

January 20, 2012 Vol. X Issue 3

Crossing The Data Chasm: A Data Extraction Roadmap For Healthcare Providers

By David Ivers, J.D., Mitchell, Blackstock, Ivers & Sneddon, PLLC

Across the country, payors are pushing providers to “bend the cost curve” through accountable care organizations, bundled payments, global budgets, or other models in which providers share risk.

Providers, for their part, are often unaware of exactly where the cost drivers are or how to alter them. Everyone has anecdotes and suspicions, but little hard data. Until they have that hard data, providers will be reluctant to move away from current fee-for-service or traditional payor-led managed care models, making successful reform doubtful at best. Thus, payors and providers have a mutual interest in an effective process to share data.

What follows are suggestions gleaned from experience in obtaining meaningful aggregate claims data from Medicaid, including sample data requests. Each state will have some differences in their particular approach, but the basic process outlined here can be applied in any state.

The same framework also would work with Medicare or private payors. However, Medicare traditionally has restricted its claims files to research organizations,^[1] and private payors guard their claims data with proprietary zeal. Eventually, all payors, not just Medicaid, will have to find common ground with providers on this issue, moving away from adversarial relationships to more collaborative ones for the new payment models to be successful.

While most smaller providers do not have the in-house expertise or staff time for extensive data extraction and analysis, it is a manageable and necessary task for provider associations or integrated healthcare organizations and their attorneys attempting to negotiate new payment models with state Medicaid officials.

The steps outlined in this article are presented in terms familiar to providers and their counsel, but in language that computer analysts also will understand. Simply sending an email requesting to know how many aged and disabled beneficiaries with chronic conditions a primary care provider treated last year, assumes the analysts will know exactly what the request means by “aged,” “disabled,” “chronic conditions,” and “primary care provider.” These terms can mean different things among programs, systems, or reports or to different people. Therefore, charting the course is necessary.

Most provider organizations want as much data as they can get on the population their members serve, whether those are primary care patients, heart patients, home health clients, nursing home residents, individuals with mental illness or developmental disabilities, or any other of the numerous patient populations. While a huge amount of data likely exists on the population of interest, a “data dump” would not be optimal. To find the right balance requires clearing a number of legal and practical hurdles.

1. Find the Door. When it comes to obtaining Medicaid data, the first step is finding the entry point to a state’s database. For example, in Arkansas, the databank is controlled by the Medicaid

Security Advisory Committee, which requires a formal application to obtain reports from the database.^[2] In New York, a Medicaid Confidential Data Review Committee controls access—different name but same general role. Its data request procedures are outlined on the internet as well.^[3] Some states may not publish their procedures on the web, but they will provide them upon request. Under federal regulations, state Medicaid agencies “must have criteria specifying the conditions for release and use of information.”^[4] Note that in some cases, the state may only be anticipating requests from research organizations, but the same or similar application form should suffice after some discussions with Medicaid officials on the importance of sharing the data with providers.

A number of states are creating all-payor claims databases, huge repositories of millions of claims from public and private payors, stripped of information identifying individuals. While payors want to use the databanks to detect variances in provider practice patterns, they offer great potential for providing useful information to providers, especially if providers are involved in their governance, as in Kansas’ Data Analytic Interface.^[5]

Some states may initially insist that providers use the Freedom of Information (FOI) route, which is generally not well suited for this endeavor. Typically under FOI laws, the information sought must already exist or, if the state is willing to compile it in the requested format, the requestor must pay for it, which can easily become cost prohibitive. This type of data extraction is not a quick, one-request-and-done process. Any serious effort likely will require repeated requests to get truly useful information.

An ongoing dialogue with Medicaid officials is a must, and hopefully they will name a point person from the Medicaid data warehouse to work with on an ongoing basis. If officials attempt to charge for the data, then before proceeding fundamental discussions need to take place on viewing the process as a joint effort. If those discussions are not successful, then it is time to weigh the value of the data against the cost of obtaining it and determine whether to proceed or perhaps limit the request.

2. Justify the Request. To obtain data in this collaborative fashion outside the FOI process, state officials likely will want justifications for the request that will require explaining how the information will improve administration of the Medicaid program. Federal regulations require state Medicaid agencies to “restrict the use or disclosure of information concerning applicants and recipients to purposes directly connected with the administration of the plan.”^[6] The regulation contains a short, non-exclusive list of allowable purposes. In the context of providing data to providers, the most relevant purpose would be “[p]roviding services for recipients,”^[7] though it would certainly be a good idea to supply more detail about how the data will be used to help the state improve services. When information is released, the persons to whom it is released will be “subject to standards of confidentiality that are comparable to those of the agency.”^[8] Only a few individual provider representatives and the attorney are likely to be granted access to the reports produced in response to requests. Make sure one of those persons loves to crunch and sort numbers in Excel.

3. Jump the HIPAA Hurdle. Emphasize to Medicaid officials that the request will not be seeking data on specific individuals at this system-wide level. This is important because the Health Insurance Portability and Accountability Act (HIPAA) comes into play when individually identifiable health information is involved, restricting use of any “protected health information” or PHI. You should ask Medicaid to assign unique identifiers to each beneficiary for purposes of cross-referencing data since this is not a practice-level request seeking to identify individual beneficiaries.

You should be able to avoid individually identifiable health information unless some combination of small geographic areas and small programs or unusual (but costly) diagnoses is involved. In those cases, it should be possible to resolve the issue by altering the request, e.g., by consolidating some

geographic areas. Nevertheless, government officials generally take a broad view of what might be individually identifiable health information. They may take the position that, regardless of efforts to alter the information sought, some of the data requests include PHI.

When that happens, the most useful HIPAA exception that allows payors to share this type of data is the one for “*health care operations*.”^[9] Under this provision, a covered entity, such as a payor, can share PHI with another covered entity, such as a provider, if both the payor and the provider have or had a relationship with the patient, the PHI pertains to that relationship, and the provider will use the PHI for a purpose that fits the first two paragraphs of the regulatory definition of PHI (which includes “population-based activities relating to improving health or reducing health care costs . . . case management and care coordination . . . and related functions.”) This encompasses most provider requests in the context of payment reform; provider associations will need to obtain the data on behalf of their members through a business associate agreement.^[10]

At the very least, the parties can use the HIPAA allowance for a “*limited data set*.”^[11] It permits a payor or any covered entity to disclose PHI to a provider or other third party as long as 16 direct identifiers are excluded, such as names, street addresses, telephone, and social security numbers, and as long as the parties enter into a “data use agreement.” The data can be used for research, public health, or the “health care operations” provision discussed above. Some states, such as Florida^[12] and Kansas,^[13] publish their data use agreement forms on the internet.

The protection of patient privacy is important, but this concern must be balanced with efforts to improve healthcare for patients. HIPAA was crafted with this balance in mind and should not be an obstacle to sharing detailed claims data with providers. However, the best course is to frame the request in a way that avoids reasonable HIPAA concerns.

4. Review Existing Data. Get familiar with data in the public domain, especially when dealing with government programs. Check the state Medicaid program’s annual report, along with any data sets, research studies, or white papers that may be available in hard copy or online. These public records will provide a good idea of how the state collects information and how a client’s program fits into the overall budget, but they most likely will not be detailed or specific enough. For instance, they may lump all physician services together, without distinguishing among specialties. They may place all inpatient mental healthcare with general acute-care hospitals. Or they may report all home-and-community-based waiver services together and not distinguish between elder care and developmental disabilities.

However, such public records can serve as a starting point for requesting more specific information and offer ideas on how to construct the parameters. In addition, they may serve as a baseline of sorts to double check your own data responses when you get them.

5. Get the Big Picture. Before honing in on a particular population, obtain some basic information on the Medicaid program overall, to provide context when assessing that population later. For example, make sure to know the total Medicaid expenditures and the total for each program of interest. Also, ask for the expenditure cutoff for the top 10%, 25%, and 40% of beneficiaries to know how many individuals in the particular population are in those high-cost groups. (Because high-cost beneficiaries often have multiple diagnoses and may access several different types of services, it is often easier to determine which of the beneficiaries in a particular program are in that high-cost group rather than ask which of those high-cost beneficiaries are in a particular program.) Some of this information may be in the annual report. Most of this information will be reported by state fiscal year, which is usually how it should be requested.

Do not forget that total Medicaid expenditures may include non-claims expenditures, i.e., an array of payments made to hospitals under various payment mechanisms that are not tied to a particular patient-day or a particular outpatient service. These payments are over and above their claims

payments. They include cost settlements, upper payment limit payments, and disproportionate share payments, and in some cases, payments related to provider tax collections. This is where having a contact in the program staff is helpful. That person can help identify these and other anomalies so that you do not rely on data that is incomplete or inaccurate.

6. Find the Categories of Services. Ask for the list of state Categories of Services (COS) codes that Medicaid requires providers to use on claim forms to identify their particular program. This is an important step because it likely will be necessary to limit the data requests to certain services that are represented by different COS codes. Some states have general categories and thus short lists, while others have long lists. The former may not be discrete enough, such as New York's list, which includes all home health and non-institutional long term care under the single COS of 15. In cases like that, see if the state's claim forms contain any more detailed identifier for specific programs. If not, it may be necessary to rely heavily on specialty codes.

The longer COS lists, like the one used in Arkansas, are extremely helpful, but sometimes they may be too narrowly defined. For instance, the categories may separate independent laboratory from physician and hospital laboratories when you want all laboratory services. They may separate out not only speech, physical, and occupational therapy, but also speech therapy evaluations from speech therapy treatment when you just want to know how much was spent on therapy. Ask the state to group claims from those categories that you want to consider together, but be careful to specify "unduplicated" beneficiaries; otherwise, beneficiaries who received services under more than one category of service in the data will be counted twice (or even more).

7. Ask for specialty codes. Often different types of providers bill services in the same program. The main use for specialty codes comes when trying to distinguish between different providers of the same service. "Place of service" codes and other information on the claim form also can help narrow the request.

8. Define Your Universe. You are finally ready to define your population universe. There are at least two major ways to go about this: diagnosis and services.

Diagnosis: Do you want to capture all beneficiaries by diagnosis, e.g., all those with a mental health diagnosis or developmental disability or certain chronic conditions? Be aware that the diagnosis is entered by the treating provider so will reflect the reason for that particular service, which may or may not be the "primary diagnosis." For instance, if a child with Down syndrome goes to the ER with pneumonia—the diagnosis will be pneumonia, not Down syndrome. So, if searching for potentially avoidable hospitalizations, using the primary diagnosis might work, but if searching for beneficiaries with developmental disabilities, that may not be the best approach.

Also, many times a diagnosis has changed, but no one has updated the records so the old diagnosis continues to be used, e.g., a child with ADHD as primary diagnosis receiving inpatient psychiatric care—something else is going on but that is the only diagnosis that shows up in the claims. Finally, there is great interest in obesity and tobacco cessation, but those often will not show up under diagnoses.

In general, first, second, and maybe even third diagnoses are good to request, but their usefulness depends on your purpose. Note also that seeking claims by diagnosis may require a more precise request by listing the corresponding ICD-9 or -10 codes that can fall within that diagnosis.[\[14\]](#)

Services. To identify all beneficiaries who received a particular type of services, e.g., physician services or hospitalizations or ER visits or mental health services, will require first identifying the correct COS and perhaps provider specialty codes. Every beneficiary with at least one billed service in that category will populate the requested universe.

At this stage the request might begin to look something like this one submitted by a statewide provider association working with Arkansas Medicaid on a new payment model for developmental disabilities. It uses the services approach:

Locate all unduplicated beneficiaries (de-identified) designated as U21 and 21+ who in SFY 2011 incurred a DD expenditure in the state Categories of Services (COS) defined as D1, 71, E2, E3, E4, 79, B3, B4, B5, C5, C6, C7, 08, T7, 27, D3 and 62.

This locates all child and adult beneficiaries who received one service in one of the developmental disabilities COS during the year. Now you have your population universe.

9. Extract the Jewels. Once the universe is defined, you can overlay different search parameters on that population. Many providers first want to obtain demographic data, expenditures, and other basic information on their patients, chiefly to determine the manner in which they are affecting the cost curve. Thus, using the developmental disabilities example above, the next request, applied to the universe, might look like this:

Report 1: Omnibus Report of Medicaid Expenditures by COS per DD Beneficiary for SFY 2011

- **Beneficiary ID # [randomly assigned for this report, not Medicaid ID no.]**
- **Age or Year of Birth**
- **Primary Diagnosis**
- **Secondary Diagnosis**
- **Third Diagnosis**
- **Beneficiary County**
- **Provider County**
- **Performing Provider Name**
- **State Category of Service**
- **Total Paid Per State Category of Service**
- **Total of All State Category Expenditures SFY 2011 for Beneficiary**

This request provides baseline data by de-identified beneficiary, which allows you to answer a number of important questions about the population, e.g., what is the average cost of a beneficiary's care by COS and for all Medicaid services? And you can now determine how beneficiaries in the defined universe relate to the total Medicaid budget, e.g., how many high-cost beneficiaries are in your universe? What are their most common diagnoses?

After this point, most providers will want to go beyond the baseline and evaluate opportunities for better care coordination or new payment models. This could involve requesting data on chronic conditions, potentially avoidable hospitalizations, ER visits, prescription drug costs, and the like. Thus, in the same example, a request might be submitted to obtain data needed for care coordination between developmental disabilities and acute care settings:

Report 2: Targeted Care Coordination Groupings of Expenditures per DD Beneficiary SFY 2011 (utilizing the data from Report 1) (separate U21 and 21+):

- **Beneficiary ID #**
- **Age or Year of Birth**
- **DD Habilitation (COS 08)**
- **DDS Waiver (COS D1)**
- **Outpatient Therapy (COS B3, B4, B5, C5, C6, C7, 79)**
- **DD Transportation (COS T7)**
- **Private ICF (COS 27)**
- **Public ICF (COS 62)**
- **Developmental Rehab Svcs (COS D3)**
- **Prescriptions (COS 56)**
- **Inpatient Hospital (COS 30, 31, 75, 76, 77, 84, 87, 88, CA, CB)**
- **Emergency Room (COS)**
- **Ambulance (COS 02)**
- **Other (All other COS which are not listed above)**
- **Total SFY 2011 Expenditures**

There are all sorts of variations on this idea that could produce more reports and different ways of sorting the data once the reports are obtained. For example, what percentage of these beneficiaries has a mental illness diagnosis or multiple chronic conditions? What types of beneficiaries are frequent users of emergency rooms? In what geographic regions are they concentrated? How many are dual beneficiaries? It all depends on what information providers need for their discussions with the state.

Cautionary Notes

- Do not expect the initial report to come back perfectly. It can take several tries to get it right, refining and adjusting the request until you finally get what you need.
- Look for traps. Do any of the results look suspicious or seem unusually high or low in terms of dollars, numbers of beneficiaries or in any other aspect? Often as programs evolved some services were categorized in ways that no longer make sense, which can lead the requestor down some rabbit holes if not detected and adjusted accordingly.
- When the request is made also can alter the dollar totals even though the same time period is specified. This can occur due to cost settlements, audits, and other events that change expenditure data for time periods that have already passed.
- Similarly, the data is only as reliable as the provider who codes it. Do not confuse reluctance on the part of the data analyst with the situation where the data simply does not exist in the system.

Conclusion

Providers need good claims data to evaluate and improve new payment models. When it comes to Medicaid, try to find a knowledgeable contact inside the agency who will champion your efforts.

Realize that you will have to do much of the work yourself. Strive for perfection, but accept there will be limitations no matter how good the database is. The possibilities for data analysis are almost endless, though your time and the state's patience may not be.

David Ivers is an attorney with Mitchell, Blackstock, Ivers & Sneddon, PLLC, in Little Rock, AK. David's health law practice includes extensive healthcare legislation and regulatory work, healthcare reform, Medicaid, Medicare, fraud and abuse, and HIPAA/HITECH, and significant public policy work on behalf of clients. He represents physicians, behavioral health, assisted living, developmental disability, and various other types of healthcare providers. He is a past chair of the Health Law Section of the Arkansas Bar Association.

[1] The Centers for Medicare and Medicaid Services (CMS) issued December 7 a final rule ([76 Fed. Reg. 76542](#)) that gives "qualified entities" access to Medicare claims data for purposes of aggregating the information with private sector data and publicly disseminating healthcare provider and supplier performance reports.

[2] [Application for Access to Arkansas Medicaid Database.](#)

[3] [Medicaid Confidentiality Regulations and Standards Question 3.](#)

[4] 42 C.F.R. § 306(a).

[5] See Phil Cauthon, KHI News Service, [How the Data Analytic Interface Works](#); Kansas Dep't of Health & Environment, [Health Care Market Reports](#).

[6] 42 C.F.R. § 431.300(a).

[7] 42 C.F.R. § 431.302(c).

[8] 42 C.F.R. § 306(b).

[9] 45 C.F.R. § 164.506

[10] See Modifications to the HIPAA Privacy Rule, 67 Fed. Reg. 53182 (Aug. 14, 2002).

[11] 45 C.F.R. § 164.514(e)

[12] See [Florida Limited Data Set Use Agreement](#).

[13] See [Kansas Data Analytic Interface Use Agreement](#).

[14] See, e.g., Edith G. Walsh, et al, [Cost Drivers for Dually Eligible Beneficiaries: Potentially Avoidable Hospitalizations from Nursing Facility, Skilled Nursing Facility, and Home and Community-Based Services Waiver Programs](#), Appx. B.

