

Leveraging Electronic Health Record Data to Complement Data from Federal Surveys

In the wake of a dynamic environment for federal data infrastructure, state health policy professionals have grown concerned about whether they will continue to have access to data they need for evidence-based policymaking. Uncertainties include the potential cancellation or scaling back of critical surveys, reduced federal survey staffing that could undermine data quality or technical assistance, and the possibility that public data releases may be curtailed or restricted.

To identify strategies to fill potential data gaps caused by federal survey disinvestments, the State Health Access Data Assistance Center (SHADAC), as part of its [State Solutions for Health Data Continuity](#) project funded by the Robert Wood Johnson Foundation, conducted discussions with 46 diverse experts and stakeholders, including state data analysts, survey data collectors, researchers, and foundations. During these discussions, SHADAC was able to gather insights on people's top concerns about emerging and potential gaps in federal surveys, and explore various potential alternatives and strategies for filling state-level data gaps.

We have created a [series of "Strategy Briefs"](#) that review a selection of strategies or "alternatives" for filling federal data gaps, identified both through these discussions and from SHADAC's expertise.

This brief provides a high-level overview of the strategy: **Leveraging Electronic Health Record (EHR) Data to Complement Federal Surveys**. Along with a general description of this gap filling strategy, SHADAC experts will also summarize identified advantages, disadvantages, and implementation considerations for data users and other interested parties.

Complement, not Substitute

A recurring theme across our stakeholder discussions was the **potential for non-survey data to fill emerging gaps**. Important to note: This strategy proposes **using non-survey data as a complement to, rather than a substitute for, federal survey data**. While both individuals we spoke to as part of this project and SHADAC experts agree that non-survey data cannot replace federal survey data's standardized, population-wide view that enables national benchmarking and cross-state comparability, it's also universally agreed that non-survey health data sources could **help fill gaps** and **complement federal health data** that is collected and released.

Sometimes referred to as "real-world" data due to their basis on actual patient encounters with the healthcare delivery system, access to non-survey health datasets has greatly improved over the last two decades due to advances in health information technology, increased consumer demand for health information, legislation to promote and protect electronic exchange of health information, and digitizing health program data.

While many types of health-related non-survey data exist (i.e., hospital discharge data, disease registries, etc.), in our work and through the informative expert conversations, SHADAC has identified two non-survey data types that could be particularly effective in helping to fill emerging health data gaps: **Health Insurance Claims Data** and **Electronic Health Record Data**. This brief focuses specifically on **leveraging Electronic Health Record (EHR) data as a gap filling strategy**; a second brief focuses on the [strategy of using Health Insurance Claims data](#).

Strategy Description

Electronic health record (EHR) data includes information on patient self-reported characteristics, provider diagnoses and treatment, as well as information about related testing and pharmaceutical prescriptions. Over the last 10 years, national, state, and local initiatives have expanded the use of electronic health record (EHR) data to track measures of and trends in public health and to support evidence-based decisionmaking.

While the shift to and use of electronic medical records has improved clinical coordination and patient access to timely information, EHR data's greatest potential for state researchers, analysts, and other state data users as a gap filling strategy depends on its **interoperability**: the ability for data users to securely exchange, standardize, and use electronic health information for broader population health research and goals. EHR data can be an effective alternative to fill federal survey data gaps if it is securely integrated across disparate provider systems via a Health Information Exchange (HIE) or third-party aggregator for use beyond the point of care for monitoring and research.

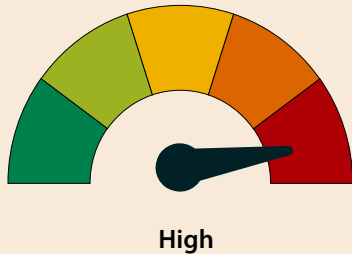
Leveraging Integrated EHR Data: Strategy Advantages

- **Timeliness**
Health information exchange and other integrated EHRs enable access to data in near real-time, avoiding the “run-out” periods and administrative lags that come with using claims or federal health survey data.
- **Population and Clinical Granularity**
All patients, including the uninsured, that present for services are represented in EHR datasets; information consists of not only diagnoses, but also patient vitals, imaging, and laboratory results.
- **Longitudinal Patient Tracking**
Integrated EHR data include unique identifiers, which support longitudinal analysis within integrated systems, something cross-sectional health surveys cannot do.

Leveraging Integrated EHR Data: Strategy Disadvantages

- **Healthcare Seeking Bias**
EHR data only captures individuals who actively engage with the healthcare system, meaning it lacks information on those who do not seek services. It also generally fails to capture information on health behaviors or social determinants that occur outside the clinical setting.
- **Uneven Inclusion**
The denominator population is limited to the service areas of participating, integrated providers. In some states, certain provider types (e.g., federally qualified health centers (FQHCs), rural providers) may not have the infrastructure needed to participate in integrated data sharing. Consequently, data may over- or underrepresent certain geographic areas, provider types, etc.
- **Limited Validation or Standardization**
Data structures vary significantly across different providers and EHR software platforms (e.g., Epic vs. Oracle). This lack of standardization makes EHR data labor-intensive to validate, clean, and harmonize for cross-system analysis.

Resource Intensity



Actors



Health information technology
or health information exchange
vendors



Researchers



Funders

Timeline



Short-term stopgap for
national trends



Possible long-term alternative for
state-level and cross-state
monitoring

Integrated Electronic Health Record (EHR) Data for Understanding Patient Health Characteristics, Care, and Quality: Examples and Implementation Considerations

There are a number of both public and private efforts that are working to implement and facilitate access to and sharing of EHR data across systems for use alongside other health-related datasets such as state and federal survey data. Such efforts demonstrate how securely integrated EHR data can be used for research, monitoring prevalence and treatment of specific health conditions, and tracking public health trends.

The state of Michigan spearheaded an effort called the Chronic Disease Registry Linking Electronic Health Record Data (CHRONICLE), one of the most advanced examples of a state enabling broad access to an EHR-integrated product for health research.¹ CHRONICLE is a near-real-time surveillance system that leverages clinical chronic disease information from electronic health records (EHRs) to support longitudinal monitoring of chronic conditions statewide.

PopHIVE (the Population Health Information and Visualization Exchange), housed at the Yale School of Public Health, is an initiative that brings together near-real-time EHR data with other datasets in order for users to observe county, state, and national trends, also allowing users to gather a more complete picture on a particular topic.² PopHIVE's chronic disease dashboard, for example, includes state-level data on obesity and diabetes "prevalence" from EPIC Cosmos, a de-identified database of EHR data from providers who subscribe to EPIC software, as well from the BRFSS survey and Medicare claims data.³ Because measure definitions vary across sources, PopHIVE provides detailed information on each source as well as notes on how to interpret the data.

Healthcare technology company Komodo Health has developed a product, the Healthcare Map, that de-identifies and integrates longitudinal patient-level EHR data and other encounter data for use by decisionmakers and researchers. In 2024, the Robert Wood Johnson Foundation, in collaboration with Komodo Health and Mathematica, launched a grant program to facilitate access to these near-real-time data for research on diet and health. Eight researchers were awarded the opportunity to use proprietary data to study diet-related conditions and factors that get in the way of nutrition equity, providing them with a timely alternative to federal survey data, such as those available in the National Health and Nutrition Examination Survey.⁴

Additionally, it is important to point out other solutions that link clinical data with survey data to improve data recency and reduce the data collection burden on survey respondents. For example, as part of federal survey modernization efforts, there has been exploration of using [clinical information as a complement to BRFSS for more timely public health surveillance](#).

Also, note how many of these examples utilizing EHR data for research, monitoring, and more, are often used alongside, or as a complement, to other data sources such as federal datasets. Like we mentioned earlier, this gap filling strategy utilizes EHR data as a complement to other types of data and sources—not as a one-to-one alternative to federal data.

While the infrastructure and precedent for using and integrating EHR data is expanding, significant operational and technical hurdles remain for those looking to utilize EHR data as a strategy for filling emerging data gaps, perhaps especially for state health services research on patient health characteristics, care, and quality. While HIEs currently operate in almost every state in the U.S., their utility is often hampered by inconsistent incentives for participation and stringent data access restrictions. Other barriers include data ownership and data governance issues associated with housing and sharing data across organizations and boundaries, as well as data privacy and security concerns. Identifying and scaling models from state HIEs that have successfully navigated these privacy and security barriers is a critical next step for the field.

Beyond governance, financial and technical barriers also deter widespread adoption of this gap-filling strategy. Not only are many EHR data products paywalled, but those paywalls also tend to be expensive (not to mention expenses incurred from security protocols, staff training, and other associated costs). In the case of the use of EHR data to assess diet-related conditions described above, philanthropic dollars made it possible for researchers to access these data free of charge.

Furthermore, the notions of poor data quality, due in part to limited control over data collection and the lack of EHR data standardization across providers and software platforms, deters use. Initiatives like PopHIVE, mentioned above, are addressing these gaps by providing clear documentation about limitations and exploring data validation techniques.

Continued monitoring, experimentation, and technical support for data users are essential parts of implementing data gap filling alternatives and strategies; SHADAC plans to continue research and provide technical assistance in this area—[contact us with questions or feedback](#).

Interested in learning about other alternatives and strategies for filling potential data gaps? Find and read the other iterations of [SHADAC's Strategy Briefs here](#).

¹ Michigan Department of Health & Human Services. (n.d.). CHRONICLE.

<https://www.michigan.gov/mdhhs/keep-mi-healthy/communicablediseases/epidemiology/chronicepi/chronicle>

² PopHIVE. (n.d.). Yale School of Public Health. <https://www.pophive.org/about>

³ PopHIVE (n.d.). *Chronic Diseases: Dashboard*. Yale School of Public Health. <https://www.pophive.org/chronic-diseases>

⁴ Komodo Health. (2024, January 8). *Komodo Health Collaborates With Robert Wood Johnson Foundation To Advance New Research on Diet and Health*. <https://www.komodohealth.com/press/komodo-health-collaborates-with-robert-wood-johnson-foundation-to-advance-new-research-on-diet-and-health>