Improving the Utility and Comparability of Health Care Data for Health Services Research, Policy Decisions, and Transparency Reports

Year 2 Report:

“Improving the Health Care Data Ecosystem Through Data Acquisition and Enhancement”

National Association of Health Data Organizations

Little Rock, Arkansas
November 7-9, 2019

FAIN: R13HS026663

Principal Investigator: Denise Love; Norman Thurston, Ph.D.

Abstract

Purpose

The purpose of the practice domain sessions during the National Association of Health Data Organizations' (NAHDO) 34th Annual Meeting was to highlight existing technical support approaches states use to fill information gaps and identify opportunities for improvements in enhancing and sharing data. Data acquisition and enhancement are used by states to support population health, research, policy, and health system use of statewide data sets.

Leveraging NAHDO's meeting infrastructure and national network, this project is designed to improve the utility and comparability of the statewide hospital and claims-based data sets. This year's meeting featured sessions structured to promote state-to-state sharing of lessons learned and best practices. This is the second of the functional practice areas proposed as the core domains of the three-year project: 1) data quality assurance and improvement, 2) data acquisition and enhancement, and 3) analytics and actionable reporting.

Scope

The sessions on data acquisition and enhancement were held November 7-9, 2019 in Little Rock, Arkansas as a part of NAHDO's 34th Annual Meeting. The meeting brought together 180 health data stewards, experts, users, and vendors to learn the latest developments in health care data collection and analytics, including enhancement of data using data linkage and methods for acquiring additional public data sets.

Methods

We used NAHDO's 34th Annual Meeting in two ways for this project. First, we embedded data enhancement topics into general plenary and abstract sessions held throughout the meeting. We also included facilitated breakout sessions on three aspects of data enhancement: 1) data linkage, 2) missing populations, and 3) missing transactions. The facilitated discussions were led by NAHDO Board and Project Advisory Committee members who were supported by session scribes to document the two-part discussion for each topic. NAHDO staff worked with each facilitator prior to the meeting to identify the discussion framework, questions, and expected outcomes from the discussion (SEE DISCUSSION GUIDES). An evaluation form was distributed at the conclusion of the discussion sessions and results were presented on the final day of the workshop, November 7. (The meeting agenda, Roundtable Discussion Guides, and roundtable evaluation results are included in Appendix 1)

Results

The proceedings from these discussions provided the Project Advisory Committee and NAHDO staff guidance on the types of resources and collaborative priorities desired by data practitioners in data agencies across the country. Filling data and information gaps is a major issue for state data agencies and is a way to add value to existing data sets. Exploring options for timely and cost-effective data enhancement becomes more urgent as demand for information about vulnerable populations, health equity, and health outcomes increases. The conference platform and in-person interaction laid the foundation for virtual educational and technical activities until the next in-person meeting during Year 3.

Project Executive Summary

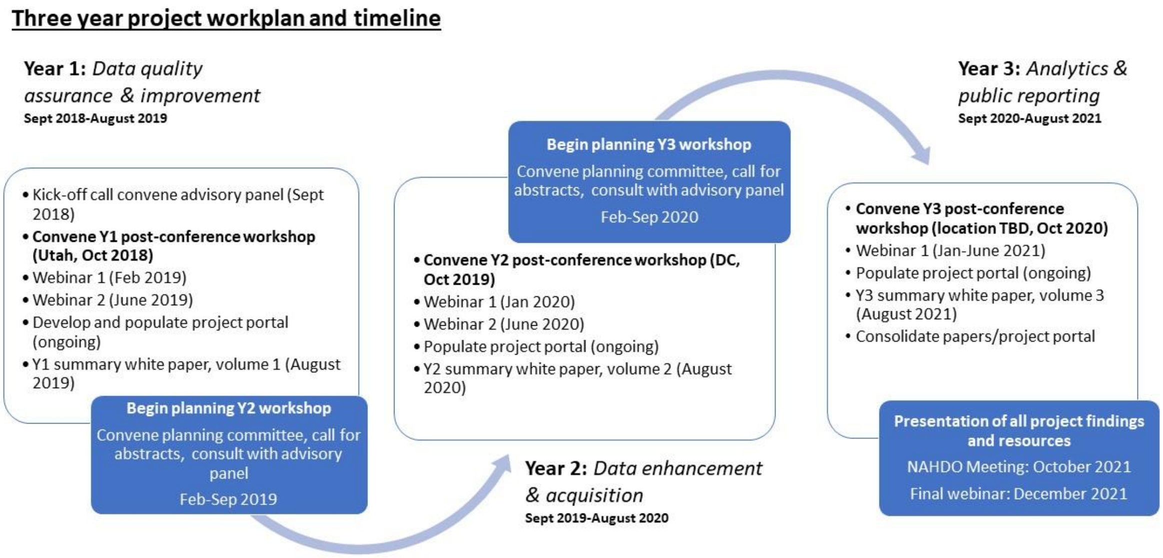
State Health Data Organizations (SHDOs) maintain 48 inpatient hospital discharge data reporting programs. Over 20 states are in various stages of All-Payer Claims Database (APCD) implementation. These SHDOs face unique challenges related to statewide data collection and public dissemination of the data.

This report[1] titled “Improving the Health Care Data Ecosystem through Data Acquisition and Enhancement” is the second in a series of three on topics related to domains of the SHDO practice:

1. data quality assurance and improvement (project year 1),
2. data enhancement and linkage (project year 2), and
3. analytics and public reporting (project year 3).

Figure 1 illustrates the project framework over the three-year period of funding.

Figure 1. Three Year Project Workplan and Timeline



Data Acquisition and Enhancement

States have a great deal of experience in implementing the acquisition of public data sets and linkage to existing data sources. This meeting provides opportunities for shared learning and shared resources to advance and harmonize these practices, especially ones pertaining to the interstate exchange of cross-border health care data. This report highlights notable practices and suggests activities that NAHDO can undertake to assist state enhancement approaches.

---

[1] This report, the first in a series of three topics, was written by the National Association of Health Data Organizations (NAHDO) as an activity under a three-year conference grant project award titled *“Improving the Utility and Comparability of Health Care Data for Health Services Research, Policy Decisions and Transparency Reports” (R13HS026663)* in the AHRQ category of “Dissemination and Implementation.”

Linking state health care data with other data sets poses technical challenges in applying probabilistic and deterministic methods. Legal and privacy issues are more important than technical challenges, especially when data sources to be linked are maintained by different agencies. Legal and privacy barriers to obtaining identifiable data include negotiations, paperwork, and agreements that add time and resources to the linkage initiatives and ultimately hinder timeliness. Information that can then be shared or made public must be negotiated upfront and may be limited.

The group cited a “Russian Doll” example: with each data set acquired for linkage, additional and different restrictions may leave very little remaining data to use or release. Data elements pertaining to Medicaid, behavioral health, and social determinants of health (SDoH) pose additional and unique barriers due to restrictions to data access and uneven capture of essential data elements. While transparency in the purpose and the process are important, developing relevant information about data submitter’s own populations will help build support for linkage projects.

At NAHDO’s 33rd Annual Meeting in 2018, during Year 1, a roundtable discussion on data acquisition and enhancement identified areas where NAHDO can provide support:

- Work with states to develop guidance around encryption/hashing of identifiers and design of “stand-alone” data sets that are pre-linked internally and anonymized so that users don’t need access to raw data sets
- Maintain the APCD-CDL™ and develop guidance for linking clinical data fields to administrative data
- Guidance around “core code of conduct” for release of linked data

The 2019 roundtables build on these themes and expand on them, as summarized below.

Plenary Keynote: Gopal Khanna, Director of the Agency for Healthcare Research and Quality (AHRQ)

A plenary session keynote on November 7 featured Gopal Khanna, Director of the Agency for Healthcare Research and Quality (AHRQ), which funds this conference grant. This plenary address set the stage for three roundtable discussions that followed.

Director Khanna described how AHRQ is leveraging its core competencies in data and analytics, health services research, and practice improvement to drive innovation forward. Predictive analytic models and linked data sets are ways the health system can have an effect on intractable health and social ills, epidemics, disparities, maternal mortality, etc. By unlocking the power of data and analytics it is possible to give policymakers real-time information they need urgently to respond to emergencies and disasters and assess the impact of events and policies on communities.

AHRQ is building a data platform for communities. Linking from multiple sources across public and private sectors promises transformational impact, but the challenges, such as establishing data use agreements across entities, cannot be understated. Director Khanna shared evidence of real-world applications. One example is how West Virginia used data-driven strategies to reduce opioid-related mortality. It required data sharing between Corrections, Public Health, Emergency Medical Systems, Prescription Drug Monitoring Program, and others to create more timely and accurate data about prescribing patterns, circumstances around patients needing medications, and approaches being used for other disease outbreaks.

Director Khanna discussed this NAHDO-AHRQ project which is happening during a time of growing need for health care data. Strategies to improve data quality, advanced analytics and public reporting are needed to produce high-quality data and information for policy, industry, and research uses. Year 3 of this project will address innovative use-cases and methods that can be replicated across states, including claims-based measures. AHRQ will learn with and from NAHDO during this project.

Director Khanna wanted to emphasize the importance of SDoH in improving health outcomes. AHRQ is developing an SDoH database for linkage with other national and state databases that can provide insights about food stability, housing, education. Another AHRQ initiative is the development of a synthetic database that is nationally representative and will be available to state and local governments for use. AHRQ will support NAHDO's efforts and work with NAHDO to improve data, develop new products, and create new data platforms to meet future challenges. Director Khanna left attendees with three considerations to move the conversation from barriers to practical solutions:

1. You must rely on the data you and your stakeholders have during a crisis. Understanding data sharing and access between entities prior to a crisis is recommended.
2. Think nationally and act locally. Find new ways to use AHRQ's platforms to obtain a more complete picture of populations in your own communities and how to link your own state data to these data. Let AHRQ know how it can help.
3. Learn from each other and with the private sector to find strategies to improve actionable data across the system.

The plenary keynote led to the project roundtables. Roundtable topics and discussion summaries are included in the section below.

Summary of Facilitated Discussion

The three facilitated discussion roundtables were conducted for the following topical areas:

- Missing Populations
- Missing Transactions
- Data Linkage

Each roundtable topic was divided into two facilitated sessions:

- Breakout 1: Issues and Challenges
- Breakout 2: Policy Issues and Next Steps

The following sections summarize these facilitated discussions.

Missing Populations: Options for Obtaining Data on Missing Populations (FEHBP, SUD, ERISA) – Facilitator: Kenley Money, Arkansas Center for Health Information

APCD systems provide previously unavailable valuable system-wide data about privately and publicly-insured populations which yield essential information about patterns of utilization, price variation, and health system performance. However, there are important gaps in information due to missing populations, limiting the completeness of the data collected. These gaps reflect the fragmentation of the U.S. healthcare system and inhibit a full understanding of health system performance, costs, and outcomes. Missing data on commercially self-insured, substance use disorder (SUD), and federal employees are occurring due to legal restrictions. States are seeking to fill these information gaps through a combination of administrative workarounds (voluntary opt-in activities) and advocating for federal policy changes governing these data.

The group identified the following missing sources of data and information about important groups within the population: SUD, Employee Retirement Income Security Act (ERISA), homeless, incarcerated, self-pay, Federal Employee Health Benefits Program (FEHBP), Indian Health Services, Veterans Administration (VA), Tricare, Medicare Advantage, Health Information Exchanges (HIE), all encounter data, workers' compensation, dental, accident insurance, Medicaid fee-for-service and managed care, cross-state data, sub-services, Prescription Drug Monitoring Programs (PDMP), Emergency Medical Services (EMS), master provider file, railroad retirees, religious groups, uninsured, out-of-pocket services.

Missing Transactions: Non-claims Data Capture (Population-Based Payment and Alternative Payment Models) – Facilitator: Maria Diaz-Perez, Center for Improving Value in Health Care (CIVHC)

APCDs are based on claims transactions that provide system-wide information about health care costs and utilization. As health care payment arrangements evolve away from fee-for-service claims payments, there is a corresponding decrease in claims containing useful financial fields such as charged amount, paid amount, and patient liability. These financial data elements are important to states who establish APCDs to support price transparency, value-based care, and price benchmarking.

Missing data examples include missing financial fields in capitated or 'global payment' arrangements, non-medical expenses such as housing, and pharmacy rebate data.

What are the challenges faced by states in capturing missing data/transactions?
There's often no place in the claims or data feeds to capture this information, further complicating the capture of financial fields. Challenges related to missing transactions depend on the use-case. It is difficult to accurately determine total spend when a significant percentage of spending in a state is based on "global payments". In some cases, data are missing that are not recorded. Plans may make the network panel complicated in order to inflate Per Member Per Month (PMPM).

States suggested there should be a framework of expenses needed to be captured for Alternative Payment Models (APMs) and methods for better attribution for those patients enrolled in multiple Accountable Care Organizations (ACOs). One state creates a separate data feed to capture ACO information that is integrated into their APCD.

Capturing financial and physician-level data fields in the myriad of evolving payment arrangements between providers and payers are complicated. Discussants recommended that the use-case will need to dictate the level of collection and reporting while recognizing the realities and limitations of provider-payer arrangements.

States vary in their approaches to capturing APM data, with some (e.g. Massachusetts) collecting under statutory requirements, some collecting for value-based contracting purposes, some for benchmarking cost trends, and some through rate setting.

Linkage of APCD, Hospital Discharge, Clinical, Census Data – Facilitator: Kyle Russell, Virginia Health Information

Why does this matter?

- Claims data by itself is good but is more meaningful when linked with other data sets
- States are wanting to link APCD data, HIE data, and Health Department data to get a more complete picture of Health for their state, regions, and counties.
- Arkansas, Oregon, Maine APCDs, Collect Hospital Discharge and Emergency Department (ED) data
- Maine APCD gets hospital encounter data
- California does not get address but gets Social Security Number (SSN)
- Policy issue: Different states are legally allowed to collect different data

Conclusion

A common theme was that the governance of public data sets needs to be modernized or the private sector, with superior resources and perhaps less accountability, will dominate the data ecosystem of the future. Therefore, it is essential that states and stakeholders considering new approaches, models, and tools that preserve data integrity and protections while also expanding the user base and utility of the data they manage.

Data will never be perfect or perfectly complete. The objective is to identify best practices in technical and policy solutions to fill these gaps and also to know at the point when the data are “good enough” for the use-cases it supports. As the health care delivery system structures evolve, the data systems also must adapt to apply new technologies, methods, and tools to augment and fill data gaps in order to remain relevant to data users.

Timeliness of the data will continue to pose challenges for SHDOs and may require new approaches to the typical data submission and reporting cycle. Data acquisition and linkage are proven ways to fill data and information gaps but also introduce additional legal and privacy considerations. States working together to influence national governance policies is a long-term but critical need. Data release, enhancement, and exchange are all resource-intensive at a time when the SHDO is constrained by stagnant or diminished funding, thus requiring innovation and collaboration across states, sectors, and stakeholders to test and implement creative solutions.

Appendix 1: Missing Populations

Issues and Challenges

What are the challenges you face in incorporating missing populations?

- Limitations in providing real-time information (example: When police are called to an emergency, having real-time information would help to identify a bad street drug outbreak).
- Incomplete datasets force assumptions: We need enough of a complete data set for use so we are not forced to make assumptions. (example: ERISA not included in a statewide data set introduces bias as it leaves a younger population).
- Analytic challenges: We don’t know if the missing data is different from the non-missing
- Unable to support continuity of care, especially for mental health and SUD, for entities outside of the healthcare system (e.g., justice system)
- Challenges to linking studies: Linked data are only as good as the less complete data set. This can limit the overall value of a complete health care utilization data set linked to another less complete data set.
- Patient attribution
- State data release rules are different creating cross-border data sharing challenges
- Data restrictions
- Data not included in the mandate
- Prison data: there are a lot of county jails and not a unified data system for this data

When collecting needed data is not possible, what adjustments are states making to compensate for these gaps?

- We use similar data sources with a disclaimer about substitute data.
- Workarounds: To substitute missing ERISA claims, we look for large insurers of comparable size and leverage fully-insured data and regression analysis to predict those claims and create a stable model. We use additional resources to compare firm size, number of employees, and percentage of self-insured.
- We backtrack into the claims.

What techniques and approaches do you use to obtain missing population data?

- Texas: homeless data has been obtained from homeless shelters. They register the homeless population and have clinics in the shelter (FQHC).
- Also for homeless data, check hospital discharge data to see if the ICD code is used for the homeless indicator. (Often under-used by physicians due to extra administrative burden.)
- Onboarding VA data through eHealth exchange.
- Obtain information from HIE/clinical data sources; however, there is missing financial analysis in clinical data
- There is movement in New Mexico to receive Indian Health Service (HIS) data but there is some sensitivity, so transparency and partnership are important. A lot of services are being captured through Medicaid, not IHS.

- States are working with employer groups to show how useful the APCD is to encourage voluntary submission for ERISA data
- Use hashed IDs to represent individuals when you don’t collect identifiable information, combining elements of demographic fields.
- One state uses a master patient index to link another demographic file (with identifiable information). We send data to CRSP vendor to match.
- Obtain agency authority to mandate reporting when possible

What challenges, when do you go dormant?
- One state said public health will not share data, period.
- Education and workforce often will not share.
- Many states find it difficult to get mental health data.

What are some lessons learned and best practices that you can share?
- Be creative and persistent
- Bring multiple stakeholders around a common goal to provide more leverage
- Be specific as to what data you need and for what purpose(s)
- Start small and achieve success to build trust
- Involve the attorneys early
- Speculative, be careful
- Build a story, personalize it
- Have a good/solid use-case
- Be thorough and consistent
- Data collectors are risk-averse so respect the risks and be responsive
- Be patient

How are you evaluating the quality of your data and your users’ experience after adjusting for missing population data?
- Texas: Driven by outcomes. Did we achieve the goal? Did it move the needle? Assuming everyone agrees on the needle.

Any last comments/thoughts/questions about the technical issues regarding missing population data?
- West Virginia: We passed legislation that requires detailed opioid-use reporting. We did this by requiring a single incidence of overdose reporting, which does not indicate substance use disorder. Providers agreed to this and report.
- Clinical is deep but narrow, proxy data elements to claims make it broader

Policy Issues and Next Steps

Why will agencies or other data stewards not share?
- Specific funding streams cause perpetuated silos. Sharing may reveal redundancy at the risk of losing money.
- Collaboration may look dodgy to the legislature and create an opportunity for spillage.

- You don’t know how the data are being used and bad things can happen. We need to build platforms for partnerships to address data privacy and an appropriate understanding of the safeguards. Also, develop personal relationships.
- Immunization records: the department of health thinks everyone should have to come through them to release information and they make the data unusable to protect their data reporting numbers. In the end, it keeps the registries limited. States often blame the CDC as the reason they cannot share.
- There is a sense of threat which causes data agencies to stonewall the release of data

What are the analytic challenges of missing data?

- Arkansas: Discovered substance use disorder treatment data collection issues due to US Code 42 CFR Part 2. Fortunately, the largest state player did not redact the data.
- Missing ERISA data. The number of procedures and services are gone which compromises the data. This is impactful in states where the ERISA population makes up a large portion of the data.
- Minnesota no longer reports total spending because of the missing ERISA data which affects chronic condition reporting. We need to access the volume of missing data and determine precision. There is a need to connect with employers and get them “back in the family.” Requires hitting the pavement more than small shops have the time and ability to do.
- We hope that the fully-funded plans are similar to the self-funded. Fortunately, some states can compare data before and after the Supreme Court decision in *Gobeille v Liberty Mutual* which disallowed mandatory submission of self-funded insurance data to states.
- Minnesota is looking at the data before and after the *Gobeille* decision to determine demographics and adjust where needed. How have the estimates changed pre/post-*Gobeille*? Then we will look at the scope to better understand the proportion of data. We work closely with TPAs and brokers to have them report the enrollment for which they are responsible. It helps to know the largest employees to start. Also, consider how coverage has changed and how insurance markets have changed to allow comparison to pre-*Gobeille*.
- The private self-funded market is growing in New Hampshire. Monitoring market size and coverage in other places like Medicaid. They have to keep asking that question. Plan design isn’t that different between fully-insured and self-insured. In New Hampshire, the insurance department has a market review hearing every year which provides helpful information.
- Minnesota: 40% of the commercial population is under an ERISA self-funded plan
- Maryland: Self-insured ERISA population is younger. Looks like population health indicators show declines in health outcomes, but due to missing data, it may not be representative of the entire covered population. Are public and private self-insured populations comparable? This is important to know.
- We need VA data to capture the full opioid epidemic.

What should be done about ERISA data?

- Congress could create an exception for state claims databases
- The U.S. Department of Labor could: 1) collect and distribute the claims or 2) states could collect and send claims to DOL
- We need to rally big employers. Lobbyists are not concerned about APCDs but don't want to change anything because adding exceptions would have implications outside of APCD data collection.
- Some states have a small percentage reporting voluntarily, to opt-in.

Other comments or thoughts regarding missing population data?

- We need out-of-pocket information to look at the aging population.
- We would like accident insurance to look at why people are initiating opioid treatment. The transportation department may have some of this information. Also, the medical costs paid through accident insurance.

What are the policy/legal issues that support obtaining missing population data?

- The impact of the decision to use outside data on medical services to cut down wait time at the VA may be a way to receive VA data.
- Cancer registries are a good area for collaboration. We have care patterns and registries have staging of the cancer.

Appendix 2: Missing Transactions

Issues and Challenges

- What are the challenges?
    - “Missing data” entails data that aren’t compatible with a traditional APCD claims reporting format
    - Fee for service equivalents are not comprehensive
    - Issues trying to “retrofit” APM data to APCD claims layouts
- APMs are also in constant flux, the group agreed that a technical approach should be to used determine what is feasible for reporting
- Some states (e.g. Massachusetts, Oregon, and California) reported out how they are capturing APM data and working with payers to collect this information
- Discussed flags for global payment in APCD to identify claims
- Broad themes/conclusions from the group:
    - Consider limitations of capturing member-level data based on the inherent structure of APMs
    - Develop broad, workable categories for APM definitions, understanding that these payment models are constantly evolving and vary from state to state, provider to provider, etc.
    - Work with payers and other stakeholders to develop a methodology for APM reporting that is meaningful but not overly burdensome
        - As Massachusetts noted, build in opportunities to provide feedback to data submitters and share trend data; also, ongoing feedback on data submissions is important to work through issues as they are identified
        - Insight into accounting principles/concepts of payers would be helpful in determining reporting categories
    - Differences in data capture and reporting; end use-case often drives requested data elements

Policy Issues and Next Steps

Policy priorities:

- Define broad categories for APM reporting
- Define specific use-cases to determine the granularity of reporting
- Develop talking points to use with policymakers for APM reporting requirements
- Survey and engage stakeholders early and throughout the process
- Get technical assistance from payers to under financial nuances and components to create more meaningful reporting

Appendix 3: Linkage of APCD, Hospital Discharge, Clinical, Census Data

Issues and Challenges

What are the Challenges?

- Misspelling, nicknames, married names, and maiden names complicate matching
- Policy issue: Different states are legally allowed to collect different data
    - Oregon: statutes forbid collecting SSN but allow name and address.
    - Arkansas: laws do not allow the collection of any Personally Identifiable Information (PII)
    - California does not get the address but gets SSN
    - Kentucky: APCD is just now getting a name, address, and phone number. The hospital association collects the data but pushed back against allowing it to be included in the APCD.
    - Arkansas Hospital Discharge system collects all PII. Questions arise about how it can be released or why it cannot?
- Linkages based on provider are an issue because of multiple NPIs and multiple clinic affiliations
- Multiple addresses for the place of service. A solution based on the use-case and mailing address is identifiable.
- Wisconsin: Collects all PII. Even when you get addresses there are still issues. Use what elements you have to get a 60-65% match.

What are the techniques?

- How reliant on SSN are state data systems? Is it going away?
- If SSN is not used, is this a risk to false-positive linkages?
- Massachusetts: Case data and APCD SSN is hashed. Actual names & addresses are not released and MA is exploring mixing the two elements for a hashed number. Only charge data are collected in the hospital case mix data set while APCD includes payment data.
- Hospital discharge data support risk adjustment in many states; APCDs are not used for this purpose.
- Oregon has a voluntary initiative to link APCD to other data sets (hospital discharge, PDMP, EMS, etc.) using the first and last names, date of birth, postal code in a probabilistic matching algorithm. FastLink is a software package that includes phonetic matching. They are now looking at the thresholds for matching.
- What is the cost of linkage software? There is free software but programming ability is important. Avoid using SQL in Access. R is good. There are packages available for a price.
- Wisconsin: When matching persons there is a tradeoff between false negatives and false positives.
- Oregon SAS shop using “The Link King” to match records.

- Other available software packages: Alterix includes a fuzzy match, ChoiceMaker (open source), IBM Initiate (paid), and Red Point.
- Arkansas: data linkage vs. entity resolution; linking data sets together is different than finding and resolving records to a particular person but resolution needs to be done well in order to appropriately link data sets together.
- California: Uses an open-source package using machine learning. Writing clues (rules) to automate 97% of their linkages. SSN was the primary weight but it was not very reliable. Over time, the clue deck grows organically and improves precision.
- California uses some referential data sources. They do not collect addresses but do get geocodes.
- New York: Looking for a gold standard but is it achievable? Geocoding becomes conflicting when multiple data sources are used with different addresses at different time points.

How do you clean the data?

- Are United States Postal Services (USPS) address services useful at all? No one has tried this.
- If the data user “corrects” the data, are you changing what the data is actually saying?

Lessons learned / Best practices?

- The percent of records that don’t match varies by states and data set.
- Some states hash everything coming in even if they have the legal authority to collect PII. This approach makes volunteer submitters more comfortable. Large employers are beginning to come on board.
- Perfection is not always needed. We must understand what is good enough for various use-cases. Don’t spend extra time and money pursuing perfection when it is not needed.
- California: Validated data set by comparing pre-linked data set to MediCal data and validated by putting the linked data into analytic use.
- Researchers: The threshold of precision is useful for a researcher to know upfront when deciding how to go about their work. Postal codes are bad. Use Census Bureau data. Census tract is most useful depending on the use-case. Population weighted centroids are promising.
- There are five levels of geocoding and uses are based on data governance Census-defined geographic areas roll up census tracts.
- Provider matching continues to be a challenge as NPIs are difficult to use due to the one-to-many relationships.
- Proprietary tools are out there. For example, CarePrecise sells provider data that is cross-walked and cleaned. It is updated once a year but may be costly.
- One state realized a 60% linkage for ambulatory services with a mandatory requirement to put service location on the form.
- From a payer perspective, regulating what comes in on a claim only works to a point. If the claim is filled out completely, not correctly, the payers have to pay it. Updates compound the problem. The actual physician is not always the one that is on the form within a hospital or clinic. It depends on what will benefit the patient or clinic most.

- What address is the most important? Billing address is what payers worry about. The payer is not worried about exactly where the service took place.
- Publishing data is one way to improve data. In Maine, facility location and costs conversations were triggered by clinics not seeing their data on the cost webpage.
- Invalid data is disqualified. It is good enough.
- HIE mix with claims data needs a provider directory. In Delaware, the HIE is also the administrator of the APCD so the same NPI is assigned. They also use clinical data to clean the master provider file but cannot share that file.
- For telemedicine, Medicare rules use the location where the provider is sitting, not where the patient is receiving care.
- Information on the claim: Providers are required to put a physical address. Gets lost during electronic adjudications etc. The doctors are submitting it.
- Other issues: Rebates
- Different data formats: who stages and normalizes
- Duplicate records
- Colorado University Record Linkage (CURL): Link events to events APCD to electronic health records (EHR). The University of Colorado is creating a linkage product and is open to partnerships prior to full release. Medicare claims individuals to EHR individuals. Hashing can be applied to this product.
- There are lots of products and lots of algorithms but how well do they scale?
- Rhode Islands uses Power BI for website reports
- Strategic linking can reduce the reporting burden but must balance with privacy controls.
- The more reporting that is demanded upfront reduces efficiency in the doctor's office. They need time to do patient care.
- Consider how technology may change as processes are developed. Be adaptive going forward.

How do you evaluate data quality?

- Arkansas: Birth data is the cleanest data set in our state
- For other states, data quality is a big issue.
- Self-coded, birth data is hard because of the baby name change when looking at claims data only. Home address increases the match rate in this case.

What are the policy/legal issues that support obtaining data linkage? What are the barriers/obstacles to overcome?

- Arkansas hashes identified data to prohibit re-identification
- Oregon: different release policies. Trade secret protection in legislation. Hospital discharge data is federal privacy and security of PHI. Hospital discharge data ownership is an issue. Data sharing is governed by contract.
- Hospital associations: some collect hospital data under the law and control it according to that law. Some have no law so the hospital association makes up their own governance.
- Virginia: APCD is more controlled than hospital discharge data release.

- Utah: APCD can't be used for health surveillance. ED and inpatient discharge can be used for surveillance. Internal requests in Utah are pretty straightforward, mainly using institutional review boards. Utah's Health Data Committee and its data-use sub-committee are appointed by the governor. Easier for external entities to purchase the data from the APCD than to go through no-cost data sharing contracts. If they bring money to the table, the whole system goes faster.

Have you overcome them?

- Oregon: Link data at Oregon site, data must be unidentified when it leaves the Oregon APCD site. Oregon wants to verify that identifiers were removed before it leaves their control.
- Cloud-based: the University of Colorado uses virtual machines on a database housed at the university. Cannot link out to the internet. This model is easier to control than DUAs on data that completely leave the system.
- Can VMs be the data release plan of the future?
- Virginia: Used to be a self-service model: users did not feel confident using the data. Have shifted to a data extract model. Virginia cuts the data and early linking and analysis instead of giving everything to the requester which can overwhelm them.

Federal and State government steps to facilitate linkage

- Federal requirements state that foreign nationals cannot work in the data center. Sometimes Canada is okay.
- FedRamp Standard for data security.

What are the privacy /security and transparency issues? How do we reconcile those issues?

- CDC makes requests using multiple APCDs.
- Is there a better, more timely solution for data exchange? Similar to vital statistics, a software solution in which state restrictions and rules are embedded in the software platform, facilitating automation and cross-state exchange. For example, if Utah and Florida agree, then approved data elements from patients being seen in the opposite state are shared.
- Border states are ripe for sharing because of market overlaps and border crossings for care. We have learned that even states motivated to share data have found it difficult in terms of time and effort.
- Should epidemiologic and public health needs supersede the policy concerns? Using public health authority for urgent conditions should be permissible.
- Use reciprocity instead of trying to get data from other states.
- Who signs the agreements? Should there be a standard agreement NAHDO has a template made available to other states? A general template master can be a starting place for conversation. One has been developed for hospital discharge data.
- You don't have to exchange everything. You can exchange exactly what is needed.
- There are two sharing agreement types: Project-specific sharing and sharing all data in an ongoing agreement.

- Consider new data governance models like establishing an “honest broker model” in which a trusted third-party serves as the exchange agent based on state laws and requirements which assures independence and neutrality in the process. A state sends its list of pre-approved people or uses it to the honest broker then the honest broker is responsible for only giving out the minimum data needed to meet each request.
- Issues in data exchange or linkage: What will be the standard unique individual identifier? States vary in the collection and handling of PII so this must be considered in any solution. The honest broker model only issues data from true person matches, not false positives that are sent. The broker can define the standard of the unique identifier.
- More and more researchers are wanting to use multiple APCDs in their research but obtaining APCD data from states is a daunting process with mixed success. New models of multi-state analytics will evolve, such as enclaves or broker models.
- Data release laws vary by state. We don’t always have release power over all the data we have.
- Why use the cloud? Clients may not have the infrastructure to house the data. You can provide partners/users with data already aggregated etc. The identified data never have to leave the security of the solution.
- Some states (e.g. Mississippi) are prohibited from putting data on the cloud so any solution should accommodate these restrictions.
- Domestic servers: Utah, like other states, prohibits any data stored outside the United States. Where is cloud data stored? You have to define with your vendor where the data will be stored and if it should never physically leave the US.
- How would a universal health identification number (like an SSN) change how data is handled and released?

Last thoughts:

- Is there or could there be a standard release layout? Like a common de-identified analytic layout states could use to aggregate and share non-PII data?
- Resource tools and checklists to facilitate limited sharing between state APCDs?
- Honest broker models might be worth exploring and piloting
- HCUP model of aggregating data across states using AHRQ’s infrastructure?
- Could data use agreement (DUA) templates use NAHDO’s discharge data template as a starting point?
- Provider lists (NPIs) need to be frequently updated. Provide incentives to keep providers lists updated.
- Update state regulations to permit the collection of non-claims data.