Research Opportunities of a New Private Health Insurance Claims Data Set

Sean Nicholson

In July 2012 FAIR Health, Inc. began allowing researchers to obtain data from the National Private Insurance Claims (NPIC) Database, the largest independent private claims collection in the United States. A great deal of work has been done examining how Medicare patients are treated, and how the quantity, cost, and quality of that treatment vary across markets. Less is known regarding how privately insured patients are treated, even though they represent the majority of patients in the U.S. And few studies combine data on Medicare (or Medicaid) and privately insured patients to compile a more complete picture of providers' treatment decisions at the market level. The recent release of the NPIC helps address this imbalance by focusing on the privately insured market. For the past two years I have served as the Research Director of the Upstate Health Research Network (UHRN), a group of researchers from many universities that have developed recommendations for how FAIR Health can improve the accuracy and transparency of the products it offers to health insurers. In this article I describe the characteristics of the NPIC data and discuss the research opportunities it presents for health economists, health services researchers, and policy makers.

FAIR Health is an independent non-profit organization created as part of a settlement between the State of New York and major healthcare insurers who, in determining out-of-network reimbursements, used data products from a database owned by Ingenix, Inc., a subsidiary of United HealthGroup, Inc. The primary objectives of the 2009 negotiated settlement were to establish an independent not-for-profit organization, create a new database and enhanced data products, create a free website that allows individual consumers to look-up expected medical and dental care costs for specific procedures in their local areas, and promote academic and health policy research. FAIR Health collects medical and dental claims from over 70 payor-contributors on a rolling basis, aggregates the data to the level of a medical or dental service by current procedural terminology (CPT) code, healthcare common procedure coding system (HCPCS) code, or current dental terminology (CDT) CDT code in each of 491 "geozips"—i.e., a geographic area usually defined as the first three digits of a zip code. These data are assembled in modules that provide percentile information on the distribution of provider charges or allowed payments at the service-market level to help its health insurer clients establish out-of-network payment rules for its enrollees.

Health insurers must be certified by FAIR Health in order to participate in the data contribution program and they are required to submit all of their claims data for each submission period. Each claim in the NPIC Database must contain the following information in order to allow FAIR Health to provide an accurate depiction of the distribution of charges to its health insurer clients: procedure code(s), zip code where the service was provided, date of service, the provider's charge, days/units of service where relevant, anesthesia time where relevant, and any modifiers (e.g., procedure performed on the same day as the patient's evaluation and

Sean Nicholson is a professor in the Department of Policy Analysis and Management at Cornell University.

management visit). Many health insurers submit claims that also contain optional fields. Of particular interest to health economists, about one-half of the claims report the "allowed charge," which is the actual payment negotiated by a health insurer and an in-network provider, and is the sum of what the patient and health insurer agree to pay the provider. Other optional fields include: a patient's claim ID, the physician's national provider identifier (NPI), the provider's tax identification number, patient's date of birth, age, and/or gender, and the assigned diagnostic code(s) such as the ICD-9 code.

There are a number of characteristics of the NPIC Database that make it attractive to health services researchers and health economists. The NPIC Database contains a large number of claims that cover a wide variety of medical services, a broad geographic area, and a long time period (2002 through today). Currently, over 15 billion claims from payors whose plans cover 126 million people are in the NPIC Database, including claims for physician services delivered in office-based settings, hospitals, and ambulatory surgery centers (including anesthesia services); imaging and laboratory; durable medical equipment; dental services; and prescription drugs administered in physician offices and clinics (e.g., chemotherapy). The NPIC Database also has geographic breadth, with claims data covering every locality in the United States as well as Puerto Rico and the Virgin Islands; the data, as noted above, are divided into 491 geozips, which are generally defined as the initial three digits of a zip code.

FAIR Health does not require its contributors to indicate how many covered lives are represented in the claims contributed with each submission. Although this makes it difficult to determine the precise percentage of the privately-insured U.S. population that is represented in the NPIC Database, there are several ways to try to estimate this percentage. For example, the FAIR Health data set contains about 900 million claims in 2009 for physician services across the full spectrum of settings. The sum of the allowed charges across all of these professional services claims is $33.4 billion. If the claims without an allowed charge have the same mean allowed charge as the claims that do contain this variable, then the total physician payments in the data set would have been $66.9 billion. The Centers for Medicare and Medicaid Services reported that private health insurers and privately-insured patients paid a total of $285.6 billion for physician and clinical services in 2009. Thus, the payments for physician services in the FAIR Health data set represent an estimated 23.4 percent of national payments by privately insured patients and their insurers. The analysis above represents an underestimate of the percentage of medical services provided to privately-insured individuals that is contained in the NPIC Database if the services with a missing allowed payment tend to be relatively expensive services.

The comparative advantage of the NPIC Database for research purposes is likely to be for questions that require market-level information on private health insurance transaction prices at a point in time or over a long period for a wide variety of medical services. For example, Jonathan Ketcham and I are using FAIR Health data to examine whether private fees respond to changes in Medicare fees across services and markets. Specifically, we study whether physicians in markets that experienced relatively large increases in Medicare physician fees between 2003 and 2009 responded by negotiating relatively large increases in private insurance payments for the 1,000 most common services (Ketcham and Nicholson, 2012).

With geographic and service breadth and data back to 2002, the NPIC Database is also likely to be an excellent data source for difference-in-differences analyses that examine whether provider payments have responded to state- or county-level policies. Andrew Friedson is using FAIR Health claims data to see whether physician payments fell in South Carolina after malpractice reform was instituted there in 2004 relative to North Carolina for a common radiology procedure for prenatal visits (Friedson, 2012). Cary Gross, a member of FAIR Health Scientific Advisory Board, is interested in studying whether the type of breast cancer screening changed over time in response to state policies regarding mandatory patient notification.[1] Because about 50 percent of the claims report a physician's NPI or tax ID, researchers could study practitioners' patterns of treatment and changes in these patterns over time. The NPIC Database also can be used to examine whether differences in physician payments across markets are correlated with differences in the size or concentration of physician groups, and differences in health insurer market power. It is important to note, however, that although researchers can obtain a consistent encrypted physician identifier when requesting NPIC claims, they will not be able to merge in physician information from outside data sources based on a physician's NPI or name.

The NPIC Database would also be able to support studies on utilization of healthcare services over time as well as disparities in utilization across geographic markets. For example, one could study the impact on utilization following a legislative or regulatory intervention or change in health benefits design either favoring a particular procedure or limiting reimbursement for such a procedure. Likewise, one could study the impact on utilization of procedures associated with the "Choosing Wisely" campaign recently launched by the ABIM Foundation in conjunction with Consumer Reports. Similarly, one could study the impact of comparative effectiveness research publications on the use of the featured procedures or services.

While the NPIC Database is well suited for market-level analyses and utilization studies, it currently presents challenges for tracking the same patient over time. Health insurers are not required to include a consistent synthetic patient ID when submitting claims and, as a result, one cannot always stitch together a patient's complete episode of care for a particular chronic health problem or link an outcome from one claim to the care that a patient received previously.

A second limitation of the NPIC Database is, as mentioned earlier, that it currently does not identify how many covered lives generated the claims for a particular data contributor. For example, one might observe that claims were submitted for 10 million office visits in Dallas in 2010. But, without knowing how many Dallas enrollees were covered by the private health insurers who submitted claims data to FAIR Health, one cannot determine with certainty the average number of office visits per person per year. The number of covered lives by insurer is ascertainable from the department of insurance in the particular state, and would thus have to be merged with the FAIR Health data to derive medical service use rates. Regardless, the NPIC Database will allow one to examine changes over time in the use of medical services as long as the number of lives covered by the contributors remains fairly constant over time.

---

[1] Copies of selected academic papers that use the FAIR Health claims data are available at: http://research.fairhealth.org/Publications.

So far I have focused on research applications involving the patient-level NPIC claims data. FAIR Health is also willing to make the aggregated (across providers within a market) data available to researchers for a relatively low price. For example, some research projects may require the mean or median (or 75th percentile) physician charge or allowed payment in all 3-digit zip codes for all CPT codes for a particular year or set of years. If so, researchers can order the same data on the distribution of charges and allowed payments by CPT code and market area that FAIR Health sells to its health insurer customers.

In order to expedite use by academic researchers, FAIR Health, in consultation with its national scientific advisory board, has established a streamlined application and licensing process for researchers wishing to obtain the NPIC data. Researchers can submit applications on-line and can customize data requests by clinical area, geographic area, timeframe of data sought, length of license, and other criteria according to the needs of their studies. A simplified data use agreement applies to individual licensees of the data, and institutional licenses are available for academic organizations to allow groups of researchers to conduct multiple projects. Pricing for all offerings is oriented to the research community.   Further information about FAIR Health's Research Support Program, including the on-line application form, is available at http://research.fairhealth.org. FAIR Health personnel are available by e-mail and phone to answer questions about the data and to assist in the application process at all stages.

References

Friedson, Andrew I., "Medical Malpractice Damage Caps and the Price of Medical Procedures," working paper.

Ketcham, Jonathan, and Sean Nicholson, "Relative Prices, Payer Mix and Regional Variations in Medical Care," working paper.