The new ecosystem of health data keeps getting **Bigger**

Charlie Ornstein, ProPublica, charles.ornstein@propublica.org, @charlesornstein

Updated December 2016

Each year, the release of new datasets makes it more exciting to cover health care. No longer are we limited to comparing states to one another to look for differences. No longer are doctors' practice patterns protected by outdated privacy rules. We have entered an era in which we can compare one doctor to another. And what we're learning is that there are huge, seemingly unexplainable, differences among them.

This tip sheet offers both very broad data sources, as well as more granular ones. None of the data sets cover individual claims-level data, which require special permissions and often cost a lot of money.

Sign up for data updates from CMS:
CMS is the Centers for Medicare and Medicaid Services (yes, the acronym doesn't quite work). https://dnav.cms.gov/. Click on receive email updates in the lower right. CMS's Data Navigator is also a resource to look up data sources, but I find it overly complex to use. Socrata powers its own tool that allows you to look up all datasets available through its network: http://www.opendatanetwork.com/.

CMS Fast Facts:
This stat sheet is at the highest level, but it is the most current one the agency produces. https://www.cms.gov/fastfacts/ (Updated July 2016)

Medicare and Medicaid Statistics (top level, super useful):
https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/CMS-Statistics-Reference-Booklet/2015.html. Last released: January 2016. Released annually.

Table I.1
Medicare enrollment/trends

	Total persons	Aged persons	Disabled persons
July		In millions	
1966	19.1	19.1	--
1970	20.4	20.4	--
1975	24.9	22.7	2.2
1980	28.4	25.5	3.0
1985	31.1	28.1	2.9
1990	34.3	31.0	3.3
1995	37.6	33.2	4.4
Average monthly			
2000	39.7	34.3	5.4
2005	42.6	35.8	6.8
2010	47.7	39.6	8.1
2011	48.9	40.5	8.4
2012	50.9	42.2	8.6
2013	52.3	43.5	8.8
2014	54.0	45.0	9.0

NOTES: Represents those enrolled in HI (Part A) and/or SMI (Part B and Part D) of Medicare. Data for 1966-1995 are as of July. Data for 2000-2014 represent average actual or projected monthly enrollment. Numbers may not add to totals because of rounding. Based on 2014 Trustees Report.

SOURCE: CMS, Office of the Actuary.

Additional summary data and charts can be found: https://www.ccwdata.org/web/guest/medicare-data and http://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare MedicaidStatSupp/index.html. In all cases, there is a data lag of at least a year, sometimes longer.

CMS Program Statistics: Includes summary statistics on national health care, Medicare populations, utilization, and expenditures, as well as counts for Medicare-certified institutional and non-institutional providers. CMS Program Statistics is organized into sections which can be downloaded and viewed separately. The Medicare Enrollment Dashboard is an interactive online tool presenting monthly enrollment figures and yearly trends at several geographical areas, including national, state/territory, and county. https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/CMSProgramStatistics/index.html

Medicaid enrollment: https://www.medicaid.gov/medicaid/program-information/medicaid-and-chip-enrollment-data/index.html

Geographic variation: How Medicare utilization differs from state to state and from county to county. Data is available from 2007 to 2014. http://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Geographic-Variation/GV_PUF.html. This will help put other data sets in context. Last released: January 2016. Released annually.

Mapping geographic variation: Choose your conditions, choose your geography, choose your unit of measurement. Neat tool. https://data.cms.gov/mapping-medicare-disparities Last updated September 2016.

National health expenditures: Released at least annually. Last update December 2016.
https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NationalHealthAccountsHistorical.html

How much doctors are paid by Medicare and what they do for that money:
https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Provider-Charge-Data/Physician-and-Other-Supplier.html. ProPublica has created a nice visualization for each provider: www.propublica.org/treatment and so does the Wall Street Journal: http://graphics.wsj.com/medicare-billing/. Last released: May 2016; covers the year 2014. The site includes historical data for 2012 and 2013 as well. CMS has pledged to release this data annually. (Important caveat with this data: It only contains claims for services rendered in Medicare fee-for-service, not for those in Medicare Advantage. As a result, it will leave out lots of services delivered in regions of the country where Medicare HMO enrollment is high, and thus may not be representative.)

Another wonderful thing about this dataset is that it now includes demographic data on each doctor's patients, including breakdowns by gender; age ranges; race; dual eligible status; prevalence of common chronic conditions; and an aggregate risk score that allows you to compare how "sick" a doctor's patients are relative to peers.

If you want an earlier glimpse at national numbers, you may be able to find them a few months earlier at these two links:
https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/Part-B-National-Summary-Data-File/Overview.html or
https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/PhysicianFeeSched/.

Payments to hospitals for inpatient and outpatient visits:
http://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Provider-Charge-Data/index.html.

What drugs doctors prescribe most: In 2015, Medicare began publicly releasing prescribing data in its Part D program.
https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Provider-Charge-Data/Part-D-Prescriber.html The data now covers 2013 and 2014 and is spliced and diced in a variety of ways. ProPublica filed Freedom of Information Act requests for earlier data and has put the data online for free: https://projects.propublica.org/data-store/. You can see our app at www.propublica.org/checkup. This accounts for all prescriptions covered in Medicare Part D, but it's important to note that not every Medicare recipient is enrolled in Part D. Also, data is redacted when a doctor prescribed fewer than 11 prescriptions for a particular drug. Last released: August 2016. Updated annually.

You can look up Medicaid utilization data as well (but this is summarized by state, not by doctor): https://www.medicaid.gov/medicaid/prescription-drugs/state-drug-utilization-data/index.html.

Medicare drug spending dashboard: CMS released an online dashboard to look at drug spending in Medicare and Medicaid. https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Information-on-Prescription-Drugs/ And it released five years worth of data (total scripts, cost, cost for patients) for every drug used in Part B and Part D. https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Information-on-Prescription-Drugs/2015MedicareData.html (Updated December 2016)

Medicare opioid drug mapping tool: The tool shows geographic comparisons, at the state, county, and ZIP code levels, of opioid prescription claims in Medicare Part D. It allows you to see both the number and percentage of opioid claims at the local level. Obvious drawback: The data is from 2013 and, well, we're now in 2016. For an evolving epidemic, this data sadly lags far behind. https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Provider-Charge-Data/OpioidMap.html. Data for 2014 is included in the broader Part D file, mentioned higher on this page.

Who doctors refer patients to: Health care data guru Fred Trotter of DocGraph has requested so-called referral data from CMS for a few years now under FOIA. The agency has posted it online in a hidden corner: https://questions.cms.gov/faq.php?faqId=7977. Again, data is redacted when the cell size is less than 11. This includes data through 2015.

What durable medical equipment doctors order: In October 2015, Medicare began releasing data on durable medical equipment and supplies ordered and dispensed in Medicare Part B, organized by the ordering physician. There is an aggregate files (high level) and a detail file (aggregating by type of DME ordered by each doctor). https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Provider-Charge-Data/DME.html (Last updated November 2016 with 2014 data)

Utilization of home health services: In December 2015, Medicare began releasing data on services provided to Medicare beneficiaries by home health agencies. The Home Health Agency Utilization and Payment Public Use File (Home Health Agency PUF) contains information on utilization, payments, and submitted charges organized by provider, state and home health resource group. It is organized by home health provider, not by doctor who ordered the services. https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Provider-Charge-Data/HHA.html (Last updated December 2016 with 2014 data)

Therapy services in nursing homes: The Skilled Nursing Facility Utilization and Payment Public Use File (SNF PUF) contains information on utilization, payments, and submitted charges organized by provider, state, and resource utilization group (RUG). The data include information on 15,055 skilled nursing facilities, over 2.5 million stays, and almost $27 billion in Medicare payments for 2013. https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Provider-Charge-Data/SNF.html (Released March 2016)

Utilization of hospice services: In October 2016, Medicare released for the first time data on hospice services to Medicare beneficiaries and information on hospice agencies. The Hospice Utilization and

Payment Public Use File (Hospice PUF) contains information on utilization, payment (Medicare payment and standard payment), submitted charges, primary diagnoses, sites of service, and hospice beneficiary demographics organized by provider and state. https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Medicare-Provider-Charge-Data/Hospice.html (Released October 2016)

Additional **hospice** quality data can be found here: https://data.medicare.gov/data/hospice-directory.

Chronic conditions among Medicare beneficiaries: Tools to help you gain a better understanding of the burden of chronic conditions among Medicare fee-for-service beneficiaries, including prevalence, utilization of health services, and Medicare spending for specific chronic conditions and multiple chronic conditions. https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Chronic-Conditions/index.html (Updated January 2016)

What incentive payments doctors have received for using electronic medical records: http://www.cms.gov/Regulations-and-Guidance/Legislation/EHRIncentivePrograms/DataAndReports.html. Updated monthly.

Doctor addresses and specialties: These are part of the National Provider Identifier system, updated weekly/monthly, based on health professionals' self-reported information. Download the file here: https://www.cms.gov/Regulations-and-Guidance/Administrative-Simplification/NationalProvIdentStand/DataDissemination.html (link to download is at the bottom of the page.) An open-source API for the data is available here: http://www.bloomapi.com/ (Here's the GitHub link: https://github.com/untoldone/bloomapi). Physician Compare has additional information on doctors who participate in Medicare: https://data.medicare.gov/data/physician-compare. NPI data updated monthly. Physician Compare data last updated in September 2016.

Hospital characteristics: The CMS Provider of Services file, updated annually, contains data on characteristics of hospitals and other types of healthcare facilities, including the name and address of the facility and the type of Medicare services the facility provides, among other information (number of beds, ownership information). https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/Provider-of-Services/POS2015.html

Pharma/device company payments to doctors: Data for the last five months of 2013 and all of 2014 and 2015 is available at http://www.cms.gov/openpayments/index.html. Every June, CMS must release data from the prior calendar year. ProPublica has an easy-to-use website to search the data at www.propublica.org/docdollars. ProPublica's site also includes data from companies that were required to report this data earlier (because of corporate integrity agreements).

Quality measures: https://data.medicare.gov/ includes information about hospitals, nursing homes, doctors, dialysis facilities and home health care companies. Data includes demographic information (names, addresses, phone numbers), as well as quality and satisfaction measures. ProPublica used the

emergency room quality measures to power its ER Wait Watcher app (https://projects.propublica.org/emergency/).

Hospital, ambulatory surgery center and nursing home inspection reports: The federal government has also started making available online its deficiency reports for health facilities. It currently does so for nursing homes (http://www.cms.gov/Medicare/Provider-Enrollment-and-Certification/CertificationandComplianc/FSQRS.html, scroll down under "related links"), for hospitals (http://www.cms.gov/Medicare/Provider-Enrollment-and-Certification/CertificationandComplianc/Hospitals.html, scroll down to downloads) and for ambulatory surgery centers (https://www.cms.gov/Medicare/Provider-Enrollment-and-Certification/CertificationandComplianc/ASCs.html, scroll down to downloads). The Association of Health Care Journalists makes hospital inspection reports available in an easy-to-use interface: www.hospitalinspections.org and ProPublica does the same for nursing home reports: www.propublica.org/nursing-homes. Nursing home inspection reports are updated monthly; hospitals and surgery centers quarterly.

Speaking of which, here's a data-rich report from CMS on how the "great recession" led to a serious decline in nursing home oversight. (Released June 2016)

Custom calculations: Use this link to estimate Medicare cohorts for diseases, etc.: https://www.ccwdata.org/web/guest/pricing/estimate-study-size.

Accountable Care Organizations: Data from the shared savings program. https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/SSPACO/Overview.html

Obamacare health insurance exchange data

Which insurance networks do doctors participate in? Under the Affordable Care Act, insurance companies participating in the Obamacare marketplaces are required to provide up-to-date information about which doctors and facilities are in their networks. Links to insurers' directories can be found here: http://download.cms.gov/marketplace-puf/2016/machine-readable-url-puf.zip

Health Insurance Marketplace Public Use Files (Marketplace PUFs): These files are available for plan years 2014 through 2017 for states on the federal exchange--those that use Healthcare.gov. They include information on: Benefits and Cost Sharing PUF (BenCS-PUF) – Plan-level data on essential health benefits, coverage limits, and cost sharing; Rate PUF (Rate-PUF) – Plan-level data on individual rates based on an eligible subscriber's age, tobacco use, and geographic location; Plan Attributes PUF (Plan-PUF) – Plan-level data on maximum out of pocket payments, deductibles, cost sharing, HSA eligibility, formulary ID, and other plan attributes; Business Rules PUF (BR-PUF) – Plan-level data on the application of rates, such as allowed relationships (e.g., spouse, dependents) and tobacco use; Service Area PUF (SA-PUF) – Issuer-level data on the geographic coverage or service area (i.e., where the plan is offered) including state, county, and zip code; Network (Ntwrk-PUF) – Issuer-level data identifying provider network URLs; Plan ID Crosswalk PUF (CW-PUF) – Plan-level data mapping plans

offered in 2014 to plans offered in 2015; Machine-readable URL PUF (MR-PUF)—Issuer-level URL locations for machine-readable plan network provider and formulary information; Transparency in Coverage PUF (TC-PUF) – Issuer-level claims, appeals, and active URL data. This PUF contains data from PY2015 for issuers participating in the Marketplace in PY2017; Quality PUF (Qual-PUF) – Quality ratings data for plans in PY2017 pilot states (VA and WI).

Health Insurance State-based Marketplace Public Use Files: Health plan information including benefits, copayments, premiums, and geographic coverage is publically available on the various State-based Marketplace (SBM) websites, but CMS has also released downloadable public use files (PUF) so that researchers and other stakeholders can more easily access SBM data. Currently, the Health Insurance State-based Marketplace Public Use Files (SBM PUF) are available for plan year 2016 https://www.cms.gov/CCIIO/Resources/Data-Resources/sbm-puf.html

Enrollment in Obamacare plans: For methodology, click here: https://www.cms.gov/CCIIO/Resources/Data-Resources/Downloads/2014-Issuer_Enrollment-Data-Methodology-Final-10-19-16-508.pdf and for full data release, click here: https://www.cms.gov/CCIIO/Resources/Data-Resources/Downloads/2014-Issuer-Data-Final.xlsx

Other health agencies

Drug side effects: Now available on the OpenFDA website, https://open.fda.gov/, along with other public use files from the agency. Check out http://www.researchae.com/ to look at how this data can be visualized. Enigma.io also has a data visualization tool: http://enigma.io/labs/healthcare-data/.

Doctor discipline: The Health Resources and Services Administration has summary data by state and profession and year on doctor discipline, as well as malpractice payouts. http://www.npdb.hrsa.gov/resources/npdbstats/npdbStatistics.jsp.

Centers for Disease Control and Prevention: https://data.cdc.gov/ and wwwn.cdc.gov/communityhealth and http://wonder.cdc.gov/.

Agency for Healthcare Research and Quality maintains a free, on-line query system based on data from the Healthcare Cost and Utilization Project (HCUP). It provides access to health statistics and information on hospital inpatient and emergency department utilization. http://hcupnet.ahrq.gov/

Health Resources and Services Administration releases county-level health care on an array of topics, including the number health care professionals by specialty, hospital usage, air quality, and demographics. http://ahrf.hrsa.gov/index.htm