducing further delays in data use. For example, STAR reported that it took 6 to 12 months to obtain CMS approval to amend DUA requests seeking to reuse existing data. STAR felt that the time intensive process is arduous and a deterrent for research teams to request data reuse.

Challenge 5: Privacy.

Some PCORnet participants are hesitant to share identifiable data with other institutions (including CMS) without entering DUAs with the other institution(s) or using privacy-preserving record linkage (PPRL) methods that encrypt patient identifiers to acquire CMS data. Most PCORnet® CRN Principal Investigators expressed concerns regarding privacy and security with moving their health system data into the CMS VRDC, as it is a data management infrastructure they do not control.

For example, PaTH reported its partner institutions have concerns with moving their data to the VRDC due to privacy and data security concerns. OneFlorida's partner institutions will not send protected health information (PHI) to the CRN prime institution for the purposes of creating one combined finder file for CMS, although they may be willing to do so if PPRL could be used to share encrypted identifiers with the CRN prime institution and CMS.

3.2.3. Medicaid-Specific Challenges

Challenge 6: Medicaid data heterogeneity.

The process for accessing Medicaid data directly from state Medicaid agencies differs by state. Most of PCORnet® CRNs operate in multiple states, which requires them to navigate individual agreements with each state—a challenge especially for CRNs that do not have established relationships with their state Medicaid agencies.

PEDSnet, with a presence in eight states, has turned to other avenues to obtain Medicaid data on a study-by-study basis, such as APCDs and agreements with local commercial insurers that offer Medicaid products.

ADVANCE noted the challenge of obtaining Medicaid data across multiple states; overall, ADVANCE includes partner institutions across 31 states, with over half of the patients served covered by Medicaid. In 2012, OCHIN, the prime institution of ADVANCE, started requesting Medicaid data from the state of Oregon to support certain research studies, each of which requires approval by a state IRB. ADVANCE has sought Medicaid data from other states with partner institutions, but each state has its unique requirements, which creates further difficulties in obtaining Medicaid data.

To alleviate the burden of applying for Medicaid data across two states (California and Massachusetts), OCHIN joined CMS' QE program with the expectation that this program would allow for access to Medicaid data. The QE program is currently only providing CMS Medicare data to QEs; thus, while OCHIN was able to obtain Medicare claims data for its population as a CMS QE, it was not able to obtain Medicaid claims data under this CMS program.

There is variability in the data quality, availability, timeliness, and standardization of Medicaid data across states, which creates additional challenges when trying to aggregate Medicaid data from multiple states. One CRN key informant noted that Medicaid claims data heterogeneity across multiple state Medicaid agencies has presented data harmonization challenges and created barriers for using Medicaid data to fulfill PCORnet research goals.

CRNs have also noted that due to the varying latency of state-sourced Medicaid data (with some states having particularly outdated data), they have opted for a study-by-study approach for obtaining the data rather than requesting data from the states annually.

3.2.4. Implications of Challenges for Multi-Network Studies

PCORnet® CRN Principal Investigators reported challenges with CMS DUA requirements and constraints that limit the distribution of CMS data between CRN prime and partner institutions and between institutions across CRNs in the context of multi-network studies. Challenges with costs and timeliness of data, as well as barriers to acquiring Medicaid data from multiple states, present challenges for conducting multi-network studies as well, especially for CRNs with patient populations spanning multiple states. Addressing such challenges could facilitate a more efficient process for conducting multi-network studies with Medicare and Medicaid claims data.

3.3. Opportunities to Enhance Access to Medicare and Medicaid Data for Multi-Network CER Studies

At the Roundtable with PCORnet® CRNs on March 30, 2022, CMS described opportunities to improve access to Medicare and Medicaid data. These opportunities include: 1) improvements to research data access, and 2) new programs for Medicare and Medicaid data access.

3.3.1. Leverage Improvements in CMS Data Access

CMS is engaged in several efforts to improve access to Medicare and Medicaid claims data for research use.

Opportunity 1: Research Data Request Process Improvements. CMS is revamping existing DUA forms and processes to streamline the amount of information entities are required to provide and reduce burden related to documenting data security plans. Under the new Data Management Plan Self-Attestation Questionnaire (DMP SAQ) process, organizations with one information technology environment will be allowed to complete one DMP for multiple DUAs; as noted previously, a separate DMP is currently required for each DUA. The DMP SAQ will be more straightforward to complete and will include the use of checklists. Overall, this process will substantially reduce the burden on CRN partner institutions.

This new DMP SAQ process reduces the administrative burden for individual CRN partner institutions when applying to acquire claims data or reuse existing claims for new research purposes. For multi-network studies, CRN partner institutions with approved DMPs may be well positioned to serve as the coordinating site for the CER study and manage the data and data access within the approved environment.

New CMS Data Access Processes Reduce Administrative Burden for Data Requestors

Reduces burden associated with claims data acquisition

Opportunity 2: T-MSIS Medicaid Data. CMS has developed a Medicaid data collection system, T-MSIS, to improve access to high-quality, timely Medicaid and Children's Health Insurance Program (CHIP) data to enable research that improves quality of care and the health of vulnerable populations. While the transition initially resulted in significant data lags, the timeliness and quality of T-MSIS data continues to improve. CMS has recently announced the availability of preliminary 2020 T-MSIS research data files for purchase. Release of subsequent annual preliminary files are anticipated within one year from the calendar year in which the claims were submitted to CMS.

T-MSIS Addresses Several Challenges to Accessing Medicaid Data

Reduces costs of Medicaid data acquisition. Costs for final files are approximately 50 percent less than the fees paid for the preliminary file, and there are no additional fees if data are accessed through the VRDC

Data latency is mitigated through access to preliminary data files for research

The need to navigate individual state DUA processes are eliminated

Reduces ambiguity in assessing data quality and heterogeneity across individual states through the T-MSIS Data Dictionary and the Medicaid Data Quality Atlas

T-MSIS provides CRN prime and partner institutions with a single mechanism for accessing Medicaid data versus negotiating individual agreements with states. T-MSIS also addresses many of the data quality and timeliness concerns expressed by CRN Principal Investigators during the informant interviews. Additionally, the recent decision to release preliminary annual Medicaid claim files increases the timeliness of national Medicaid research data. CMS is currently developing user guidance on the completeness of annual claims data included in preliminary data files. For CRN prime and partner institutions that do not have established DUAs with state Medicaid agencies, acquiring T-MSIS data provides a single path to access standardized Medicaid data for research use. To further support the use of these data for multi-network studies, the Coordinating Center for PCORnet could explore mapping T-MSIS data to the PCORnet® CDM.

Opportunity 3: Explore New Approaches to CMS Data Access Between CRNs and Partner Institutions. CMS now offers two new options to acquire claims data that PCORnet participants may take advantage of which would allow both the CRN prime institution and partner institutions to receive linked claims data. Specifically, these two options allow for more than one entity to be the data recipient for the relevant files. How the options may work in the context of PCORnet® CRNs is described below:

<u>Option 1: Data Access Requests are Coordinated by a Coordinating Site</u>. Within a single CRN, Option 1 enables a coordinating site and additional institutions participating in a research study to coordinate submission of a single finder file, including patient identifiers, to CMS. CMS conducts the linkage and returns the full extract to the coordinating site and partner-specific extracts to each partner institution. The coordinating site pays the usual cost of the CMS data; the partner institution pays a flat fee of $5,000 to obtain a copy of their data extract.

Novel CMS Claims Data Access Address Several Data Acquisition Challenges

Offers an option for more than one institution to receive linked claims data

Can reduce data acquisition costs

<u>Option 2: Data Access Requests are Submitted by Individual Institutions Participating in PCORnet</u>. Within a single CRN, a coordinating site and additional institutions participating in a research study can opt to each send their own finder file directly to CMS, rather than having the coordinating site submit the finder files. CMS compiles the individual finder files, conducts the linkage, then returns the full extract to the coordinating site, and partner-specific extracts to each partner institution. The coordinating site pays the usual cost for the CMS data. If partner institutions coordinate to send their finder files within a few weeks of one another, CMS will maintain the $5,000 fee for each partner institution extract. If partner institution finder files are received after the group submission period, the partner institution will be charged the full data extraction fee. Consequently, coordination amongst CRN partner institutions is important to reducing data acquisition costs.

These two options may provide added value to partner institutions by providing them with their own data.

Leveraging the VRDC for Multi-Network Studies Reduces Data Acquisition Costs

Reduces the cost of obtaining physical data from CMS

Can reduce institutional data acquisition costs by leveraging transferrable VRDC seat licenses

Opportunity 4: Leverage the CMS VRDC for Multi-network Studies. The VRDC offers an alternative option to acquiring and analyzing claims data. Utilizing the CMS VRDC environment would allow multi-network studies to conduct consolidated centralized analyses, which can improve the statistical power of these types of studies. To use the VRDC, CRN prime and/or partner institutions would enter a joint DUA and obtain the desired number of analyst seats in the VRDC. Analysts would securely upload patient-level clinical data. Then CMS would link to the requested claims data and make the full extract of data available in the secure VRDC user space. Multiple partner institutions would be allowed to access data within the same user space and upload code and other analytic data as needed. Additionally, the seat license is transferable, thus reducing the cost of access across PCORnet by reducing the total number of licenses needed at any given time. Another significant benefit of using the CMS VRDC is access to more timely data, including quarterly extracts of annual fee-for-service claims data. Finally, the overall cost to purchasing a VRDC seat can be significantly lower than the cost of purchasing physical data, depending on the amount of data requested (see Figure C-1 for additional information on CMS data fee structures).

While the VRDC presents a cost-effective, more timely data option for accessing claims data for multi-network studies, its use poses other challenges. During the interview with key informants, CRN Principal Investigators shared that health systems participating in PCORnet are hesitant to share patient data that are hosted in an environment outside of the partner institutions' control. For many partner institutions, the PCORnet distributed infrastructure assuaged these concerns, while facilitating an approach to distributed analytics. Additionally, CRN prime institutions that offer a centralized platform for analytics may find the VRDC duplicative of their own environments, introducing a disincentive to use the VRDC. Despite these challenges, use of the VRDC was identified as a potential opportunity for conducting analytics of comparison populations for both multi-network studies and studies conducted within a single CRN. In addition, CRNs may explore partnership opportunities with APCDs as well; however, this alternative approach is not covered in depth in this paper.

3.3.2. Explore Use of New CMS Programs for Medicare and Medicaid Data Access

CMS has developed two FHIR-based APIs to enable provisioning of Medicare fee-for-service claims data, which could be leveraged by PCORnet® CRNs in the longer term. The CMS Data at the Point of Care and Blue Button 2.0 APIs allow for data to become part of a prime or partner institutions' designated record set, enabling the partner institution to make determinations regarding approved research uses. Additionally, these new APIs have the capability to provide Medicare data on a timelier basis, reducing the claims data lag inherent in the CMS research data request process.

CMS does not charge users of these services for data, although these pathways do require upfront technologic investments to enable access via these API options. However, these APIs use the HL7 FHIR standard[17] for Medicare claims data rather than the traditional RIF file format, which CRN partner institutions would need to map to the PCORnet® CDM.

Data at the Point of Care API Pilot Augments Clinical Data through Access to Medicare Data for Research

- Costs are significantly minimized as CMS does not charge for Medicare data access through the API
- Use of the API removes lags in data acquisition

Opportunity 5: Data at the Point of Care API Pilot. The Data at the Point of Care Pilot API program offers a novel approach to improve access to Medicare data for both quality improvement and PCORnet research goals. The pilot program aims to provide a complete picture of a patient's medical history and help fill data gaps by providing providers and vendors access to Medicare fee-for-services and Part D claims data using bulk FHIR data calls. Because these data become part of the prime or partner institutions' designated record set, no CMS research data DUA is required and there are no fees paid to CMS to acquire this data. This pilot program allows providers participating in the Data at the Point of Care program to receive formatted Medicare fee-for-service and Part D claims data for their patient population to improve the quality of Medicare beneficiary care.[18] Data sharing is permitted under the HIPAA treatment provision.

As HIPAA-covered entities, providers participating in the CMS Data at the Point of Care pilot program can incorporate Medicare beneficiary data into their designated record set, which enables them to use the data for other purposes as permitted under the law (e.g., for research with the individual's authorization or an IRB-approved waiver of authorization). However, because these data sharing programs are focused on providers and efforts to improve quality of care rather than support for research networks, approaches to partnerships and agreements would need to be refocused to work with the care delivery component of CRN institutions. The pilot is not accepting additional participants at this time, however interested parties can test the program in the pilot's sandbox at https://dpc.cms.gov/.

While Data at the Point of Care is currently in a pilot phase, CRN prime institutions are encouraged to explore the opportunity by working with their partner institutions to review the program documentation[19], sample files[20], discuss the benefits of participation and potential research use cases, and consider applying to future phases of the pilot program.

Opportunity 6: Blue Button 2.0 API. The Blue Button 2.0 program enables beneficiaries to access their own Medicare fee-for-service data and connect it to applications, services, and research programs they trust. Beneficiaries can grant permission to access their data for applications, services, and research programs. Beneficiaries may authorize access their four most current years of data and can revoke permission at any time—potentially enabling researchers using PCORnet to conduct patient-enrolled research studies to gain access to a complementary data source: patient-directed Medicare claims data. While there are no data acquisition fees paid to CMS for data obtained via the Blue Button program, there would be costs associated with the development of technology to deliver and aggregate Blue Button data and the utility of this approach is limited to studies that can elicit consent from participants.

Blue Button 2.0 API Offers Another Approach for Access to Medicare Data

- Access to patient-directed Medicare claims data is available at no cost
- Use of APIs reduces time to acquire data compared to the CMS data access request process

The use of Blue Button 2.0 to support study participation is increasing. There are currently 18 Medicare-authorized Blue Button applications currently supporting research functions—from managing study participants' consent to data donation.[21] Identifying research use cases where Blue Button 2.0 could facilitate increased access to Medicare data should be further explored. This could include further exploration and guidance around harmonizing the data available for each consented patient, given each patient may have different years of data available for research based on the date of consent.

Opportunity #7: Privacy-preserving Record Linkage. CMS is considering implementing PPRL as part of a larger initiative within the U.S. Department of Health and Human Services. The Agency is currently investigating which of the numerous PPRL tools would be the best path forward, though the timeline for selection is unclear.

PPRL Approaches Offer Alternative Approaches to Requesting CMS Data

May provide additional assurances to CRN prime and partner institutions regarding secure data exchange

As most PCORnet participants are familiar with the current CMS research data request process and are currently willing to exchange unencrypted patient identifiers to facilitate linkages, the use of PPRL by CMS may not materially improve access to Medicare or Medicaid claims. CRN partner institutions are encouraged to stay abreast of CMS' decisions on the use of PPRL and to consider any implications of the additional option to use PPRL for data linking with CMS for internal data use polices.

In Figure 13 below, we provide a summary of the opportunities and the specific challenges they address.

Figure 13. CMS Claims Data Opportunities Mapped to Challenges

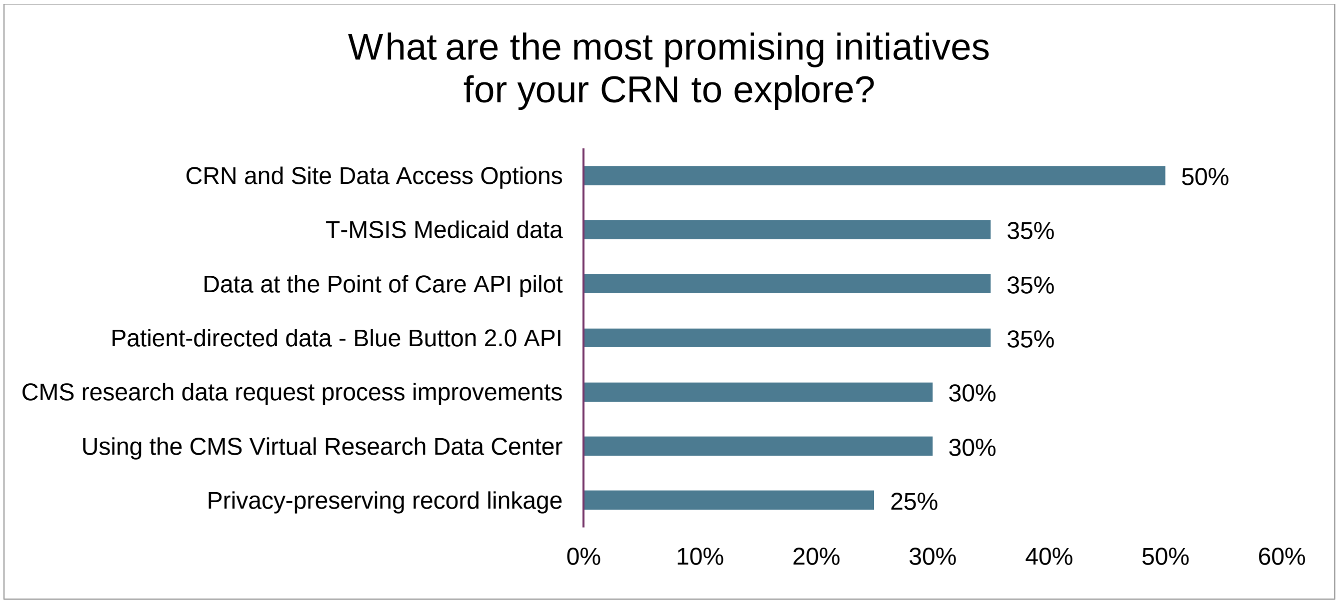
Opportunity	Challenges Identified
CMS redesign of research data request process improvements to streamline the amount of information entities are required to provide, and reduce burden related to data security plans	**DUA processes** (Challenge 4) CMS is revamping the existing DUA forms and process to streamline the amount of information requestors are required to provide and reduce burden related to describing data security plans.
Obtaining T-MSIS data vs. Medicaid-state based data to improve access to high-quality Medicaid and CHIP	**Cost** (Challenge 1) Fees for physical shipment of T-MSIS data are based on year requested and cohort size. Costs for final files are approximately 50 percent less than the fees paid for the preliminary file, and there are no additional fees if accessed through the VRDC. **Medicaid data heterogeneity** (Challenge 6) T-MSIS provides a single mechanism for accessing Medicaid data for multiple states versus negotiating individual agreements with states. CMS provides the data documentation as well as the Medicaid Data Quality Atlas for detailed information about the quality of the data by year and state.
New CRN prime and partner institution data access offer an option for more than one institution to receive linked claims data	**Cost** (Challenge 1) With prior coordination initiated by a coordinating site, CRN partner institutions can pay a flat fee for a copy of their data extract. Coordination amongst partner institutions is important to reducing data acquisition costs. **DUA restrictions** (Challenge 3) New research data access options allow for CRN partner institutions to receive a copy of their data from CMS in addition to the coordinating site.
Leveraging the CMS VRDC to support analysis for multi-network studies	**DUA restrictions** (Challenge 3) Utilizing the CMS VRDC environment would allow multi-network studies to conduct centralized analytics, which can improve the statistical power of these types of studies.
Data at the Point of Care API Pilot **Blue Button 2.0** programs that provide timely, low cost or free Medicare fee-for-service claims data	**Cost** (Challenge 1) Medicare fee-for-service claims data obtained through the Data at the Point of Care API Pilot and Blue Button 2.0 programs do not incur CMS research data costs and do not require CMS DUAs. **Timeliness** (Challenge 2) These APIs allow for Medicare data to be accessed more quickly than the traditional research data access program.

Opportunity	Challenges Identified
CMS implementation of privacy-preserving record linkage to address data privacy concerns	**Data privacy concerns** (Challenge 5) CRNs are encouraged to stay abreast of CMS' decisions on the use of PPRL and consider any changes to internal data use polices among partner institutions.

3.3.3. CRN Prime Institution Perspectives on Opportunities to Improve Access to Medicare and Medicaid Data

At the end of the Roundtable discussion on opportunities, NORC conducted a poll to identify which of the CMS opportunities PCORnet participants saw as most promising (see Figure 14). Among the 20 respondents, all seven of the initiatives in the poll solicited some interest. The highest level of interest was in learning more about new CRN prime and partner institutions' access options for CMS research data (50 percent of respondents). Exploring new CMS data access options for T-MSIS data; Data at the Point of Care API pilot; and Blue Button 2.0 were mentioned by 35 percent of respondents. The remaining initiatives were interesting to at least 25 percent of respondents.

Figure 14. Poll Results of Most Promising CMS Data Access Initiatives*



*Respondents were able to select multiple options.

These responses point to opportunities that can be prioritized in the immediate- and mid-term. Specifically, the new data access options available to CRN prime and partner institutions can be pursued immediately for new claims data research requests. Likewise, accessing T-MSIS Medicaid data represents an immediate-term opportunity for new requests for Medicaid research data. The two CMS API projects each offer novel approaches to supplement clinical data with claims data for CER studies. Interest in these programs should be prioritized as mid-term opportunities, as the Data at the Point of Care project is a pilot and additional considerations were raised regarding the use of Blue Button 2.0 data.

The CRN prime institutions described limited interest in learning more about the CMS research data request process improvements, incorporating PPRL processes in linking to CMS data, and leveraging the CMS VRDC to conduct multi-network studies. The responses indicate these three opportunities do not need to be prioritized in the near-term.

Next Steps

Given the interest expressed in these opportunities by the Network, there are several actionable steps that can be pursued in the immediate- and mid-term to support data infrastructure enhancement to PCORnet.

In the near-term, the Coordinating Center for PCORnet can begin exploration of incorporating the T-MSIS research data into the PCORnet® CDM. Furthermore, the Coordinating Center and the PCORnet® governance structure (i.e., PCORnet® Steering Committee) can begin developing guidance for CRN prime and partner institutions on developing new DUAs with CMS that include data acquisitions for both coordinating sites and partner institutions. The PCORnet® Steering Committee may also wish to form a working group of interested members to further explore Network options for the Data at the Point of Care and/or Blue Button 2.0 pilots, depending on the various CRN partner institutions' research needs. Long-term, accessing claims data via new CMS programs and resources can equip the Network to consider a range of multi-network CER questions for which linked claims and clinical data are needed to support robust and definitive national studies.

Conclusions

This White Paper articulates findings on the current state of CMS claims data access across PCORnet® CRNs, along with challenges and opportunities to expand use of claims data for PCOR, particularly for multi-network studies.

Engaging with CMS to explore ways PCORnet can maximize the utility and value of claims data shows great promise. PCORnet partners express strong interest in exploring the new approaches CMS has developed to increase the availability and accessibility of Medicare and Medicaid data. Doing so has great potential to address many of the challenges identified through the interviews with key informants.

Of the opportunities CMS presented to Roundtable attendees, the CRN prime institutions chose the following four as the most promising:

1. Leverage new distributed data access options for CRN prime and partner institutions
2. Utilize T-MSIS to access Medicaid data
3. Explore the Data at the Point of Care pilot

4. Explore the Blue Button 2.0 API

These opportunities for data infrastructure enhancement have potential to increase Medicare and Medicaid data capture, availability, and use across PCORnet® CRNs to enable learning and national scale research based on the healthcare experiences of millions of Americans with enhanced accuracy and efficiency. The information presented in this White Paper should inform PCORI of the CMS data access opportunities and facilitate discussion with health systems participating in PCORnet and relevant Network governance as they work to expand access to CMS claims data for multi-network studies.

Appendix A. Overview of Clinical Research Networks

The table below summarizes the CRN prime and partner institutions participating in PCORnet at the time of the data collection and analysis, July to December 2021.

Clinical Research Network Prime Institution	Network Composition	Partner Institution
Accelerating Data Value Across a National Community Health Center Network (ADVANCE)	Led by OCHIN, Inc., ADVANCE represents more than 5 million patients at community health centers (CHCs) nationwide. Partners contribute ambulatory patient and community-level data from more than 44.6 million clinic encounters for more than 5 million patients.	OCHIN, Inc. (Lead) Fenway Health Health Choice Network HealthLandscape Oregon Health & Science University Robert Graham Center
Greater Plains Collaborative (GPC)	GPC collected data for more than 20 million patients. Led by the University of Kansas Medical Center Research Institute, the network used its data to focus on three disease cohorts: amyotrophic lateral sclerosis (ALS), breast cancer, and obesity. The network's de-identified database gathered data primarily from electronic health record and billing systems and merged these data with claims from CMS to increase data completeness.	University of Kansas Medical Center (Lead) Allina Health, Intermountain Healthcare Marshfield Clinic Research Institute Medical College of Wisconsin University of Iowa Healthcare University of Missouri University of Nebraska Medical Center University of Texas Health Science Center at San Antonio University of Texas Health Science Center at Houston University of Texas Southwestern Medical Center University of Utah Washington University in St. Louis

Clinical Research Network Prime Institution	Network Composition	Partner Institution
INSIGHT	INSIGHT brought together seven partner institutions in and around New York City to collect data on 12 million unique patients. Led by Weill Medical College of Cornell University, INSIGHT has six additional partners in the New York area and collects data from EHRs, and Medicare, Medicaid, and commercial insurance claims.	Weill Cornell Medicine (Lead) Albert Einstein College of Medicine Columbia University Irving Medical Center Icahn School of Medicine at Mount Sinai New York-Presbyterian Hospital New York University School of Medicine
OneFlorida	Led by the University of Florida, the network includes 11 additional partners populating a database containing information on more than 5 million Florida residents. Data are available since 2012. Data come from patients who consented to release their health information as well as EHRs from partner institutions, Medicaid and Medicare claims, the Florida Cancer registry, and state vital statistics.	University of Florida – UFHealth (Lead) AdventHealth Bond Community Health Center, Inc. Emory University Florida State University Nicklaus Children's Hospital Orlando Health System Tallahassee Memorial Healthcare Tampa General Hospital University of Alabama at Birmingham University of Miami and UHealth University of South Florida and USF Health
Path Towards a Learning Health System (PaTH)	Led by the University of Pittsburgh, PaTH comprises seven partner institutions in the Mid-Atlantic and midwestern regions. The network had data on more than 13 million patients as of July 30, 2019. Data sources included Medicare and Medicaid claims, EHR data, relevant laboratory results, and patient-reported surveys.	University of Pittsburgh (Lead) Geisinger Health System Johns Hopkins Institute for Clinical & Translational Research Johns Hopkins Medicine Penn State College of Medicine Penn State Health Medical Center Temple University School of Medicine The Ohio State University The Ohio State University Medical Center University of Michigan Institute for Clinical & Health Research UPMC

Clinical Research Network Prime Institution	Network Composition	Partner Institution
PEDSnet	PEDSnet is a national pediatric learning health system. Led by Children's Hospital of Philadelphia, PEDSnet is made up of eight pediatric medical centers, whose primary markets are located in eight states, with data on more than six million children, adolescents, and young adults. PEDSnet's data encompasses all pediatric diseases and specialties. The main data sources are EHRs. Additional data are collected by individual studies from health insurance and medication claims, birth records, and child and parent surveys.	Children's Hospital of Philadelphia (Lead) Ann & Robert H. Lurie Children's Hospital of Chicago Children's Hospital Colorado Cincinnati Children's Hospital Medical Center Nationwide Children's Hospital Nemours Children's Health System Seattle Children's Research Institute Stanford Children's Health
Research Action for Health (REACHnet)	Led by the LPHI, REACHnet is a CRN that represented almost 6.4 million patients at four healthcare systems in Louisiana and Texas.	Louisiana Public Health Institute (Lead) Baylor Scott & White Health Research Institute Ochsner Health System Tulane Medical Center University Medical Center New Orleans
Stakeholder, Technology and Research (STAR)	Led by VUMC, the STAR Network has collected EHR data for more than 20 million patients as of August 2018. The network collected data from academic health centers; community hospitals; and a variety of outpatient practices including primary care, specialty practices, and safety net facilities serving diverse populations.	Vanderbilt University Medical Center (Lead) Duke University Health Sciences South Carolina Mayo Clinic Meharry Medical College The University of North Carolina at Chapel Hill Vanderbilt Health Affiliated Network Wake Forest Baptist Health

*CAPriCORN is no longer participating in PCORnet

Appendix B. Clinical Research Network Interview Summaries

Accelerating Data Value Across a National Community Health Center Network (ADVANCE)

Principal Investigators	Jon Puro John Heintzman
Lead Partner	OCHIN, Inc.
Other Partners	Fenway Health Health Choice Network HealthLandscape Oregon Health & Science University Robert Graham Center
Area Covered	26 states
Area(s) of focus	Safety net care

Led by OCHIN, Inc., ADVANCE represents more than 5 million patients at CHCs nationwide. Partners contribute ambulatory patient and community-level data from more than 44.6 million clinic encounters for more than five million patients. ADVANCE's patient population are largely underserved patients that seek care at safety net clinics. Roughly half of the patients are Medicaid enrollees and about a quarter are uninsured. Of the remaining 25 percent, 8 or 9 percent have Medicare coverage, and the remaining have commercial coverage.

Medicaid Data

ADVANCE obtains Medicaid data on a state-by-state basis. The prime CRN institution primarily receives Medicaid data from Oregon. The data are stored centrally at OCHIN, separated from the research data warehouse. They are sequestered for specific protocols. Medicaid data coming from the state of Oregon must be approved by a state IRB on a protocol-by-protocol basis. For other states where ADVANCE has a large presence, it has been challenging to obtain a regular feed of state Medicaid data.

Medicare Data

OCHIN became a QE and has started getting Medicare data for certain OCHIN members. Only 8 to 9 percent of ADVANCE patients are covered by Medicare, so obtaining Medicare claims data has not been a priority.

The Greater Plains Collaborative

Principal Investigators	Russ Waitman Rick Barohn
Lead Partner	University of Kansas Medical Center
Other Partners	Allina Health Intermountain Healthcare Marshfield Clinic Research Institute Medical College of Wisconsin University of Iowa Healthcare University of Missouri University of Nebraska Medical Center University of Texas Health Science Center at San Antonio University of Texas Health Science Center at Houston University of Texas Southwestern Medical Center University of Utah Washington University in St. Louis
Area Covered	Iowa, Kansas, Minnesota, Missouri, Nebraska, Texas, Utah, Wisconsin
Area(s) of focus	ALS, breast cancer, and obesity

The Greater Plains Collaborative (GPC) collects data for more than 20 million patients. Led by the UKMC Research Institute, the network uses its data to focus on three disease cohorts: ALS, breast cancer, and obesity. GPC operates a central data repository called GROUSE, which is built with Amazon Web Services (AWS) and the Snowflake data platform. The environment consists of 1) a data lake, where data are loaded into secure S3 buckets via Secure Shell File Transfer Protocol or Transport Layer Security 1.2 Protocol; 2) a data warehouse, where data are extracted and loaded into Snowflake for data transformation into the PCORnet® CDM and de-identified; and 3) and analytic workbenches, where self-service workspaces are deployed that give researchers the ability to query data. GPC partner institutions send finder files to CMS for record linkage. The GPC Coordinating Center links hashes CMS data with limited data sets received from each partner institution. Linked data sets are stored centrally in GROUSE.

Medicaid Data

GPC has two years of Medicaid data (2011 and 2012) for all patients covered by Medicaid in states served by GPC. Medicaid data is provided by CMS through the T-MSIS program.

Medicare Data

GPC has Medicare data up to 2017 for all patients covered by Medicare in states served by GPC. These data are provided by CMS.

INSIGHT

INSIGHT brings together seven partner institutions in and around New York City to collect data on 12 million unique patients. Led by Weill Medical College of Cornell University, INSIGHT has six additional partners in the New York area and collects data from partner institutions EHRs and Medicare, Medicaid, and commercial insurance claims.

Principal Investigators	Rainu Kaushal Mark Weiner
Lead Partner	Weill Cornell Medicine
Other Partners	Albert Einstein College of Medicine Columbia University Irving Medical Center Icahn School of Medicine at Mount Sinai New York-Presbyterian Hospital New York University School of Medicine
Area Covered	New York City, New York

Medicaid Data

INSIGHT acquires Medicaid data for all patients from the state of New York. These data are stored separately from patient population data at the prime institutions. INSIGHT has a broad DUA with the New York state Medicaid agency to conduct research regarding high cost and high need patients.

Medicare Data

INSIGHT primarily obtains Medicare data quarterly through the prime institutions' ACO for studies that qualify as quality improvement projects; the ACO data includes Medicare data for 1 million of the 14 million patients they serve. INSIGHT has an approved CMS DUA to conduct research on high cost and high need beneficiaries.

OneFlorida +

Principal Investigators	Betsy Shenkman Bill Hogan
Lead Partner	University of Florida – UF Health
Other Partners	AdventHealth Bond Community Health Center, Inc. Emory University Florida State University Nicklaus Children’s Hospital Orlando Health System Tallahassee Memorial Healthcare Tampa General Hospital University of Alabama at Birmingham University of Miami and UHealth University of South Florida and USF Health
Area Covered	Florida

Led by the University of Florida, the network includes 11 partners populating a database containing information on more than 5 million Florida residents. Data are available since 2012. Data include clinical data from partner institutions, state Medicaid and CMS Medicare claims, the Florida Cancer registry, and state vital statistics.

Medicaid Data

OneFlorida has fully identified Medicaid claims data and fully identifiable information on dual-eligibles for all patients in Florida. They also have a fully identified APCD of linked mother-baby data. Early on through stakeholder engagement, the OneFlorida prime CRN institution determined how their access to Medicaid data would be mutually beneficial to the state. They collaborate with the state on program evaluation work, data quality work, and aligned research interests. OneFlorida is approved to use the state-provided Medicaid data for observational studies; but must seek state approval for each proposed interventional study. Overall, they have 80-85 percent of interventional study requests approved.

Medicare Data

OneFlorida operates a single, secure data warehouse managed by the OneFlorida Coordinating Center (OFCC) at the University of Florida, called the OneFlorida Data Trust. For Medicare claims data, OneFlorida participating partner institutions voluntarily send a finder file to ResDAC for record linkage. The linked datasets are then stored in the Data Trust.

PaTH

Principal Investigators	Kathleen McTigue Michael Becich
Lead Partner	University of Pittsburgh
Other Partners	Geisinger Health System Johns Hopkins Institute for Clinical & Translational Research Johns Hopkins Medicine Penn State College of Medicine Penn State Health Medical Center Temple University School of Medicine The Ohio State University The Ohio State University Medical Center University of Michigan University of Michigan Institute for Clinical & Health Research UPMC
Area Covered	More than 13 million patients in the New England, Mid-Atlantic, and midwestern regions served by a partner institution
Area(s) of Focus	Healthy lifestyles and weight, idiopathic pulmonary fibrosis, and atrial fibrillation

Led by the University of Pittsburgh, PaTH comprises seven partner institutions in the Mid-Atlantic and midwestern regions. The network has data on more than 13 million patients as of July 30, 2019. Data sources included Medicare and Medicaid claims, EHRs, relevant laboratory results, and patient-reported surveys.

Medicaid Data

PaTH has not used Medicaid data.

Medicare Data

PaTH has Medicare claims data for all patients covered by Medicare from 2011-2015 in Pennsylvania, Washington, D.C., and Maryland. These data are provided by CMS. PaTH stores CRN prime and partner institutions' Medicare claims data via a virtual platform available through the University of Pittsburgh called the Health Services Research Data Center (HSRDC). Researchers access the data via virtual desktops through the HSRDC. Investigators fill out a survey with information about their study, then are invited to a weekly meeting for further discussion with CRN partner institution clinical researchers and informaticians. PaTH then provides a data consult and provides the investigator with a project manager to help the study launch (e.g., secure IRB approval, secure supplemental DUAs). PaTH enters DUAs with investigators to allow access to the HSRDC.

PEDSnet

PEDSnet is a national pediatric learning health system. Led by Children's Hospital of Philadelphia, PEDSnet is made up of eight pediatric medical centers, whose primary markets are located in eight, with data on more than six million children, adolescents, and young adults. PEDSnet's data encompasses all pediatric diseases and specialties. The main data sources are EHRs. Additional data are collected for individual studies and include claims, birth records, and child and parent surveys.

Principal Investigators	Christopher Forrest Charles Bailey
Lead Partner	Children's Hospital of Philadelphia
Other Partners	Ann & Robert H. Lurie Children's Hospital of Chicago Children's Hospital Colorado Cincinnati Children's Hospital Medical Center Nationwide Children's Hospital Nemours Children's Health System Seattle Children's Research Institute Stanford Children's Health
Area Covered	8 states
Area(s) of Focus	Children

Medicaid Data

PEDSnet has set up a variety of pathways to obtain Medicaid data, including establishing agreements with state health departments, and previous efforts to obtain data through APCDs.

Medicare Data

PEDSnet is working to acquire Medicare data for children with ESRD through USRDS for the PRESERVE study.

Research Action for Health Network (REACHnet)

Led by the Louisiana Public Health Institute, REACHnet includes EHR data for more than 5 million patients from partner institutions in Louisiana and Texas. Their mission is to enable multi-site research with enhanced efficiency in real-world health care delivery systems. REACHnet offers linkage with claims data through partnerships with Blue Cross Blue Shield of Louisiana, Humana, and Louisiana Medicaid.

Medicaid Data

REACHnet has Medicaid data for patients in Louisiana; they request data on a study-by-study basis and provided by the state. REACHnet has an infrastructure to link on demand for Medicaid data rather than storing large amounts of Medicaid data centrally.

Principal Investigators	Thomas Carton Beth Nauman
Lead Partner	Louisiana Public Health Institute
Other Partners	Baylor Scott & White Health Research Institute Ochsner Health System Tulane Medical Center University Medical Center New Orleans
Area Covered	Louisiana and Texas
Area(s) of focus	Weight/body mass index, diabetes, and sickle cell disease

Medicare Data

REACHnet acquires Medicare data on a study-by-study basis from CMS. REACHnet works with investigators to create and send finder files to CMS for their patient cohort of interest for record linkage. CMS returns Medicare data to REACHnet, who in turn links the claims and clinical data and sends a file, with patient identifiers removed, to the approved investigators.

Stakeholder, Technology and Research Network (STAR)

Led by Vanderbilt University Medical Center, the STAR Network collects EHR data for more than 20 million patients as of August 2018. The network collects data from academic health centers, community hospitals, and a variety of outpatient practices including primary care, specialty practices, and safety net facilities serving diverse populations.

Medicaid Data

STAR acquires Medicaid data on a study-by-study basis. STAR partner institutions may also have local access and local linkage to Medicaid data. These data are provided by each individual state.

Medicare Data

STAR acquires Medicare data on a study-by-study basis. STAR partner institutions may also have local access and local linkage to Medicare data. STAR works with investigators to create and send finder files to CMS for their patient cohort of interest for record linkage.

Principal Investigators	Russel Rothman Michael Kappelman
Lead Partner	Vanderbilt University Medical Center
Other Partners	Duke University Health Sciences South Carolina Mayo Clinic Meharry Medical College The University of North Carolina at Chapel Hill Vanderbilt Health Affiliated Network Wake Forest Baptist Health
Area Covered	Florida, Minnesota, New Mexico, North Carolina, South Carolina, Tennessee
Area(s) of focus	Chronic conditions, including overweight/obesity, diabetes, cardiovascular disease, sickle cell disease, rheumatic conditions, inflammatory bowel disease, and certain cancers

Appendix C. Current Mechanisms and Fee Structures to Access Claims Data

In this section, we provide background on the current mechanisms for acquiring CMS Medicare and Medicaid claims data.

Medicare Claims Data. Medicare coverage includes four parts, each with their own structures for claims reimbursement. Medicare claims data refer to beneficiary enrollment data which include dates of Medicare enrollment, enrolled services, and patient information such as age, sex, and race.

Medicare Part A (inpatient services) and Part B (outpatient services) claims data are considered "traditional Medicare" under a fee-for-service payment structure. Medicare Part C includes claims from Medicare Advantage plans where patient enrollment and provider reimbursement are managed by private companies; these data are not included in Medicare claims data requested from CMS. Medicare Part D includes prescription drug benefits. Medicare data particularly includes information from prescriptions provided in outpatient settings. Appendix F, Figure F-1 contains further descriptions about each of the data access mechanisms.

Transformed Medicaid Statistical Information System (T-MSIS) for Medicaid Claims Data. T-MSIS is the new Medicaid data submission system for all 50 U.S. states. Researchers can request access to T-MSIS data to obtain Medicaid and Children's Health Insurance Program (CHIP) enrollee data. T-MSIS analytic files (TAFs) are available from 2014 to 2019, with a preliminary file currently available for 2020.

To improve the timeliness of T-MSIS data, CMS has begun offering annual preliminary T-MSIS claims data, which will not include all final adjudicated paid claims for the calendar year. Though preliminary files are produced prior to states completing their annual submissions, these files are considered adequate quality for most uses, according to the Center for Medicaid and CHIP Services (CMCS). CMCS plans to publish completion rates by claim type to inform data requesters. The TAFs include demographic and eligibility information and claims data for inpatient hospital, prescription drug, long-term care, annual provider, annual plan, and other services. As of 2021, all states and two territories are regularly submitting T-MSIS data. The CMS Data Quality Atlas is a tool that assesses data quality by topic and state and is available at https://www.medicaid.gov/dq-atlas/welcome.

Request Process with CMS for Medicare and Medicaid Claims Data. To obtain Medicare and Medicaid claims data from CMS, data requestors (i.e., researchers) must complete a DUA that establishes the terms of the use of data for the study and submit this to the Research Data Assistance Center (ResDAC).[22] ResDAC serves as the CMS research contractor to provide academic, non-profit, and government researchers assistance with requesting Medicare and Medicaid claims data, responding to questions or issues with data, and providing data training.

As outlined in Figure 4 in the body of the paper, Researchers can request preliminary or final research identifiable files (RIFs).[23] RIFs contain PHI at the beneficiary level. Requests for RIFs require a DUA and review by CMS' Privacy Board. The DUA also allows CMS to track the disclosure of Protected

Health Information (PHI) and Personally Identifiable Information (PII) as required by law. In addition to the DUA, data requestors requesting physical copies of data must complete a Data Management Plan Self-Attestation Questionnaire (DMP SAQ) which serves to demonstrate the requesting organization's compliance with CMS data safeguarding requirements, notably in the CMS Acceptable Risk Safeguards (ARS) 3.1 Publication. Other materials may be required in the request packet depending on the organization and purposes for data use. Additionally, data requestors can submit a finder file to ResDAC with their request, or for an additional fee, request to have a finder file created. A finder file identifies all people for which the requestor wants claims data. The finder file contains either Social Security Numbers or Health Insurance Claim Account Numbers for linkage to the claims data.

Once the request packet has been received by ResDAC and has undergone an additional review by a data management contractor, ResDAC submits the packet to the CMS Privacy Board for a final determination before the DUA is signed (see textbox for key review criteria). Upon approval from CMS and fulfillment of payment from the requestor, the data are processed securely through virtual access to the Chronic Conditions Warehouse (CCW) Virtual Research Data Center (VRDC). The CCW VRDC is a virtual research environment that provides researchers access to the requested Medicare and Medicaid claims data (i.e., RIFs). Researchers with access to the CCW VRDC can conduct their analysis within this secure environment.[24] Alternatively, the RIFs can be shipped to the requestor. Once CMS claims data are sent to a site, the DUA restricts the receiving site from sharing or moving the data elsewhere.

Key CMS Privacy Board Review Criteria

Are the files appropriate to meet study design objectives?

Has minimum data necessary policy been met?

Are the data being stored and handled securely?

Does the research include a scientific question and contribute to generalizable knowledge?

CMS permits researchers to amend their study to make minor modifications to their research goals or request additional data—there is no fee for an amendment. CMS permits researchers to request to use data approved for one study for another study without requesting a second copy of the data (reuse)—reuse requires a second DUA and an additional $2,000 fee.

Alternative options for reusing Medicare data obtained through other CMS programs include:

Qualified Entity (QE) program. Through the Qualified Entity (QE) program, participants can obtain Medicare Parts A, B, and D data from CMS at a reduced cost for the primary purpose of evaluating provider performance. Participants can reuse the Medicare data for research. As a QE, a participant must meet the annual requirements of the CMS QE program, including public provider performance reporting.

Accountable Care Organization (ACO) data. Participants of this program can work with ACOs affiliated with their health systems to obtain data, but they must follow each ACO's data access requirements, which can vary by ACO. CMS data obtained through the ACO program include data use restrictions and must be used to support quality assurance.

Fee Structures. In Appendix C, Figure C-1 we describe the fee structures for each type of CMS data. There are varying price structures based on the CMS data type and access method.

General Challenges with Using CMS Claims Data for Research. The process for requesting Medicare or Medicaid data from CMS can take up to 3-5 months to complete,[25] and depending on the type and amount of data requested, the costs of the data can be tens of thousands of dollars.[26] Furthermore, CMS claims data are intended for the purposes of healthcare payments and not research. Consequently, these data may be subject to misclassifications and influenced by financial incentives,[27,28] which can impact estimates of incidence and prevalence and require significant computing resources to analyze.[29] Medicare fee-for-service and Medicare Advantage RIFs release dates are not synchronized. Therefore, researchers interested in obtaining both data files for the same year would need to submit a DUA amendment as new data files are released. Additionally, Medicare fee-for services files are available on a quarterly basis, whereas Medicare Advantage files are only released annually. This White Paper investigates some of these challenges as well as others identified through key informant interviews with Principal Investigators from PCORnet® CRNs.

There are varying price structures based on the CMS data type and access point. These are further described in Figure C-1.

Figure C-1. Overview of Fee Structures for Claims Data from CMS

CMS Data Type	Fee Structures
Physical Copy Pricing	There are initial fees for creation of the finder files if the requestor opts to have ResDAC perform this function for them. The fee for the finder file and crosswalk files is up to $10,500. Medicare files, including enrollment data, are priced based on the types of claims (e.g., inpatient claims, outpatient claims) and number of beneficiaries. Medicaid files include the Medicaid Analytic eXtract (MAX) files and are priced based on the types of claims (e.g., MAX inpatient, MAX prescription drug) and number of beneficiaries.[30]
VRDC Pricing	The user access fees include CMS onboarding, a seat license, operational support, and administrative fees. The user access fees are charged per user per year and vary slightly based on a preference for the "SAS only" option, in which researchers access and manipulate the data using SAS programming code for $20,000, or the "full VRDC" option, which offers other programming options for $22,000. Project fees per DUA per year include ResDAC assistance, privacy board review, DUA review system tracking, data extract processing and access setup, space allocation, and output review. Like the user access fee, the project fee is slightly discounted for the "SAS only" option at $15,000 compared to the "full VRDC" option of $18,000. There are renewal fees per year for the user access fee and project fee.[31]
T-MSIS (Medicaid Claims Data)	Data fees for physical shipment are charged per file, per year, and vary by cohort size Researchers who purchase preliminary T-MSIS files can purchase final files for a reduced fee Researchers do not pay additional fees to access final T-MSIS files in the Virtual Research Data Center.

Appendix D. Prior CMS Data Access Efforts

In November 2018, partners within PCORnet conducted a claims data landscape analysis to understand each CRN's readiness to engage in health services research, including data availability. As part of this analysis, the partners conducted interviews with representatives from each CRN to discuss health services research, data sources, and linkage; and reviewed current data sets and capabilities among the CRNs. At the time of the review, data sets within seven of the CRNs in PCORnet included Medicaid data and six included Medicare data (Figure D-1).

Figure D-1. Types of Data in Data Sets within PCORnet, November 2018

CRN	Medicaid	Medicare	Government Health Plans	Commercial Health Plans
ADVANCE				
CAPriCORN*				
GPC				
INSIGHT				
Mid-South/STAR				
OneFlorida				
PaTH				
PEDSnet				
REACHnet				

All CRN partner institutions have this data source
Some CRN partner institutions have this data source
CAPriCORN was not funded in Phase 3 and as of January 2022 is no longer participating in PCORnet.

Source: Cohen, A., & Thoman, A. (2018, November). Health Services Research CRG Landscape Analysis: Claims

The 2018 review concluded that many CRN prime and partner institutions only pursue claims data in response to a study-specific research request. As a result, researchers at either the CRN prime institution or partner institution are granted study-specific approval for claims data and those data can only be reused for new research purposes if the partner institution receives permission from CMS to reuse the data.

The review found that many CRN partner institutions sought approval to receive physical claims data from CMS, rather than access these data via the VRDC, and store these claims data at the institution that initiated the request per the approved Data Use Agreement (DUA). Once these data are received and stored in the data requestors' environment, they cannot be re-disclosed (i.e., shared or moved), according to CMS data use restrictions.[32] However, the CRN partner institution can request to "reuse" these data for new research studies with appropriate permissions.

Appendix E. Expanded Methods

Figure E-1. Key Informants by Stakeholder Type

Stakeholder Type	Key Informant Name and Institution
CRN Prime Institutions	Lemuel Russell Waitman and Xing Song, GPC Russell Rothman, STAR Kathleen McTigue, PaTH Charles Bailey, PEDSnet Thomas Carton, REACHnet Elizabeth Shenkman, OneFlorida Jon Puro and John Heitzman, ADVANCE Rainu Kaushal, INSIGHT
PCORI Board of Governors	Michael Herndon, Oklahoma Healthcare Authority Connie Hwang, Alliance of Community Health Plans
CMS Data	Scott Smith, ASPE Allison Oelschlaeger and Andrew Shatto, Centers for Medicare & Medicaid Services Jessie Parker, Centers for Medicare & Medicaid Services

Appendix F. Supporting Figures

Figure F-1. Mechanisms for Facilitating Medicare and Medicaid Claims Data Access

Entity	Initiative	Description
Medicare & Medicaid	Research Data Assistance Center (ResDAC)	ResDAC serves as the CMS research contractor to provide academic, non-profit, and government researchers assistance with requesting Medicare and Medicaid claims data, responding to questions or issues with data, and providing data training. Researchers can request preliminary or final RIFs through CY 2020, and the CY 2021 RIF will be available in November 2022.[33]
Medicare	Qualified Entity (QE)	Organizations apply to be CMS QEs at a reduced cost to receive Medicare Parts A, B, and D data for the purpose of evaluating provider performance. Approved organizations must produce and disseminate identifiable or non-identifiable reports on provider performance, and organizations may reuse the data they receive for other research purposes after obtaining a DUA.[34]
Medicare & Medicaid	Accountable Care Organizations (ACO)	ACOs are provider-led organizations which deliver coordinated care through a group of physicians, hospitals, and other providers. They aim to reduce costs while making care management simpler for patients with complex or chronic health conditions. CRN partner institutions can work with ACOs affiliated with their health system to obtain data, but they must follow each ACO's data access requirements and vary by ACO. Data are generally not for research purposes but specifically for quality assurance.
Medicare & Medicaid	U.S. Renal Data System (USRDS)	The USRDS collects, analyses, and disseminates data about chronic kidney disease and ESRD. Funded by the National Institute of Diabetes and digestive and Kidney Diseases, USRDS works with CMS to share datasets and improve the quality of the patient data collected.[35]
Medicaid	Transformed Medicaid Statistical Information System (T-MSIS)	CMS requires states to submit state Medicaid and CHIP enrollment and claims data via the T-MSIS. Because not every state collects their data in the same format, each state must convert their data to the standard T-MSIS format. T-MSIS data is submitted monthly and represents an expanded dataset compared to the prior MSIS data files. Raw data are converted into annual research files, including the T-MSIS Analytic Files, which organizations can request from CMS.[36]
Medicaid	State-level Agreements	Organizations can arrange agreements directly with their state Medicaid office to request and pay for Medicaid claims data without having to go through CMS to obtain the data. Many CRN prime institutions expressed difficulty in obtaining agreements with multiple states when their network spans several states.
Medicare & Medicaid	All-Payer Claims Database (APCD)	APCDs are state-run databased which collect medical, pharmacy, and dental claims in addition to eligibility and provider data from private and public papers as mandated by the state. These datasets are more complete due to containing data from private payers and more times of patient encounters. However, only 18 states have legislation mandating APCDs or are in the process of establishing an APCD.[37]

Figure F-2. Approaches to Requesting Medicare Data within CRNs

CRN	Description
ADVANCE	OCHIN, the prime CRN institution within ADVANCE, can reuse the data it receives through the CMS QE program for research, though participation in the program also comes with several requirements to evaluate provider performance.
GPC	GPC uses the centralized data governance model via a program called GROUSE. If a GPC study's main research question is covered by GPC's current IRB approval, researchers can apply to reuse GPC's claims data without needing to develop a new data management plan, under a new CMS process called the DMP SAQ. To reuse the data, GPC charges fees that depend on which data the researcher requests (e.g., the entire GROUSE data resource or Medicare data only).
OneFlorida+	OneFlorida's partner institutions voluntarily send finder files to ResDAC, which removes patient identifiers and sends the patient ID and data back to the OneFlorida prime CRN institution, which then stores the data for its partner institutions.
PaTH	PaTH has established a unique agreement with CMS allowing both the prime CRN institution and its partner institutions to receive and store Medicare data. Under a single DUA with the prime CRN institution, each CRN partner institutions sends its own finder file to CMS, and CMS agrees both to return the entire geographic claims data set to the prime CRN institution and to return an extract to each participating partner institution specific to their patients. Researchers conducting studies within PaTH can access the data via virtual desktops, and enter DUAs with PaTH.
REACHnet	REACHnet CRN partner institutions create and send finder files affixed to a REACHnet-generated study ID to CMS. CMS then returns the data to the prime CRN institution, which links the clinical data and makes the file available to the relevant research team in a central location.
STAR	STAR assists research teams in completing grant applications and obtaining a DUA with CMS, then sending a finder file for the study cohort. The research teams then receive and store the data locally.

Figure F-3. Approaches to Requesting Medicaid Data within CRNs

CRN	Description
GPC	GPC uses the same approach to governing Medicaid data as it does for Medicare data—the GPC prime institution obtains Medicaid data through CMS. The prime CRN institution establishes the DUA with CMS, CRN partner institutions send the finder files to CMS, and the prime CRN institution stores the data received from CMS in GROUSE.
OneFlorida+	OneFlorida is allowed to use Medicaid data from the state for observational studies; however, for interventional studies it must request state permission.
PEDSnet	PEDSnet facilitates linkages to Medicaid data and hosts the data centrally in addition to distributing study-specific files back to research teams. Internal research team members with certain certifications can access the PEDSnet centralized claims database.
REACHnet	REACHnet has the infrastructure to link Medicaid and clinical data on demand on a study-by-study basis.

CRN	Description
STAR	STAR has established agreements with a couple of states in its coverage area to access Medicaid data on a study-by-study basis. The prime CRN institution facilitates the linkage and distributes study-specific files to the research teams.

Figure F-4. Approaches for Facilitating Research Data Access to Medicare Data Among CRNs

CRN	Description
ADVANCE	ADVANCE stores Medicare data for certain OCHIN sites within its network centrally.
GPC	GPC developed a central, cloud-based data repository called GROUSE built with the Snowflake Data Warehouse on AWS cloud computing platform. Snowflake has a data marketplace that allows linkages to commercial data sources that could significantly improve measurement of data elements such as social determinants of health.
INSIGHT	INSIGHT stores Medicare data centrally for its New York City patient population.
OneFlorida+	OneFlorida operates a single, secure data warehouse, managed by OFCC at the University of Florida, called the OneFlorida Data Trust.
PaTH	PaTH stores data in a University of Pittsburgh on the HSRDC platform, where PaTH partner institution investigators can access the data using virtual desktops. PaTH establishes individual DUAs with these investigators to facilitate appropriate data access. The prime CRN institution pays to use the HSRDC, and investigators also plan use of the platform in their studies' budgets.
REACHnet	REACHnet stores linked datasets (without patient identifiers) for study teams in a central location.
STAR	STAR is the only CRN primarily using local storage at partner institutions.

Figure F-5. Approaches for Facilitating Research Data Access to Medicaid Data Among CRNs

CRN	Description
ADVANCE	OCHIN houses Oregon Medicaid data centrally and separately from the ADVANCE data warehouse. OCHIN can link these data only for approved protocols on request.
GPC	GPC stores Medicaid claims data in its GROUSE data resource.
INSIGHT	INSIGHT stores Medicaid data centrally for its New York City patient population as it does with its Medicare data.
OneFlorida+	OneFlorida stores Medicaid data in the OneFlorida Data Trust and uses its Datavant token for linkage to clinical data.
PEDSnet	PEDSnet facilitates linkages to Medicaid data and hosts the data centrally in addition to distributing study-specific files back to research teams.
REACHnet	REACHnet has an infrastructure to link Medicaid data on demand rather than storing large amounts of Medicaid data centrally.
STAR	STAR uses a related method for Medicaid data infrastructure related to its process for Medicare data infrastructure. Previously, STAR linked and centrally stored Tennessee Medicaid claims to EHR data annually. In recent years, demand caused STAR to switch to a project-by-project approach in which partner institutions locally store their own data.

References

[1] Getting Started with CMS Data. ResDAC. Accessed June 30, 2022. https://resdac.org/getting-started-cms-data

[2] Fee List for RIFs: Physical Research Data Request. ResDAC. Updated March 30, 2022. Accessed June 8, 2022. https://resdac.org/sites/datadocumentation.resdac.org/files/2022-05/CMS%20Fee%20List%20for%20Physical%20Research%20Data%20Request_0.pdf. Updated March 30, 2022. Accessed June 8, 2022

[3] Limited Data Set (LDS) Files. CMS.gov. Updated May 24, 2022. Accessed June 8, 2022. https://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/Data-Disclosures-Data-Agreements/DUA_-_NewLDS

[4] Blue Button 2.0 Implementation Guide. CMS.gov. Accessed July 6, 2022. https://bluebutton.cms.gov/assets/ig/index.html

[5] About PCORI's National Priorities for Health. About PCORI's National Priorities for Health | PCORI. Published June 15, 2021. Accessed July 7, 2022. https://www.pcori.org/about/about-pcori/pcori-strategic-plan/pcori-strategic-plan-national-priorities-health/about-pcoris-national-priorities-health

[6] Patient Protection and Affordable Care Act of 2010, Pub. L. No. 111–148, 124 Stat. 119 (2010), Codified as Amended 42 U.S.C. § 18001.

[7] Engagement. The National Patient-Centered Clinical Research Network. Accessed April 22, 2022. https://pcornet.org/engagement/

[8] PCORnet-DRN-OC / Medicare-Data-Transformation. Medicare to PCORnet Data Transformation Package Version 4.0. Github.com. Published February 8, 2021. Accessed June 6, 2022. https://github.com/PCORnet-DRN-OC/Medicare-Data-Transformation

[9] Courcoulas A, Coley RY, Clark JM, et al. Interventions and Operations 5 Years After Bariatric Surgery in a Cohort From the US National Patient-Centered Clinical Research Network Bariatric Study. JAMA Surg. 2020;155(3):194–204. doi:10.1001/jamasurg.2019.5470

[10] About Us. ResDAC. Accessed June 8, 2022. https://resdac.org/about-resdac

[11] Getting Started with CMS Data. ResDAC. Accessed June 30, 2022. https://resdac.org/getting-started-cms-data

[12] Fee List for RIFs: Physical Research Data Request. ResDAC. Updated March 30, 2022. Accessed June 8, 2022. https://resdac.org/sites/datadocumentation.resdac.org/files/2022-

05/CMS%20Fee%20List%20for%20Physical%20Research%20Data%20Request_0.pdf. Updated March 30, 2022. Accessed June 8, 2022

[13] Courcoulas A, Coley RY, Clark JM, McBride CL, Cirelli E, McTigue K, Arterburn D, Coleman KJ, Wellman R, Anau J, Toh S, Janning CD, Cook AJ, Williams N, Sturtevant JL., Horgan C, Tavakkoli A, for the PCORnet Bariatric Study Collaborative. *JAMA Surg*. 2020;155(3),194-204. doi:10.1001/jamasurg.2019.5470.

[14] ADAPTABLE, a PCORnet® Demonstration Project, Offers a Template for Future Pragmatic Studies. PCORnet. Published April 7, 2021. Accessed June 8, 2022. https://pcornet.org/news/adaptable-a-pcornet-demonstration-project-offers-a-template-for-future-pragmatic-studies/

[15] Fee List for RIFs: Physical Research Data Request. ResDAC. Updated March 30, 2022. Accessed June 8, 2022. https://resdac.org/sites/datadocumentation.resdac.org/files/2022-05/CMS%20Fee%20List%20for%20Physical%20Research%20Data%20Request_0.pdf

[16] RIF Medicare Quarterly Data. ResDAC. Updated September 27, 2016. Accessed June 30, 2022. https://resdac.org/articles/rif-medicare-quarterly-data

[17] Blue Button 2.0 Implementation Guide. CMS.gov. Accessed July 6, 2022. https://bluebutton.cms.gov/assets/ig/index.html

[18] Frequently asked questions. Data at the Point of Care. Accessed May 26, 2022. https://dpc.cms.gov/faq

[19] Documentation. Data and the Point of Care. Accessed May 26, 2022. https://dpc.cms.gov/docsV1

[20] Understanding DPC Data. Data at the Point of Care. Accessed May 26, 2022. https://dpc.cms.gov/data

[21] Medicare's Blue Button apps. Medicare.gov. Accessed May 26, 2022. https://www.medicare.gov/manage-your-health/medicares-blue-button-blue-button-20/blue-button-apps?label=connecttoresearch

[22] About Us. ResDAC. Accessed June 8, 2022. https://resdac.org/about-resdac

[23] Difference between RIF, LDS, and PUF Data Files. U.S. Department of Health & Human Services Guidance Portal. Published September 15, 2020. Accessed June 8, 2020. https://www.hhs.gov/guidance/document/differences-between-rif-lds-and-puf-data-files#:~:text=Research%20Identifiable%20Files%20(RIFs)%20contain,necessary%20are%20requested%20and%20justified

[24] CCW Virtual Research Data Center (VRDC). ResDAC. Accessed June 8, 2022. https://resdac.org/cms-virtual-research-data-center-vrdc

[25] Getting Started with CMS Data. ResDAC. Accessed June 30, 2022. https://resdac.org/getting-started-cms-data

[26] Fee List for RIFs: Physical Research Data Request. ResDAC. Updated March 30, 2022. Accessed June 8, 2022. https://resdac.org/sites/datadocumentation.resdac.org/files/2022-05/CMS%20Fee%20List%20for%20Physical%20Research%20Data%20Request_0.pdf. Updated March 30, 2022. Accessed June 8, 2022

[27] Mues KE, Liede A, Liu J, et al. Use of the Medicare database in epidemiologic and health services research: a valuable source of real-world evidence on the older and disabled populations in the US. *Clin Epidemiol*. 2017;9,267-277. doi:10.2147/CLEP.S105613

[28] Strengths and Limitations of CMS Administrative Data in Research. ResDAC. Updated January 1, 2018. Accessed June 30, 2022. https://resdac.org/articles/strengths-and-limitations-cms-administrative-data-research

[29] Strollo SE, Adjemian J, Adjemian MK, Prevots DR. The burden of pulmonary nontuberculous mycobacterial disease in the United States. *Ann Am Thorac Soc.* 2015;12(10),1458-1464. doi:10.1513/AnnalsATS.201503-173OC.

[30] Fee List for RIFs: Physical Research Data Request. ResDAC. Updated March 30, 2022. Accessed June 8, 2022. https://resdac.org/sites/datadocumentation.resdac.org/files/2022-05/CMS%20Fee%20List%20for%20Physical%20Research%20Data%20Request_0.pdf

[31] Chronic Conditions Warehouse (CCW) Virtual Research Data Center (VRDC) Cloud Migration Fee Changes. ResDAC. Updated December 12, 2021. Accessed June 8, 2022. https://resdac.org/sites/datadocumentation.resdac.org/files/2022-02/CMS%20Fee%20List%20for%20CCW%20VRDC%20Cloud%20Environment.pdf

[32] Data Disclosures and Data Use Agreements. CMS.gov Updated December 3. 2021. Accessed June 30, 2022. https://www.cms.gov/Research-Statistics-Data-and-Systems/Files-for-Order/Data-Disclosures-Data-Agreements/Overview

[33] Research Identifiable File (RIF) Requests. ResDAC. Accessed June 8, 2022. https://resdac.org/research-identifiable-files-rif-requests

[34] Qualified Entity Program. CMS.gov. Updated June 1, 2022. Accessed June 8, 2022. https://www.cms.gov/Research-Statistics-Data-and-Systems/Monitoring-Programs/QEMedicareData

[35] USRDS. National Institute of Diabetes and Digestive Kidney Diseases. Accessed June 8, 2022. https://www.usrds.org/

[36] Medicaid & CHIP Research Files. Medicaid.gov. Accessed June 8, 2022. https://www.medicaid.gov/medicaid/data-systems/macbis/medicaid-chip-research-files/index.html

[37] All-Payer Claims Databases. Agency for Healthcare Research and Quality. Published March 2017. Updated February 2018. Accessed June 8, 2022. https://www.ahrq.gov/data/apcd/index.html