Supplementary Online Content

Bui AL, Dieleman JL, Hamavid H, et al. Spending on children's personal health care in the United States, 1996-2013. *JAMA Pediatr*. Published online December 27, 2016. doi: 10.1001/jamapediatrics.2016.4086.

This supplementary material has been provided by the authors to give readers additional information about their work.

Note: The content in this supplement was drawn from the supplementary appendix associated with Dieleman et al.[1] The content was, in a small number of places, adjusted to reflect only content essential to the methods used to complete the analysis for this paper.

eAppendix 1. Overview of methodology

The objective of this research was to comprehensively measure and describe spending on pediatric health care in the United States (US) using granular, politically and clinically useful categories. We produced annual estimates for 1996 through 2013. These estimates were created to be as comprehensive as possible, and they aggregate to reflect the official US government estimates of US health spending, as reported in the National Health Expenditure Accounts (NHEA).[2] These estimates were produced to reflect actual spending on health, also known as expenditure or payments, rather than charges made by medical providers. In many cases, charges are not paid in full and tracking these would be an overestimate of the resources actually spent on health care[3–5]. Spending estimates were adjusted for inflation using the economy-wide consumer price index from the International Monetary Fund, and were reported in 2015 dollars.[6] In addition to health spending, volume of health goods or services was also estimated – measured as the number of visits, bed-days, or prescriptions filled.

This research focused on personal health care spending. Personal health care spending is defined in the NHEA as "the total amount spent to treat individuals with specific medical conditions," and in 2013 was 84.8% of total US health spending. For this study, personal health care spending was disaggregated into 6 types of care, including inpatient care, ambulatory care, retail pharmaceuticals, emergency department care, and nursing facilities care, and dental care.

The overarching research strategy was to use microdata to inform spending and volume estimates at the most granular level possible. For the disaggregation of personal health care spending, microdata consisted of administrative records, insurance claims, or household surveys that report health spending by cause of illness or reason for the health care event, type of good or service, and demographic information. These sources provided data at the patient, encounter, or claim level. In most cases, spending and volume estimates were disaggregated into age-, sex-, cause -, type of care-, and year-specific categories. For the disaggregation of government public health activities, government budget documents and public agency justification documents were used.

To provide a comprehensive yet granular set of health spending estimates, health spending was split into categories defined by simultaneously applying three distinct frameworks. These three frameworks reflect demography, epidemiology, and the type of health care provided.

1. **Demography:** Health spending and volume of goods and services were estimated for **both sexes** and for **5 age groups**, <1, 1-4, 5-9, 10-14, and 15-19.
2. **Epidemiology:** Health spending and volume of goods and services were estimated for **158 causes**. The cause list for this project was based on the Global Burden of Disease (GBD) 2013 study.[7] GBD 2013 classified causes of health burden at five different levels of disaggregation. Level III classification was extracted from GBD 2013 for this study. This included 144 causes of health burden. In addition to these, 14 other categories were added. Four risk factors for other underlying health causes were added because it was clear that there is substantial spending on the treatment of these risk factors, and this spending is to prevent a wide set of causes of illness. These additional causes are spending on hypertension, hyperlipidemia, obesity, and tobacco cessation. In addition to these, seven causes were added that were not associated with health burden (and are therefore not considered by GBD) but were associated with health spending. Examples of these causes were well-person care, pregnancy and postpartum care, and well-dental care. Finally, this project also tracked spending on three impairments. These impairments – heart failure, septicemia, and renal failure – are not underlying causes of health burden, but rather consequences of other underlying causes. Spending on these causes was tracked because they represent large portions of health spending and are of political interest. A description and full list of causes and how they map to the International Classification of Diseases version 9 (ICD-9) are provided in section three of this appendix.
3. **Types of goods or services:** Health care spending and volume of goods and services were estimated for **six types of goods and services**: ambulatory care, inpatient care, emergency department care, nursing care, dental care, and prescribed retail pharmaceuticals. Definitions for these types of goods and services were designed to reflect the underlying microdata.

- *Ambulatory care*: Ambulatory care included preventive, curative, and rehabilitative medical and psychiatric services, procedures, and medications provided in ambulatory care settings including physician's offices, freestanding clinics and hospital outpatient departments. Emergency room visits and dental visits are excluded from ambulatory care. For ambulatory care, volume was measured as the number of visits.
- *Inpatient care*: Inpatient care included all spending in an inpatient hospital facility, whether preventive, curative, or rehabilitative, and included all medical goods, whether pharmaceuticals, diagnostics or devices, consumed by inpatients, regardless of their length of stay. Emergency room visits that result in an inpatient stay are considered inpatient care. For inpatient care, volume was measured as the number of days spent in an inpatient setting.
- *Emergency department care*: Emergency department care included preventative, curative and rehabilitative medical and psychiatric care provided at hospital-based and freestanding emergency departments. Emergency department care excluded visits that resulted in inpatient admission. For emergency department care, volume was measured as the number of visits.
- *Nursing facilities care*: Nursing care included nursing care provided in nursing homes or other residential institutions. Home-based care and palliative or hospice care provided in inpatient settings were excluded. Spending on hoteling costs, such as room and board are included. For nursing care, volume was measured as the number of days spent in a facility.
- *Dental care*: Dental care included preventative and curative health care at a dental facility. For dental care, volume was measured as the number of visits to a dental facility.
- *Prescribed retail pharmaceuticals:* Prescribed retail pharmaceuticals (pharma) included all prescription medicines purchased in a retail pharmacy setting. This category excluded any medications consumed in inpatient, ambulatory, long-term and emergency settings during a visit. It also excluded over-the-counter (non-prescribed) medications and therapeutic devices. For prescribed retail pharmaceutical, volume was measured as the number of prescriptions filled. The cause of illness is captured by the diagnoses reported by an individual who held the prescription, not by an Anatomical Therapeutic Chemical (ATC) classification system or medication code.

For all estimates, uncertainty was propagated using a bootstrapping method.

Statistical models were used when necessary to generate a complete set of estimates, combine data sources, and adjust the data for known biases. The population weighted estimates derived from the microestimates were compared and scaled to reflect the total health expenditure for each type of care and year. A brief summary of each step, including the types of care impacted, the effect of the process, and the motivating purpose of the process are described in table 1-1 below. This table does not attempt to explain how each step was conducted. Rather, this table explains briefly why each step was conducted and how it impacted the data.

Table 1.1 Summary of steps taken to get final estimates			
Step	**Types of care**	**Motivation**	**Effect**
Format data	Ambulatory, inpatient, emergency department, nursing care, dental, prescribed retail pharmaceuticals	To enable all data sources to go through same statistical machinery	All data were structured in the same manner, and variable names and variable formats were systematized across all data sources used
Bootstrap	Ambulatory, inpatient, emergency department, nursing care, dental, prescribed retail pharmaceuticals	To obtain 1,000 bootstrap samples upon which all other steps could be run independently, in order to quantify uncertainty	1,000 samples were created for analysis based on survey adjusted bootstrapping methods
De-truncation	Ambulatory (spending data only), emergency department (spending data only), prescribed retail pharmaceuticals	To estimate more detailed four- and five-digit ICD-9 diagnoses from the three-digit diagnoses recorded in Medical Expenditure Panel Survey (MEPS)	Variation within each bootstrap draw and across draws for data from MEPS was increased
Redistribution	Ambulatory, inpatient, emergency department, nursing care, prescribed retail pharmaceuticals	To attribute all spending and volumes to causes that represent the true underlying reason for a health care encounter	Spending and volume originally attributed to ICD-9 codes that do not map to GBD causes were assigned to GBD causes based on redistribution packages developed by the IHME GBD research. This redistributions was designed to take into account age and sex. While each cause is impacted differently by the redistribution process, spending per cause, measured at the age, sex, type, and year level goes up or stays the same, while spending attributed to “garbage codes” is removed

Mapping	Ambulatory, inpatient, emergency department, nursing care, dental, prescribed retail pharmaceuticals	To divide spending into 158 medically important and policy-relevant categories	Causes were aggregated from ICD-9 codes to 158 GBD causes, leading to more data for each cause-, year-, age-, sex-, type-combination
Injury adjustment	Ambulatory, inpatient, emergency department, nursing care, prescribed retail pharmaceuticals	To have all spending and volume due to injuries be defined by external cause of injury codes, rather than less actionable nature of injury codes	All spending attributed to injuries were defined by the external cause of injury
Comorbidity adjustment	Ambulatory, inpatient, emergency department, nursing care	To redistribute resources toward the underlying cause of the health care spending, rather than merely the primary diagnosis	Spending was moved from some causes to others, based on whether, on average, the cause leads to excess spending (as comorbidity) or is a primary diagnosis that has spending increased by excess spending on comorbidities
Age-splitting	Nursing care	To have Medicare nursing care claims data be consistent with all other data sources, as Medicare aggregates younger ages to ensure patient privacy	Charges captured in Medicare claims were split up from larger age bins into the age bins used in the study
Inpatient charges-to-payments adjustment	Inpatient	To estimate total inpatient spending from the inpatient facility charges report in the National Inpatient Sample	Inpatient spending estimates were made smaller than originally reported in National Inpatient Sample, based on cause, year, payer specific payment to charge ratios

Completing the series	Ambulatory, inpatient, emergency department, nursing care, dental, prescribed retail pharmaceuticals	To have estimates for years in which data does not exist, to obtain estimates for spending that are missed due to survey designs, and to have estimates that are appropriately consistent across age and time	Multiple data sources were combined to leverage strengths across data sources, such that every type-, age-, year-, cause-, and sex-combination was estimated and “smooth” series were produced
Nursing-care adjustment	Nursing care	To estimate nationally representative spending and volume estimates for short- and long-term stays at nursing homes	Three data sources were leveraged together, two using linear regression, to create nationally representative spending and volume estimates for short-term and long-term nursing facility care
Mental health adjustment	Ambulatory, inpatient	To address the under sampling of mental health and substance abuse specialty facilities and create mental health and substance abuse health care spending aggregates that are commensurate with official US government estimates.	Spending and volume on mental illnesses were increased, relative to non-mental illness causes, for the ambulatory and inpatient types of care
Scaling	Ambulatory, inpatient, emergency department, nursing care, dental, prescribed retail pharmaceuticals	To match spending estimates that reflect the official US government numbers, as no data source offers complete census of health care spending	Estimates for spending were increased or decreased depending on type of care

eAppendix 2. Data sources

This section of the appendix describes the data sources used for this study. Many of these methods are specific to an individual data source, as they examine the process of extracting data and making it comparable to other sources.

Tables 2.1 summarizes the primary data sources and years of data used for this study.

- The **National Health Expenditure Accounts (NHEA)** is a primary data source used to provide macro-estimates of annual health spending. Produced annually by the Office of the Actuary at the US Centers for Medicare and Medicaid Services (CMS), the NHEA constitutes official estimates of total healthcare spending in the US, dating back to 1960.[2] In addition to reporting total health spending, the NHEA reports health spending by type of good and services, source of funding, and type of sponsor. Data from the National Health Expenditures tables were used. This data "measures annual U.S. expenditure for health care goods and services, public health activities, government administration, the net cost of health insurance, and investment related to health care" This study focused on generating annual spending and volume estimates that could be scaled to reflect these type-specific spending totals. Scaling to NHEA totals was necessary because no single source of microdata fully captured the NHEA type-specific envelope, due to incomplete sampling frames and biases associated with small samples. This study assumes that the portion of NHEA directly accounted for in the microdata is proportional to the portion of NHEA not accounted for in the microdata, unless otherwise adjusted. These NHEA data were extracted from the CMS website.[8]
- The **Medical Expenditure Panel Survey (MEPS)** was a primary microdata source used to estimate the distribution of annual health spending across age, sex, and disease groups.[9] MEPS is produced by the US Agency for Healthcare Research and Quality (AHRQ), and provides data on the frequency of health services, health status and conditions, payments, and methods of payment for health services. MEPS draws from an annual survey sample of between 21,000 and 37,000 non-institutionalized civilians. Survey weights included in the data were used throughout this study to make MEPS estimates nationally representative. For each health system encounter, MEPS reports information on both payments and causes of health system encounter using the International Classification of Disease version 9 (ICD9). ICD9 codes were truncated by AHRQ to include only the first three digits of codes that are often four or five digits long. To address this, three digit codes were assigned four- or five-digit codes probabilistically for patient-level data and proportionally for aggregated data. Probabilities for this re-assignment were generated from data sources that include four- and five-digit codes. MEPS is already disaggregated into types of goods and services, which generally correspond closely to the types used in the NHEA. To make NHEA and MEPS data align more completely, emergency department (ED) visits that result in inpatient stays were removed from the MEPS ED data. These visits are identified by an indicator in the survey or by assessing if an inpatient stay and ED visit occurred on the same day and share at least one ICD9 diagnosis in common.
- **Substance Abuse and Mental Health Services Administrative (SAMHSA)** data provide estimates on health spending in mental health and substance abuse specialty clinics. Estimates for spending in these settings are often not included in other data sources, and it is important to account for this to accurately capture spending on certain causes. Data were extracted from the National Expenditure for Mental Health Services & Substance Abuse Treatment: 1986—2009, and from the Projections of National Expenditure for Mental Health Services and Substance Abuse Treatment: 2004—2014. These data were used to adjust the microdata when scaling to the NHEA totals.[10,11]
- The Truven Health **MarketScan©** Commercial Claims and Encounters Database provides claim-level health care information on more than 53 million commercially insured enrollees. These data were combined with the Truven Medicare Supplemental and Coordination of Benefits Database, which covers more than 4 million Medicare-eligible retirees with employer-sponsored supplemental plans in 2012. These data were used to create health system encounter profiles by age, sex, type, and cause. These profiles then served as Bayesian priors for volume and spending estimates.
- The **National Ambulatory Medical Care Survey (NAMCS)** and **National Hospital Ambulatory Medical Care Survey (NHAMCS)** are annual surveys conducted by the US Center for Disease Control and Prevention (CDC) to collect data on the utilization and provision of outpatient and ED services. These data are collected from physicians who primarily engage in direct patient care. Together, these two surveys cover 69 and 109 thousand patients per year. These data provide age, sex, type, and cause estimates. Causes

are reported using five-digit ICD9 codes. Because this data source does not include information on costs or spending, it was used only to inform volume estimates. Survey weights were used to make estimates nationally representative.

- The **National Inpatient Sample (NIS)** is also produced by AHRQ and is the largest publically available all-payer inpatient healthcare database with nationally representative US spending estimates. The NIS covers six to eight million inpatient hospital stays per year, and includes information on age, sex, cause, days spent hospitalized, and charges. Causes are reported using five-digit ICD9 codes.
- CMS provides data with information about Medicare beneficiaries, Medicaid eligibility, **Medicare and Medicaid claims**, Medicare providers, and clinical data. These data are stripped of personally identifying information. Data on beneficiaries and claims for health care at skilled-nursing facilities were obtained from this database. Data on payments and causes of illness, reported using five-digit ICD9 codes, were used only for beneficiaries 65 years and older. These data include between two and four million claims per year.
- The **National Nursing Home Survey (NNHS)** was used to supplement information from Medicare and Medicaid claims in skilled nursing facilities. While the Medicare and Medicaid claims only provide information on patients with public funding in skilled nursing facilities, the NNHS provides information on patients regardless of payer in both skilled and unskilled nursing facilities. NNHS is nationally representative and provides information on payments and causes, which are reported using five-digit ICD9 codes. Data were provided for between 20,000 and 36,000 current long-term care residents per year.
- The **Medicare Current Beneficiary Survey (MCBS)** was used to supplement information from Medicare and Medicaid claims in skilled nursing facilities and from the NNHS. The MCBS is a nationally representative sample of those on Medicare, including spending and volume in nursing homes. The MCBS includes not only nursing care spending covered by Medicare, but also supplemental insurance and out-of-pocket spending. MCBS was received in an aggregated form from the Bureau of Economic Analysis. These spending and volume estimates were stratified by age, year, sex, and cause in Clinical Classification Software codes.

Table 2.1: List of primary data sources

Type of care	Macro spending data and years	Micro spending data and years	Micro volume data and years
Ambulatory	NHEA (1996 – 2013)*	MEPS (1996 – 2013); SAMHSA (1998, 2002, 2004, 2005, 2009); MarketScan (2000, 2010, 2012)	NAMCS (1996 – 2011); NHAMCS (1996 – 2011); MarketScan (2000, 2010, 2012)
Inpatient	NHEA (1996 – 2013)	NIS (1996 – 2012); MEPS (1996 – 2013), SAMHSA (1998, 2002, 2004, 2005, 2009); MarketScan (2000, 2010, 2012)	NIS (1996 – 2012); MarketScan (2000, 2010, 2012)
Emergency Department	NHEA (1996 – 2013)*	MEPS (1996 – 2013); MarketScan (2000, 2010, 2012)	NHAMCS (1996 – 2011); MarketScan (2000, 2010, 2012)
Nursing care	NHEA (1996 – 2013)	Medicare claims data (1999 – 2001, 2002, 2004, 2006, 2008, 2010, 2012); NNHS (1997, 1999, 2004); MCBS (1999-2011); MarketScan (2000, 2010, 2012), MCBS	Medicare claims data (1999 – 2001, 2002, 2004, 2006, 2008, 2010, 2012); NNHS (1997, 1999, 2004); MCBS (1999-2011); MarketScan

		(1999 – 2011)	(2000, 2010, 2012)
Dental	NHEA (1996 – 2013)	MEPS (1996 – 2013)	MEPS (1996 – 2013)
Prescribed retail pharmaceuticals	NHEA (1996 – 2013)	MEPS (1996 – 2013)	MEPS (1996 – 2013)
Other	NHEA (1996 – 2013)	Not disaggregated	Not disaggregated

eAppendix 3. Cause maps

Cause	Cause Name	ICD Codes	Sexes Allowed	Ages Allowed
A.1.1	Tuberculosis	010-019.9, 137-137.9, 138.0-138.9, 139.9, 320.4, 730.4-730.6, V01.1, V03.2, V12.01, V71.2, V74.1	both	0-19
A.1.2	HIV/AIDS	042-044.9, 112.4-118.9, 136.3-136.5, 279.2-279.3, 279.8-279.9, V08	both	0-19
A.2.1	Diarrheal diseases	001-001.9, 003-006.9, 007.4-007.8, 008.01-008.02, 008.04, 008.2-009.9, 787.91, V01.0, V01.83, V02.0, V02.2-V02.3, V03.0, V74.0	both	0-19
A.2.2	Intestinal infectious	002.0-002.9, 007-007.3, 007.9-008.00, 008.03, 008.09-008.1, V02.1, V03.1	both	0-19
A.2.3	Lower respiratory infections	466-469, 470.0, 480-482.89, 483.0-483.9, 484.1-484.2, 484.6-484.7, 487-489, V01.82, V03.81-V03.82, V04.7, V04.81-V04.82, V12.61	both	0-19
A.2.4	Upper respiratory infections	460-465.9, 475-475.9, 476.9	both	0-19
A.2.5	Otitis media	381-384.9	both	0-19
A.2.6	Meningitis	036-036.40, 036.5, 036.8-036.9, 047-049.9, 320.0-320.3, 320.5-320.89, 321-321.4, 321.6-322.9, V01.84	both	0-19
A.2.7	Encephalitis	062-064.9, 139.0, 323-323.9, V05.0-V05.1, V12.42	both	0-19
A.2.8	Diphtheria	032-032.9, V02.4, V03.5, V74.3	both	0-19
A.2.9	Whooping cough	033-033.9, 484.3, V03.6	both	0-19
A.2.10	Tetanus	037-037.9, 771.3, V03.7	both	0-19
A.2.11	Measles	055-055.9, 484.0, V04.2, V73.2	both	0-19
A.2.12	Varicella	052-053.9, V01.71, V05.4	both	0-19
A.3	NTDs & malaria	060-061.8, 065-066.9, 071-071.9, 076-076.1, 076.6, 076.9, 080, 080.2-084.9, 085.0-085.5, 086-088.9, 120-130.9, 139.1, V01.5, V04.4-V04.5, V05.2, V12.03, V73.4-V73.6, V75.1-V75.3, V75.5-V75.8	both	0-19
A.4.1	Maternal hemorrhage	640-641.93, 665-665.34, 666-666.9	female	10-45
A.4.2	Maternal sepsis	659.3-659.33, 670-670.9	female	10-45
A.4.3	Maternal hypertension	642-642.94	female	10-45
A.4.4	Maternal obstructed labor	660-660.93	female	10-45

A.4.5	Maternal abortive	630-636.92, 638-638.92, 646.3-646.33	female	10-45
A.4.6	Maternal indirect	646-646.24, 646.4-649, 649.00-649.9, 674-674.94	female	10-45
A.4.9	Other maternal disorders	643-644.00, 644.1-644.20, 645-645.10, 645.13-645.23, 652.0-652.20, 652.23-652.50, 652.53-652.60, 652.63-652.80, 652.83-653.40, 653.43-654.20, 654.23-655.70, 655.73-655.80, 655.83-656.10, 656.13-656.40, 656.43-656.50, 656.53-656.60, 656.63-656.80, 656.83-657.00, 657.03-658.00, 658.03-658.10, 658.13-658.20, 658.23-658.40, 658.43-659.10, 659.13-659.23, 659.4-659.40, 659.43-659.50, 659.53-659.60, 659.63-659.70, 659.73-659.80, 659.83-659.93, 661-661.00, 661.03-661.20, 661.23-661.30, 661.33-663.10, 663.13-663.20, 663.23-663.30, 663.33-663.80, 663.83-664.00, 664.04-664.80, 664.84-664.94, 665.4-665.94, 667-669.61, 669.70, 669.8-669.80, 669.82-669.94, 671-673.9, 675-679.14, 768.0-768.1, V13.1, V15.21-V15.22	female	10-50
A.5.1	Neonatal preterm birth	761.0-761.1, 765-765.9, 769-769.9, 770.2-770.9, 776.6, 777.5-777.6	both	0
A.5.2	Neonatal encephalopathy	761.7-763.9, 767-768, 768.2-768.9, 770.1-770.18, 772.1-772.9, 779.0-779.2	both	0
A.5.3	Neonatal sepsis	771.4-771.9	both	0
A.5.4	Neonatal hemolytic	773-774.9	both	0
A.5.5	Other neonatal	760-760.70, 760.72-761, 761.2-761.6, 764-764.99, 766-766.9, 770, 771, 772-772.0, 775, 775.4-776.5, 776.7-777.4, 777.7-779, 779.3-779.34, 779.6-779.89	both	0
A.6.1	Protein-energy malnutrition	260-263.9	both	0-19
A.6.2	Iodine deficiency	244.2	both	1-19
A.6.3	Vitamin A deficiency	264-264.9	both	1-19
A.6.4	Iron-deficiency anemia	280-281, 285-285.9, V18.2, V78.0-V78.1	both	0-19
A.6.5	Other nutritional	265-269.9, 281.0-281.9, 716.0-716.09	both	0-19
A.7.1	STDs	054.1, 090-099.9, 131-131.9, 614-614.9, V01.6, V02.7-V02.9, V73, V73.8, V73.88, V73.9-V73.98, V74.5-V74.6	both	0-19
A.7.2	Hepatitis	070-070.21, 070.3-070.31, 070.4-070.43, 070.49-070.53, 070.59-070.9, V02.6-V02.69, V05.3	both	0-19
A.7.3	Leprosy	030-030.9, V74.2	both	1-19

A.7.4	Other infectious	020-029, 031-031.9, 034-034.9, 039-039.4, 039.8-040, 040.1-041.89, 045-046.9, 050-051.9, 054-054.0, 054.10-054.9, 056-059.9, 072-075.9, 076.5, 076.8, 078.5-079.99, 080.0, 100-104.9, 112-112.0, 112.3, 136-136.29, 138, 139, 321.5, 357.0, 390-390.9, 391.4, 392, 392.9, 484.4-484.5, 730.7-730.99, 771.0-771.2, V01, V01.2-V01.4, V01.7, V01.79-V01.81, V01.89-V02, V02.5-V02.59, V03, V03.3-V03.4, V03.8, V03.9-V04.1, V04.3, V04.6, V04.8, V04.89-V05, V05.8-V06.8, V09-V09.91, V12.0-V12.00, V12.02, V12.04-V12.09, V18.8, V71.82-V71.83, V73.0-V73.1, V73.3, V73.81, V73.89, V73.99, V74.8-V74.9, V75.4, V75.9	both	0-19
A.7.5	Septicemia	038-038.9, 995.91-995.92	both	0-19
B.1.1	Esophageal cancer	150-150.9, 211.0, 230.1, V10.03	both	15-19
B.1.2	Stomach cancer	151-151.9, 209.23, 209.63, 211.1, 230.2, V10.04, V55.1	both	15-19
B.1.3	Liver cancer	155-155.3, 211.5, V10.07	both	5-19
B.1.4	Larynx cancer	161-161.9, 162.1, 212.1, 231.0, 235.6, V10.21	both	15-19
B.1.5	Lung cancer	162-162.0, 162.2-162.9, 163.5, 209.21, 209.61, 212.2-212.3, 231.1-231.2, 235.7, V10.1-V10.20, V16.1-V16.2, V76.0	both	15-19
B.1.6	Breast cancer	174-175.9, 217-217.8, 233.0, 238.3, 239.3, 610-610.9, V10.3, V16.3, V50.41, V51.0, V52.4, V76.1-V76.19	both	15-19
B.1.7	Cervical cancer	180-180.9, 219.0-219.1, 233.1, 622-622.2, V10.41, V13.22, V67.01, V72.32, V76.2, V88.0-V88.03	female	15-19
B.1.8	Uterine cancer	182-182.8, 218-218.9, 233.2, 621.0-621.35, V10.42	female	15-19
B.1.9	Prostate cancer	185-185.9, 222.2, 236.5, V10.46, V16.42, V76.44	male	15-19
B.1.10	Colorectal cancer	153-154.9, 155.5-155.9, 209.1-209.17, 209.5-209.57, 211.3-211.4, 230.3-230.6, 569.0, 569.43-569.44, 569.84-569.85, V10.05-V10.06, V55.3, V76.41, V76.5-V76.52	both	15-19
B.1.11	Mouth cancer	140-145.9, 210.0-210.6, 235.0, V10.01-V10.02, V76.42	both	15-19
B.1.12	Nasopharynx cancer	147-147.9, 210.7, 210.9	both	5-19
B.1.13	Other pharynx cancer	146-146.9, 148-148.9, 210.8	both	15-19
B.1.14	Gallbladder cancer	156-156.9, 209.25-209.27, 209.65-209.67	both	15-19

B.1.15	Pancreatic cancer	157-157.9, 211.6-211.7, V88.1-V88.12	both	15-19
B.1.16	Melanoma	172-172.9	both	15-19
B.1.17	Skin cancer	173-173.99, 209.31-209.36, 214-214.1, 215-216.9, 222.4, 232-232.9, 238.2, V76.43	both	15-19
B.1.18	Ovarian cancer	183-183.0, 236.2, V10.43, V16.41, V50.42, V76.46	female	15-19
B.1.19	Testicular cancer	186-186.9, 222.0, 222.3, 236.4, V10.47-V10.48, V16.43, V76.45	male	15-19
B.1.20	Kidney cancer	189.0-189.1, 209.24, 209.64, 223.0-223.1, 236.91, V10.52-V10.59, V16.51	both	1-19
B.1.21	Bladder cancer	188-188.9, 223.3, 233.7, 236.7, 239.4, V10.51, V16.52, V43.5, V55.5-V55.6, V76.3	both	15-19
B.1.22	Brain cancer	191-192.9, 225-225.9, 237-237.9, 239.6, V10.85-V10.86, V12.41	both	1-19
B.1.23	Thyroid cancer	193-193.9, 226-226.9, V10.87	both	10-19
B.1.24	Mesothelioma	163-163.3, 163.8-163.9	both	15-19
B.1.25	Hodgkin disease	201-201.98, V10.72	both	0-19
B.1.26	Lymphoma	200-200.9, 202-202.98, V10.7-V10.71, V10.79, V16.7	both	1-19
B.1.27	Myeloma	203-203.9	both	15-19
B.1.28	Leukemia	204-208.92, V10.6-V10.69, V16.6	both	1-19
B.1.29	Other neoplasms	152-152.9, 158-158.9, 160-160.9, 164-164.9, 170-171.9, 181-181.9, 182.9, 183.2-183.8, 184.0-184.4, 184.8, 187.1-187.8, 189.2-189.8, 190-190.9, 194-194.8, 209.0-209.03, 209.22, 209.4-209.43, 211.2, 211.8, 212.0, 212.4-212.8, 213-213.9, 214.2-214.9, 221.0-221.8, 222.1, 222.8, 223.2, 223.8-223.89, 224-224.9, 227-228.9, 229.0, 229.8, 230.7-230.8, 233.31-233.32, 233.4-233.5, 234.0-234.8, 235.4, 235.8, 236.1, 236.99, 238.0-238.1, 238.4-238.8, 239.2, 623.0-623.1, 623.7, V10.22-V10.29, V10.4-V10.40, V10.44-V10.45, V10.49-V10.50, V10.8-V10.84, V10.88-V10.89, V55.2, V58.0, V58.11, V67.1-V67.2, V76.4, V76.47-V76.49	both	0-19
B.2.1	Rheumatic heart disease	391-391.2, 391.8-391.9, 392.0, 393-398.99	both	1-19
B.2.2	Ischemic heart disease	410-414.9, V17.3, V81.0	both	1-19
B.2.3	Cerebrovascular disease	430-435.9, 437.0-437.2, 437.5-437.8, V12.54, V17.1	both	1-19
B.2.4	Hypertensive heart disease	402-402.91	both	1-19

B.2.5	Cardiomyopathy	036.43, 036.6, 422-422.99, 425-425.9, 429.0-429.1	both	1-19
B.2.7	Aortic aneurysm	441-441.9	both	15-19
B.2.9	Endocarditis	036.42, 421-421.9, 424.9-424.91	both	0-19
B.2.10	Other cardiovascular	036.41, 417-417.9, 420-420.99, 423, 423.1-424.8, 424.99, 427-427.2, 427.6-427.89, 442-443, 444-445.89, 447-454.9, 456, 456.3-457.9, 459, 459.1-459.39	both	0-19
B.2.11	Heart Failure	428-428.9	both	0-19
B.3.1	COPD	490-492.9, 494-494.9, 496-499	both	1-19
B.3.2	Pneumoconiosis	500-504.9, V15.84	both	1-19
B.3.3	Asthma	493-493.92, V17.5	both	1-19
B.3.4	Interstitial lung disease	135-135.9, 136.6, 515, 516-516.9	both	1-19
B.3.5	Other chronic respiratory	327.2-327.8, 470, 470.9-474.9, 476-476.1, 477-479, 495-495.9, 506-506.9, 508-509, 517-517.8, 518.6, 518.9, 519.1-519.8, 713.4, 780.57, 786.03, V07.1, V13.81, V14-V15.09, V19.6	both	1-19
B.4	Cirrhosis	070.22-070.23, 070.32-070.33, 070.44, 070.54, 456.0-456.21, 571-571.9, 572.3-572.9, 573.0-573.3, 573.8-573.9, V42.7	both	0-19
B.5.1	Peptic ulcer disease	531-534.91, V12.71	both	1-19
B.5.2	Gastritis & duodenitis	535-535.9	both	1-19
B.5.3	Appendicitis	540-542.9	both	1-19
B.5.4	Ileus & obstruction	560-560.39, 560.8-560.9	both	0-19
B.5.5	Hernia	550-551.1, 551.3-552.1, 552.3-553.03, 553.6, 555.3	both	1-19
B.5.6	Inflammatory bowel	555-555.2, 555.9-556.9, 558-558.9, 569.5, V12.72	both	1-19
B.5.7	Vascular intestinal	557-557.9	both	1-19
B.5.8	Gallbladder & biliary	574-576.9	both	1-19
B.5.9	Pancreatitis	577-577.9, 579.4	both	1-19
B.5.10	Other digestive	455-455.9, 530-530.9, 536-536.1, 537-537.6, 537.8-537.84, 538, 543-543.9, 553.1-553.3, 562-562.13, 564-564.1, 564.5-564.7, 565-566.9, 569.1-569.42, 569.7-569.71, 573.4, 579-579.2, 579.8-579.9, 713.1, 787.1	both	1-19
B.6.3	Epilepsy	345-345.91	both	0-19
B.6.4	Multiple sclerosis	340-340.9	both	5-19
B.6.5	Migraine	346-346.93	both	5-19
B.6.6	Tension headache	307.81, 339-339.12, 339.20-339.89	both	5-19

B.6.8	Other neurological	330-330.9, 331.5-331.9, 333-338.4, 341-341.9, 349, 349.2-349.8, 350-353.0, 353.5-355, 355.1-356.9, 357.1, 357.3-357.4, 357.7, 358-359.9, 713.5, 725-725.9, 728-728.11, 728.13-728.81, 728.83-729.5, 729.7-729.90, 729.92-729.99, 775.2	both	0-19
B.7.1	Schizophrenia	295-295.95, 301.0, 301.2-301.22, V11.0	both	10-19
B.7.2	Alcohol use disorders	291-291.9, 303-303.93, 305.0-305.03, 357.5, 760.71, 790.3, E86.0-E86.019, V11.3, V79.1	both	0-19
B.7.3	Drug use disorders	292-292.9, 304.0-304.83, 305, 305.2-305.93, E85.0-E85.029, E85.09-E85.439, V15.85-V15.86	both	10-19
B.7.4	Depressive disorders	296.2-296.36, 300.4, 311-311.9, V11.1-V11.2, V79.0	both	1-19
B.7.5	Bipolar disorder	296-296.16, 296.4-296.99, 301.1-301.13	both	10-19
B.7.6	Anxiety disorders	300.0-300.09, 300.2-300.3, 301.4, 308-309.9, 313.0	both	1-19
B.7.7	Eating disorders	307.1, 307.51, 307.54	both	5-19
B.7.8	Autistic spectrum	299.0-299.01, 299.8-299.81	both	0-19
B.7.9	ADHD	314.0-314.01	both	1-19
B.7.10	Conduct disorder	301, 301.3, 301.5-301.89, 312-312.9, V71.02	both	5-19
B.7.11	Intellectual disability	317-319.9, V18.4	both	0-19
B.7.12	Other mental & substance	298-298.4, 299, 299.1-299.11, 299.9-300, 300.1-300.15, 300.5-300.89, 302-302.9, 306-306.9, 307.0, 307.2-307.49, 307.6-307.7, 313, 313.1-313.83, 314, 314.1-314.2, 315-315.5, 327-327.09, 347-347.9, 780.5-780.52, 780.59, V71.01	both	1-19
B.8.1	Diabetes	250-250.39, 250.5-250.99, 357.2, 362.0-362.07, 366.41, 775.0-775.1, 790.2-790.22, V12.21, V18.0, V45.85, V53.91, V58.67, V77.1	both	0-19
B.8.2	Acute glomerulonephritis	580-580.9	both	0-19
B.8.3	Chronic kidney disease	250.4-250.49, 403-404.93, 581-583.9, 585-585.9, 589-589.9, V13.03-V13.09, V18.6, V18.69, V42.0, V45.1-V45.12, V45.73, V56-V56.8, V81.5-V81.6	both	0-19
B.8.4	Urinary diseases	588-588.9, 590-590.9, 592-593.89, 594-596.81, 596.89-598.1, 598.8-599.6, 599.8-599.89, 600-608.89, 788.0, 788.3-788.39, V13.0-V13.02, V26.5, V26.52, V45.74, V47.4, V58.76	both	0-19

B.8.5	Gynecological diseases	112.1-112.2, 220-220.9, 256.4, 611-612.1, 615-618.9, 620-620.9, 621.4-621.9, 622.3-622.7, 623, 623.2-623.6, 623.8-624.9, 625.4, 627-629.81, V07.4-V07.59, V13.2, V13.29, V18.7, V26-V26.32, V26.34-V26.39, V26.42-V26.49, V26.51, V26.8-V26.9, V43.82, V45.71, V45.83, V47.5, V49.81, V59.70-V59.74, V72.3-V72.31, V84.04	female	10-19
B.8.6	Hemoglobinopathies	282-284.9, 713.2, V12.3, V18.3, V78, V78.2-V78.9, V83.0-V83.02	both	0-19
B.8.7	Endo/metab/blood/immune	240-243.9, 245-246.9, 251-251.2, 251.4-253.6, 253.8-256.39, 256.8-259.9, 270-271.9, 273-273.9, 275-276, 277-277.2, 277.30-277.9, 278.2-279.19, 279.4-279.49, 279.6, 286-286.5, 286.7-289.9, 713.0, 775.3, V12.2, V12.29, V12.4-V12.40, V18.1-V18.19, V29.3, V77-V77.0, V77.3-V77.4, V77.6-V77.7, V77.9, V77.99, V83.81, V84.81	both	0-19
B.8.8	Renal failure	584-584.9, 586-586.9	both	0-19
B.9.1	Rheumatoid arthritis	714-714.33, 714.8-714.9	both	5-19
B.9.3	Low back & neck pain	353.1-353.4, 355.0, 720-721.1, 721.3, 721.5-721.6, 721.8-724.9, 737-737.9	both	5-19
B.9.4	Gout	274-274.9	both	15-19
B.9.5	Other musculoskeletal	416.1-416.2, 446-446.9, 695.4-695.59, 710-712.99, 716.2-716.39, 719.2-719.39, 719.8-719.89, 721.2, 721.4-721.42, 726-727.9, 730-730.39, 732-734.2, 739-739.9, V82.81	both	0-19
B.10.1	Congenital anomalies	740-758.9, 759.0-759.89, V13.6-V13.69, V18.61, V18.9, V19.5, V19.7-V19.8, V55.7, V82.3	both	0-19
B.10.2	Skin diseases	035-035.9, 078-078.4, 110-111.9, 132-134.9, 680-695.3, 695.8-709.3, 709.8-709.9, 713.3, V13.3, V19.4, V43.83, V58.77, V82.0	both	0-19
B.10.3	Sense organ diseases	077-077.99, 360-360.44, 360.8-362, 362.1-366.19, 366.3-366.4, 366.42-374.85, 374.87-376.52, 376.8-380.9, 385-385.82, 385.89-389.9, V19.0-V19.3, V41-V41.5, V42.5, V43.0-V43.1, V45.6-V45.69, V45.78, V48.4-V48.5, V52.2, V53.1-V53.2, V58.71, V59.5, V72.0-V72.19, V74.4, V80, V80.1-V80.3	both	0-19
B.10.4	Oral disorders	520-529.9, V07.31, V45.84, V49.82, V52.3, V53.4, V58.5, V72.2	both	1-19

B.10.5	SIDS	798	both	0-1
C.1.1	Road injuries	E80.03, E80.13, E80.23, E80.33, E80.43, E80.53, E80.63, E80.73, E81.00-E81.06, E81.10-E81.17, E81.20-E81.27, E81.30-E81.37, E81.40-E81.47, E81.50-E81.57, E81.60-E81.67, E81.70-E81.77, E81.80-E81.87, E81.90-E81.97, E82.00-E82.06, E82.10-E82.16, E82.20-E82.27, E82.30-E82.37, E82.40-E82.47, E82.50-E82.57, E82.60-E82.61, E82.63-E82.64, E82.70, E82.73-E82.74, E82.80, E82.84, E82.90-E82.94	both	0-19
C.1.2	Other transport injuries	E80.0-E80.02, E80.1-E80.12, E80.2-E80.22, E80.3-E80.32, E80.4-E80.42, E80.5-E80.52, E80.6-E80.62, E80.7-E80.72, E81.07, E82.07, E82.17, E82.62, E82.72, E82.82, E83.1-E83.19, E83.3-E83.89, E84.0-E84.8, E92.91	both	0-19
C.2.1	Falls	E88.0-E88.699, E88.8-E88.89, E92.93, V15.88	both	0-19
C.2.2	Drowning	E83.0-E83.09, E83.2-E83.29, E91.0-E91.099	both	0-19
C.2.3	Fire & heat	E89.0-E89.909, E92.4-E92.499, E92.94	both	0-19
C.2.4	Poisonings	E85.03-E85.089, E85.48-E85.899, E86.02-E86.939, E86.940-E86.999, E92.92, V15.6, V87.0-V87.39	both	0-19
C.2.5	Mechanical forces	E91.3-E91.319, E91.6-E92.299, E92.81-E92.87	both	0-19
C.2.7	Animal contact	E90.5-E90.699, V90.31	both	0-19
C.2.8	Foreign body	360.5-360.69, 374.86, 376.6, 385.83, 709.4, 728.82, 729.6, E91.1-E91.209, E91.38-E91.509, V15.53, V90-V90.3, V90.32-V90.9	both	0-19
C.2.9	Other unintentional	E00.0-E03.0, E90.01-E90.019, E90.11-E90.119, E90.2-E90.4, E90.41-E90.499, E91.32-E91.339, E92.3-E92.399, E92.5-E92.809, E92.88-E92.889	both	0-19
C.3.1	Self-harm	E95.0-E95.9	both	5-19
C.3.2	Interpersonal violence	E90.40-E90.409, E96.0-E96.9, V15.41, V71.5, V71.81	both	0-19
C.4.1	Forces of nature	E90.0-E90.009, E90.09-E90.109, E90.18-E90.199, E90.7-E90.99	both	0-19
C.4.2	War & legal intervention	E97.0-E97.99, E99.0-E99.91	both	0-19
D.1	Well person	V20.1-V21.9, V30-V39.2, V70-V70.0, V70.3-V70.6, V70.8-V70.9, V72, V72.5-V72.8, V72.83-V72.9, V82, V82.5-V82.8, V82.89-V83, V83.8, V83.89-V84, V84.01-V84.03, V84.8, V84.89, V86-V86.1	both	0-19

D.2	Well pregnancy	644.03, 644.21, 645.11, 650-652, 652.21, 652.51, 652.61, 652.81, 653.41, 654.21, 655.71, 655.81, 656.11, 656.41, 656.51, 656.61, 656.81, 657.01, 658.01, 658.11, 658.21, 658.41, 659.11, 659.41, 659.51, 659.61, 659.71, 659.81, 661.01, 661.21, 661.31, 663.11, 663.21, 663.31, 663.81, 664.01, 664.81, 669.7, 669.71, 669.81, V13.21, V20-V20.0, V22-V24.2, V27-V28.9, V72.4-V72.42, V82.4, V91-V91.99	female	15-19
D.3	Well newborn	Same ICD codes as cause D.1, well person, but restricted to neonates in inpatient care.	both	0
D.4	Family planning	V15.7, V25-V25.9, V26.33, V26.41, V45.5-V45.59	both	0-19
D.5	Donor	V59-V59.4, V59.6-V59.7, V59.8-V59.9	both	0-19
D.6	Counselling services	V26.4, V61.1, V61.11-V62.9, V65-V65.9, V69-V69.9	both	0-19
D.9	Social services	V60-V61.09, V61.10	both	0-19
E.1.1	Tobacco	305.1-305.13, 649.0, 989.84, E86.94, V15.82	both	0-19
E.2.1	Obesity	278.0-278.1, V45.86, V77.8, V77.91, V85, V85.2-V85.54	both	0-19
E.2.2	Hypertension	401-401.9, 405-405.99, V81.1	both	0-19
E.2.3	Hyperlipidemia	272-272.9	both	0-19

eAppendix 4. Adjustments

Adjusting for Comorbidities.

We based the comorbidity adjustment on the National Inpatient Sample survey (NIS), primarily because this dataset is large and contains information on multiple secondary diagnoses in addition to the primary diagnosis. On average, 5.2 secondary diagnoses appear with each primary diagnosis in the NIS. These data were analyzed at the encounter level, where each observation in the data corresponds to a single hospital stay.

The input data used in the comorbidity analysis were mapped from ICD-9 codes to GBD causes, but still contains Not-elsewhere-classified (NEC) codes, N-codes for injuries and garbage codes. NEC codes are ICD-9 codes with a level-two or level-one mapping but no specific level-three mapping. Garbage codes are ICD-9 codes that represent ill-defined or non-underlying causes.

The data also included demographic information associated with each encounter: namely the sex and age of the patient, with ages binned into 5-year groups.

Select diagnoses were reassigned to alternative ones that were considered more informative, cause-restrictions were applied, data were divided into four age categories, causes with very few observations were dropped from the analysis, and bootstrap draws were merged on.

A probabilistic replacement was used to reassign certain injury causes (N-codes) and Not-Elsewhere-Classified (NEC) causes to alternative related diagnoses that were more relevant for this analysis. Probability maps were created for the injury adjustment by using data from years that provided both N-codes and E-codes to calculate the proportions of multiple N-codes to each E-code. These data were combined across all years to make probability maps specific to data source and age group. The maps were created at the source-age level, because disease burden and the distribution of causes are a function of age. Thus, the maps capture the variability in disease patterns across ages.

If multiple E-codes were listed for a given encounter, the first one was used to create the map. If multiple N-codes were listed for a given encounter, the most severe injury N-code was used to create the map, based on a severity hierarchy developed in GBD 2013. This means that if an encounter presented with multiple injuries coded as N-codes, the diagnosis that was likely to be responsible for the largest cost and burden was the one selected.[18]

For the NEC-adjustment, spending for each NEC-cause was probabilistically reassigned to a non-NEC cause in the same family. For instance, NEC cardiovascular disease might be reassigned to ischemic heart disease, cerebrovascular disease, or a number of other cardiovascular sibling-causes. The probability of being reassigned to a given sibling cause was based on the relative proportions of spending for each sibling cause in the data. For instance, IHD comprised 69% of the non-NEC cardiovascular spending, whereas heart failure comprised only 8%. Spending for NEC cardiovascular disease would therefore have a 69% probability of being reassigned to IHD, or an 8% chance of being reassigned to heart failure.

After removing N-codes and NEC causes from the data set, GBD restrictions were applied in the same manner as described in section three. All observations with a garbage code as the primary diagnosis were dropped from the dataset. If a primary diagnosis was not a garbage code, but a secondary diagnosis was a garbage code, that secondary diagnosis was replaced as missing. If a single observation had multiple diagnoses with ICD-9 codes that mapped to the same GBD cause (for example, two or more secondary diagnoses of “septicemia”), the duplicates were replaced as missing in the diagnosis list. All missing secondary diagnoses were removed from the data.

Encounters were divided into four age categories and all analysis was done at the source-age category-level. The four age categories were: (i) 0-14 years, (ii) 15-44 years, (iii) 45-64 years, and (iv) 65 years and above. These age groupings were chosen to reflect the observed age-delineations in patterns of disease burden and in the distribution of comorbidities. Because burden and comorbidity distributions differ across these age categories, four age category-specific lists of primary diagnoses and comorbidities were used in the analysis. Although the analysis was only conducted at the age category level, the sex and year variables were retained to inform the regression.

Even after pooling the data across all years and both sexes, there were still several causes that appeared as a primary diagnosis on only a relatively small number of encounters. These causes, such as leprosy, were conditions with low prevalence in the US. Because these conditions accounted for a negligible share of the total sample size, a lower bound on the reported number of encounters necessary for inclusion of a cause in analysis was set. Causes with fewer than 1,000 reported encounters across all years and both sexes within an age category were excluded from analysis.

One thousand draw frequencies were merged on to the cleaned input data by source, year, age, sex, and primary diagnosis. In order to integrate the comorbidity analysis with the rest of the disease expenditure analysis, the same bootstrap frequencies were used as in the rest of the study. All subsequent steps in comorbidity analysis were carried out 1,000 times; separately for each draw. This bootstrapping method was used to generate the uncertainty interval around point estimates. All reported comorbidity results are the mean estimates across the 1,000 bootstrap sample draws.

Comorbidity selection

To maintain the comprehensive nature of the analysis, nearly all conditions present in the data as primary diagnoses and as comorbidities were included. However, the list of comorbidities allowed for a given primary diagnosis was restricted because of the aims of the research and data availability.

Infrequently occurring comorbidities can present as merely noise in the dataset. For this framework, the comorbidities for each primary cause were defined by their probability of occurrence. For a given primary diagnosis, any secondary diagnosis with a probability of occurring greater than or equal to a lower bound threshold of 10% was considered as a viable comorbidity threshold. This threshold ensured that only the most pertinent and robust primary diagnosis-comorbidity pairs were considered in analysis.

After setting the comorbidity threshold, several secondary diagnoses still remained that were not viable comorbidities. These secondary diagnoses were manifestations of underlying causes rather than true comorbidities. To account for these false comorbidities, the following were excluded as comorbidities:

- All intermediate causes, such as skin and subcutaneous disease as a comorbidity for diabetes and heart failure as a comorbidity for CVD
- All residual "other" categories, such as other indirect maternal causes and other infectious diseases
- All risk factors, impairments, and well care causes, such as hyperlipidemia, renal failure, and well pregnancies

These restrictions were set in consultation with medical professionals who have an understanding of ICD-9 coding in clinical settings.

Modeling risk of excess spending

A log-linear regression model was used to generate estimates of the risk of excess spending due to comorbidities. Log-linear regression is one of the most commonly used methods for modeling health care spending data. A log-linear regression was estimated separately for each primary condition and age category, with the expenditure for a health system encounter as the dependent variable and all of the relevant comorbidities as binary independent variables indicating whether a patient was co-diagnosed with these comorbidities. The simplest form of the model is illustrated by Equation (1):

$$\log(expenditure_i) = \beta_{i0} + \sum_{j=1}^{J} \beta_{ij} comorbidity_{ij} + \varepsilon_i \qquad (1)$$

In this equation, excess spending was estimated independently for each primary diagnosis *i*, using age category-specific encounter-level data, and included the set of comorbidities that spanned from *j* to *J*. Binary indicators were included to control for the effects of heterogeneity between sexes and in spending across time. The relative risk of excess spending for *i* induced by comorbidity *j* was given by the coefficient on the respective primary diagnosis-comorbidity pair (β_{ij}). Only statistically significant pairs ($p < 0.05$) were included in the final comorbidity adjustment.

The presence of a comorbidity generally led to increased health spending for a given primary diagnosis. In these cases, $\beta_{ij} > 0$ and, on average, the comorbid condition raised the cost of managing the primary condition. However, a relative risk less than zero was a possible regression outcome. This result implied that the costs of managing the primary condition were lowered due to the coexistence of a given comorbid condition. While empirically rare, this would occur when a comorbid condition rendered standard treatment for the primary condition ineffective, unsafe, or poorly tolerated, necessitating less aggressive, intensive, or complex, and therefore less expensive treatment.

Calculating attributable fractions

The relative risk of excess spending due to comorbidities was then used to calculate the attributable fraction for each primary diagnosis-comorbidity pair. Attributable fractions are the proportions of disease expenditure attributable from the primary diagnosis to each comorbidity. The share of total expenditure for primary condition *i* attributable to comorbidity *j* is the product of the pair-specific relative risk of excess expenditure and the conditional probability of *i* and *j* co-occurring. This is illustrated by Equation (2):

$$AF_{ij} = p_{ij}\left(e^{\beta_{ij}} - 1\right) \qquad (2)$$

Generating flows and adjustment scalars

The attributable fractions for all primary diagnosis-comorbidity pairs were then used to reallocate spending from primary diagnoses to comorbidities. The comorbidity adjustment was applied to spending data that had been mapped from ICD-9 codes to GBD causes and had gone through redistribution and post-redistribution cleaning. However, the data had not yet been smoothed over age and sex. The spending data were disaggregated by five-year age groups, sex, year, cause, and source. Conversely, attributable fractions were calculated at the age category-cause-source level. Expenditure fractions for cause *i*, age group a_5, sex *s*, and time *t* within cause *i*, age category a_{cat} were calculated as shown in Equation (3):

$$expenditure\ fraction_{ia_5st} = \frac{expenditure_{ia_5st}}{expenditure_{ia_{cat}}} \qquad (3)$$

After calculating expenditure fractions, the total spending was collapsed down to the age category-cause level. This aggregated expenditure was used to calculate the comorbidity-adjusted spending. After adjustment, the expenditure fractions were used to disaggregate the age category-cause-specific expenditure to the age group-sex-year-cause level. The outflows are the resources transferred away from the primary condition to comorbidities. The outflow from primary diagnosis *i* to comorbidity *j* is the product of the attributable fraction AF_{ij} and the total spending of *i*. The total outflow of resources from primary condition *i* due to all comorbidities is the sum of the outflows from *i* to all comorbidities under consideration (vector of *j*), illustrated in Equation (4):

$$outflow_i = total\ expenditure_i * \sum_j AF_{ij} \qquad (4)$$

Within this framework, a primary diagnosis for one health system encounter can be, and generally is, a comorbidity for another primary diagnosis for a different health system encounter. Thus, it was important to not only calculate the share of primary diagnosis *i* attributable to comorbidity *j*, but also to calculate the share of primary diagnosis *j* attributable to comorbidity *i*. These funds are inflows, or the resources transferred to *i* when it is listed as a comorbidity for each of the *j* other causes. The total inflow of resources from all comorbidities to primary diagnosis *i* is the sum of the product of the total spending for *j* and the attributable fractions. Equation (5) illustrates the calculation of inflows:

$$inflow_i = \sum_j (total\ expenditure_j * AF_{ij}) \qquad (5)$$

Because the comorbidity adjustment was a true redistribution of resources, the total outflows across all causes in an age category should have been equal to the total inflows in that age category. That is, the same amount of money should have been flowing out of the primary diagnoses as was flowing into the comorbidities. This assumption was used to check the calculations of outflows and inflows by age category.

The netflow of resources for a primary condition is the net transfer of resources to and from that cause. That is, the netflow for cause *i* is the difference between the total inflows and total outflows for *i*, as illustrated in Equation (6). The netflow can be positive or negative. A positive netflow meant that the given cause had more inflow than outflow. Causes with positive netflows generally appeared often as comorbidities and saw increases in spending as a result of comorbidity adjustment. A negative netflow indicated that the given cause had less inflow than outflow. Causes that appeared often as primary diagnoses, but rarely as comorbidities, often had negative netflows. These causes saw decreases in spending after comorbidity adjustment, relative to their pre-adjustment spending.

$$netflow_i = inflow_i - outflow_i \qquad (6)$$

The final, comorbidity-adjusted expenditure for cause i was the sum of the pre-comorbidity adjusted expenditure for i and its corresponding netflow, as shown in Equation (7):

$$adjusted\ expenditure_i = total\ expenditure_i + \ netflow_i \qquad (7)$$

Relative increases and decreases in spending are described using comorbidity adjustment scalars. The scalar for cause i is defined as the netflow for i as a percent of the total spending on i. This is shown by Equation (8):

$$scalar_i = \frac{netflow_i}{total\ expenditure_i} + 1 \qquad (8)$$

For a given cause, a scalar greater than one represented an increase in spending, while a scalar less than one represented a decrease in spending. The value of the scalar represented the percent change in expenditure for that cause. The scalars provided a common metric for comparing comorbidity adjustments between causes and across age categories and sources.

There was on instance in which comorbidity pairs did not have associated attributable fractions and therefore were not adjusted for comorbidities. These cases were for:

Encounters for individuals under 65 years old that appeared in the CMS data; these encounters were not included due to data sparseness; and

Causes that were restricted so they did not appear as comorbidities (intermediate causes, “other” residual causes, risk factors, etc). For comorbidity pairs that did not have associated attributable fractions, it was assumed that the netflows were zero and that the pre- and post-comorbidity spending values were the same. That is, if there were missing attributable fractions, the causes were considered to have no associated comorbidities and therefore no adjustment.

Applying attributable fractions to other spending sources

Attributable fractions were only calculated for the NIS dataset because it was the only two sources of health spending that included a large enough set of multiple diagnoses. However, this methodology is flexible enough to be applied to any health spending data for age-cause-specific spending estimates. Although the attributable fractions are dependent on the observed patterns of comorbidities in the test data, the final comorbidity adjustment is a function of both these comorbidity patterns and the pre-adjustment spending. Therefore, by assuming that the comorbidity patterns observed in the NIS reflect the patterns in other health spending sources, the attributable fractions from those two sources can be applied to spending estimates from other sources that lack multiple diagnoses. This assumption was utilized to adjust most spending sources used in the wider study for the effects of comorbidities.

Adjusting for Charge Data

Much of the microdata used in this study reports on the charges for an encounter. In order to fully understand the landscape of US health care spending, charge data needed to be adjusted into payment data. An adjustment was developed to enable the use of the National Inpatient Sample (NIS) dataset over the MEPS Inpatient dataset. NIS is very large but contains only data on charges, while MEPS Inpatient provides information on both payments and charges but is substantially smaller. A regression-based framework was used to model total payment to total charge ratios in the inpatient setting. A similar regression was run to model facility charge to total charge ratios. Both regressions were run on MEPS Inpatient data. These ratios were combined to create facility charge to total payment conversion factors. The conversion factors were applied to facility charge data in NIS to produce nationally representative inpatient spending estimates. This charges to payments adjustment is documented in greater detail in other research.[5]

Data processing

Both MEPS Inpatient and NIS data were processed before making these adjustments. NIS was processed according to the methodology described in section 3. MEPS Inpatient was processed differently, because the regression used for this adjustment requires encounter-level data. For MEPS Inpatient, ages were aggregated into 5-year bins, and ICD-9 codes were mapped to GBD causes (see section 3), but the data did not go through redistribution. Consequently, MEPS Inpatient still contained N-codes for injuries, as well as garbage codes. N-codes were removed using the probabilistic replacement method described in section 5b. Garbage codes were dropped.

MEPS Inpatient data were categorized by 3 payer strata: public insurance, private insurance, and out-of-pocket. This strata variable was defined to be the primary payer. For example, if Medicare paid 75% of a patient's total payment and the other 25% was out-of-pocket, the observation was assigned to the public insurance stratum. In addition, facility charges were taken from NIS, and both spending and charge information were taken from MEPS Inpatient. These MEPS spending and charge data were then disaggregated into facility spending, doctor spending, facility charges, and doctor charges. When a patient receives treatment at an inpatient facility, they receive 2 bills: a facility bill and a doctor bill. Facility charges and spending cover basic hospital expenses and most professional fees. Doctor charges and spending cover services for certain doctors who bill separately. These bills generally come from anesthesiologists, radiologists, and pathologists.[12]

Total charges to total payments regression

The ratio of payments to charges was calculated for each encounter. Observations in which payments were greater than charges (<2% of all observations) were considered to be errors, and charges were re-coded to be equal to payments. By inspection, the ratios were found to be invariant by age and sex. Data were grouped by broader causes (GBD cause level 2), in order to increase the number of observations for each cause and payer combination. A model of the charge to payment ratio was run separately for each cause and bootstrap draw, with a binary indicator for payer and an interaction term for payer and year. The equation was as follows:

$$\left(\frac{payments}{charges}\right)_i = \beta_0 \cdot public + \beta_1 \cdot private + \beta_2 \cdot oop + \beta_3 \cdot public \cdot year + \beta_4 \cdot private \cdot year + \beta_5 \cdot oop \cdot year \qquad (9)$$

The above equation defines the payment to charge ratio as a function of cause, payer, and time. However, inspecting trends in the underlying data suggested that total charge itself also has an important influence on the payment to charge ratio, as a person is more likely to pay a smaller proportion of a large charge. In this analysis, conversion factors were applied to data that were aggregated to the age and sex level and had garbage codes redistributed. Consequently, there was no longer information on the amount an individual was charged. To incorporate the effect of charges on the payment to charge ratio at the population level, the total weighted charge was assigned to be the regression weight using the frequency weight option in Stata. The decision to use frequency weights was motivated by the fact that the regression was run to find the percentage paid for each dollar charged. Under this conceptualization, a charge of $100 with a ratio of 0.80 would be equivalent to a ratio of 0.80 for 100 separate $1 charges. By definition, a frequency weight of 100 is treated as if an observation occurred 100 times, so this weighting choice is valid.

For a given cause and draw combination, the regression was run as shown when all payers had more than 200 observations. When a cause, draw, and payer combination did not meet the 200 observation threshold, the corresponding payer-year interaction term was dropped. There are conflicting opinions concerning the number of observations needed to run a multivariate linear regression [13,14]. Several thresholds were tried, and the final decision to set the threshold at 200 best combines goodness-of-fit, trust in the data, and the literature.

Running the regression produced estimates of the payment to charge ratios by year and payer for each level 2 GBD cause. A weighted average of these ratios were taken over payer to get year- and cause-specific estimates. The weights were year-specific proportions of spending on a given level 3 GBD cause from each payer. These proportions were calculated using data from NIS. The averaging resulted in a single cause payment to charge ratio for each year and cause combination.

Facility charges to total charges regression

An additional regression was needed to apply the estimated payment to charge ratios to NIS. Hospital charges are often split into 2 components: facility charges and professional charges. MEPS Inpatient reports both types of charges, whereas NIS reports facility charges only. This paper addresses the cost of receiving inpatient care from the perspective of the patient, so total charges and total payments are the metrics of interest. These totals are equivalent to the sum of facility charges and professional charges, or the sum of facility payments and professional payments, respectively. The payment to charge regression detailed above estimates the ratio of *total* payments to *total* charges. Therefore, a facility charge to total charge conversion was needed in order to estimate the total payments in NIS. This second conversion follows a similar form. The ratio of facility charges to total charges was the dependent variable. This ratio was considered to be a function of cause and time. Inspection of the data showed that this ratio was unrelated to age, sex, and payer. Further, the listed price for a given treatment—what are considered "charges" in this study—are known to be independent of payer within a hospital. The regression was run for each cause-draw combination, with weighted total charges as the regression weight.

$$\left(\frac{facility\ charges}{total\ charges}\right)_i = \beta_0 + \beta_1 year \qquad (10)$$

The accuracy of the second model is limited by differences in how the 2 data sources define facility charges. MEPS Inpatient defines facility charges as the amount a hospital charges a patient. This number often includes fees for a physician's work, in addition to those for the use of the facility, such as bedding or cleaning. However, some physicians charge separately from the hospital, and these separate charges are labeled as professional charges. In contrast, NIS separates all physician charges from hospital charges when possible, even if they were both billed through the hospital.[15] This definitional difference means that "facility charges" in MEPS should tend to be a higher proportion of total charges than they would in NIS. Consequently, our model overestimates the ratio of facility charge to total charge.

Adjusting NIS facility charges

Finally, a cause- and payer-specific facility-charge to total-expenditure conversion factor was calculated:

$$\text{Conversion factor} = \left(\frac{total\ charges}{facility\ charges}\right)\left(\frac{total\ payment}{total\ charges}\right) \qquad (11)$$

Conversion factors are cause- and payer-specific. A weighted average across payer was taken in order to obtain a single conversion factor for each cause-draw combination. The weights were calculated as the draw- and cause-specific proportions of facility charges for each payer at GBD cause level 3 NIS. The weighted average resulted in the final conversion factor, which was applied to NIS after NIS had gone through all of the processing procedures described in section 3.

Estimating Gaps in Data

Second, a systematic model of the relationship between spending, volume and price data was used to address issues of incompleteness and irregularity in the data, fill in missingness, and leverage multiple data sets to produce the best possible estimates. Our model hinges on the following identity:

$$expenditure = volume * price \qquad (12)$$

We use a hierarchical Bayesian model to simultaneously estimate all three variables while preserving this fundamental identity.[16] Our model leveraged data from across years, ages, and datasets to produce credible spending estimates for each age, sex, cause and type of healthcare.

Model Overview

Once the data has been processed, modelling takes place on a draw-specific level. Variance, data sparsity (percentage of data missing in the domain), and age and time splines are calculated from the raw data. If the entirety of expenditure, volume, or price is missing then that draw is skipped altogether. The model is fit by finding the maximum a posteriori estimate via Powell's algorithm using the *PyMC2* package (version 2.3.6) for Python (version 3.5).[17] If Powell's algorithm fails to converge, the missing data are filled in using linear averaging (between observed data points) and missing end points are set equal to the nearest time point, and then the fit is attempted again. If fitting again fails, then the process is stopped. After fitting, all predicted data is outlier-detected using the Median Absolute Deviation (MAD) method with a threshold of 3.5.

Covariates

The backbone of the model is the linear models of the mean of the raw data. For the price model, year and age splines are used as covariates. For the MEPS data, indicators for years before 2007 are used to mark changes in survey design. The volume model includes these as well as indicators for zero and 85 year olds as well as treated prevalence data extracted from MarketScan. The splines are specifically 4th order basis-splines with 16 knots, calculated using the Cox - De Boor algorithm with three additional repeating knots at each endpoint.[18 19] If not enough points are present to generate the splines, then no splines are used. The prevalence data is an age profile that is time-invariant and generated from MarketScan data as the average of 2010 and 2012 count of visits, prescriptions or beddays (depending on type of care). Each covariate used in the linear models are mean-standardized.

Equations

Point estimates are modeled as log-normally distributed with inverse-Gamma distributed variances. An offset is calculated as one percent of the median of the data.

$$y_{expenditure} \sim LogN(\mu_{expenditure} + \rho_{offset}, \quad \sigma^2_{expenditure})$$

$$y_{volume} \sim LogN(\mu_{volume} + \rho_{offset}, \quad \sigma^2_{volume})$$

$$y_{price} \sim LogN(\mu_{price} + \rho_{offset}, \quad \sigma^2_{price})$$

$$\sigma^2_{expenditure,\ volume,\ price} \sim InvGamma(1.0, 5.0)$$

The means are modeled linearly in log space to facilitate positive predictions. Expenditure is explicitly calculated as the product of volume and expenditure to enforce consistency.

$$\mu_{expenditure} = \mu_{volume} * \mu_{price}$$

$$\log(\mu_{volume}) = \beta_0 + i_{invsparsity}\left(\sum_I year_{spline_i} * \beta_i + \sum_J age_{spline_j} * \beta_j\right) + \beta_{inter} * year * age + \beta_{prevalence} * i_{sparstiy} * prevalence + \beta_{zero} * i_{zero} + \beta_{85} * i_{85} + \beta_{MEPS} * i_{MEPS}$$

$$\log(\mu_{price}) = \beta_0 + \sum_I year_{spline_i} * \beta_i + \sum_J age_{spline_j} * \beta_j + \beta_{inter} * year * age + \beta_{MEPS} * i_{MEPS}$$

$$\beta_0 \sim N\left(0.0, \sigma_{\beta_0}^2\right), \quad \sigma_{\beta_0}^2 \sim InvGamma(1.0, 5.0)$$

$$\beta_{inter} \sim N\left(0.0, \sigma_{\beta_{inter}}^2\right), \quad \sigma_{\beta_{inter}}^2 \sim InvGamma(1.0, 5.0)$$

$$\beta_{prevalence} \sim HalfNormal\left(\sigma^2_{\beta_{prevalence}}\right), \quad \sigma_{\beta_{prevalence}}^2 \sim InvGamma(1.0, 5.0)$$

$$i_{MEPS} = \begin{cases} 1, & year < 2007 \\ \frac{1}{2}, & year = 2007 \\ 0, & year > 2007 \end{cases}$$

$$i_{sparsity} = \begin{cases} 1, & sparsity \geq 0.8 \\ 0, & sparsity < 0.8 \end{cases}$$

$$i_{invsparsity} = \begin{cases} 0, & sparsity \geq 0.8 \\ 1, & sparsity < 0.8 \end{cases}$$

Indicators are added for ages zero and 85 when present because these two age categories often represent unique trends in volume and spending, especially because the 85 year old age category is uncapped. The MEPS indicator is included at the recommendation of the survey itself. The outcome data is predicted as if MEPS data were post-2007. Within the volume model, prevalence is only used in draws with high sparsity, greater than 80%. For these draws, it is believed the data is too sparse to inform a good trend with splines, so the MarketScan data is relied upon and the splines are not included.

The coefficients for the splines are determined via a random walk. This method provides some measure of regularization and allows for the inclusion of relatively large numbers of knots compared to the amount of data being modeled while avoiding Runge's phenomenon.[18] This was especially important for this application because extrapolation is commonly performed. The first knot is initialized with an uninformative prior and a value of one, and each subsequent coefficient is walked to according to a normal distribution. For N splines:

$$\beta_0 = 1.0$$

$$for\ i\ \in N, \qquad \beta_i = \beta_{i-1} + e_{i-1}$$

$$e_{i-1} \sim N(0.0, \sigma^2)$$

The variance of the random walk parameter dictates the "smoothness" of the resulting fit, and so is tuned as a function of the sparsity of the data to provide more regularization when less data is present. This is done differently depending on the dependent and independent variable. Maximal flexibility is given to the volume age trends, while less flexibility is given to the year trends in volume and price. Here the inverse variances are represented as that is how the program is specified, so larger values represent more regularization.

$$\sigma_{volume_{year}}^{-2} = \begin{cases} 100, & sparsity \leq 0.4 \\ 333 * (sparsity - 0.4) + 200, & sparsity > 0.4 \end{cases}$$

$$\sigma_{volume_{age}}^{-2} = \begin{cases} 1, & sparsity \leq 0.4 \\ 246 * (sparsity - 0.4) + 1, & sparsity > 0.4 \end{cases}$$

$$\sigma_{price_{year}}^{-2} = \begin{cases} 100, & sparsity \leq 0.4 \\ 333 * (sparsity - 0.4) + 200, & sparsity > 0.4 \end{cases}$$

$$\sigma_{price_{age}}^{-2} = \begin{cases} 50, & sparsity \leq 0.4 \\ 246 * (sparsity - 0.4) + 50, & sparsity > 0.4 \end{cases}$$

Adjusting Mental Health Data

Spending data from the Substance Abuse and Mental Health Services Administration (SAMHSA) were used to adjust our estimates for populations and care settings that are included in the NHEA estimates but out of scope of the surveys used. Inpatient and ambulatory estimates were adjusted.

Data gaps

Goods and services provided at specialty mental health and substance abuse clinics are not accounted for in the sampling schemes of NIS and MEPS. To correct for this, two documents from the Substance Abuse and Mental Health Services Administration (SAMHSA) were used to account for the spending on visits to specialty clinics:

National Expenditures for Mental Health Services and Substance Abuse Treatment, 1986-2005

National Expenditures for Mental Health Services and Substance Abuse Treatment, 1986-2009

SAMHSA reports spending at specialty mental health centers (MHCs) and specialty substance abuse centers (SACs) broken up by type: inpatient, outpatient, and residential. The SAMHSA reports provide spending estimates by MHCs and SACs across inpatient, outpatient, and residential settings for the following years: 1986, 1992, 1998, 2002, 2004, 2005, and 2009. As the NHEA nursing care type excludes MHCs and SACs, only the inpatient and outpatient estimates from SAMHSA were included in the adjustment.

SAMHSA expenditures were converted to real 2014 USD in millions. Spending was imputed using linear regression to fill in estimates for all years from 1996 to 2013. SAMHSA estimates are reported scaled to correspond to the NHEA envelopes, so no adjustment was necessary to line up SAMHSA and the NHEA.

Applying adjustment

As covered in Section five a, the SAMHSA expenditures were first subtracted from the total NHEA envelope for each given type and year. For example, the inpatient expenditure was parsed out into "inpatient expenditure excluding specialty mental health and substance abuse expenditure" and "inpatient specialty mental health and substance abuse expenditure." The ambulatory type was divided in the same manner. Microdata estimates were scaled to the "inpatient expenditure excluding specialty mental health and substance abuse expenditure."

In order to disaggregate the specialty envelopes, cause-, year-, age-, sex-, type-proportions were created from the scaled data. First, scaled spending data was summed by year, type, and whether or not the care was for mental health or substance abuse to mirror the breakdown of the SAMHSA estimates. Then individual scaled spending estimates were divided by them to create scalars. These scalars were used to disaggregate the SAMHSA envelopes to arrive at age, sex, and cause level 3-specific spending estimates proportional to the distribution of mental health and substance abuse causes in non-specialty settings.

Volume of care is not reported in SAMHSA by specialty status. In order to account for the volume of care in specialty settings, volume was back calculated from the newly disaggregated specialty expenditure. First, age-, sex-, year-, type-, and cause-specific ratios of spending to volume were created using scaled data. After specialty spending was disaggregated, these ratios were used to back calculate specialty volume.

A few assumptions had to be made to perform this adjustment. We assume that the distribution of causes, ages and sexes treated at specialty clinics is the same as the distribution treated at non-specialty settings and captured in our micro-data. We also assume that expenditure per visit or bed day at specialty and non-specialty clinics is the same in order to back calculate volume. It is difficult to know the direction of the bias introduced by these assumptions. Assuming an equal distribution of causes, ages and sexes in specialty clinics and non-specialty clinics most likely leads to underestimates of spending on illnesses that more often cause hospitalizations, such as schizophrenia.

Adjusting Nursing-facility Care Data

Data from NNHS, CMS-SNF, and MCBS were used to estimate spending and volume for the nursing care type of service. All three data sources have limitations. NNHS is nationally representative, but it is sparse and only covers three years between 1996 and 2013. CMS-SNF is more comprehensive for short-term nursing home visits but not nationally representative, as it only tracks patients at skilled-nursing facilities (SNFs) who are Medicare-eligible. MCBS covers all nursing home care received by Medicare beneficiaries, so it includes spending at facilities other than SNFs and thus tracks a larger portion of nursing home spending than CMS-SNF, but it is still not nationally representative of all nursing home spending and volume. The goal of combining these three data sources is to apply the time trends found in CMS-SNF and MCBS to the sparse yet nationally representative estimates of NNHS. Short-term and long-term stays are known to have different disease profiles, and they are also known to have changed differently over the past 15 years. Consequently, nursing care spending was estimated separately for short-term and long-term stays [20]. The results were then aggregated to estimate all health spending in nursing homes from 1996 to 2013.

Short-term stays at nursing facilities

Most nursing home care is for people with chronic illnesses that need treatment for the indefinite future[21]. The NNHS finds that 95.5% of all nursing home spending in 2004 was for long-term visits, where people had been in the facility for more than 100 days[22]. This number may be an exaggeration of reality, since the NNHS is known to under-sample short visits, but it confirms the current understanding of who spends the most in nursing homes While long-term care makes up a significant majority of nursing care spending, nursing home care for acute conditions in SNFs has become more common in recent years[20]. These SNFs often aim to have a person leave the nursing home within 100 days, as Medicare coverage only contributes to SNF stays of 100 days or fewer[23].

In this study, short-term stays at nursing facilities were defined as stays of fewer than 100 days. This threshold was chosen to align with that of Medicare's funding policy. Additionally, in tracking nursing care spending, it was assumed that care received at SNFs and captured by CMS-SNF is comprehensive of all nursing care stays shorter than 100 days. The 2004 NNHS finds that 2.8% of all nursing home spending was for stays shorter than 100 days and for which Medicare did not contribute. Consequently, this 2.8% of spending was not accounted for in this study. Additionally, Medicare does not cover all spending for short term stays at SNFs[26]. Analysis of the 2004 MCBS finds that Medicare covers 75% of all money spent for short term stays. However, CMS-SNF provides charges data rather than spending so all charges for this population will be captured, even if Medicare does not cover the entirety of every claim. In other words, the entire charge of a service in a skilled-nursing facility will be included in CMS-SNF, even if Medicare only covers a portion of the cost and the rest must be paid out-of-pocket. However, the fact that CMS-SNF tracks charges itself is a limitation, as charges represent pre-negotiated prices, which are known not to be equal to actual spending.

To properly estimate short-term spending and volume from CMS-SNF, the data were processed similarly to all other data sources as discussed in detail in previous sections. However, placing patients into the five-year age bins used in this study required additional methodology. For those aged 65 and older, CMS-SNF data files categorize patients into the same five-year age bins used in this study. However, due to privacy concerns, CMS-SNF places younger patients into broader age bins. For years 1999 to 2001, all patients in CMS-SNF data under 65 are aggregated into one age bin. Starting in 2002, CMS-SNF files changed the format. These files have more granular estimates, with three age bins for those under 65 years old: ages less than 25, ages 25 to 44, and ages 45 to 64. The assumption was made that, for a given sex and cause, the breakdown of spending and volume across ages is similar for all payers. Therefore, spending and volume for these younger ages were disaggregated into five-year age bins using age-specific proportions of treated prevalence in the long-term healthcare setting, which was estimated by MarketScan.

Data on the number of treated cases in each age, sex, and cause were extracted from MarketScan for the years 2010 and 2012. These data were available for all of the five-year age bins of interest. The number of people within an age and sex group who were treated for a specific cause was summed over the two years. Next, proportions were calculated that described the age distribution of these treated-case data within the wider age bins found in CMS-SNF. The CMS-SNF spending and volume estimates for the wide younger age bins were then broken out into more granular age groups using these age-specific proportions. Each proportion was also matched by cause and sex.

In review, CMS-SNF is a good source to estimate spending in short-term nursing care visits, as it is a census of all claims received by Medicare beneficiaries at SNFs, and it covers many years, However, CMS-SNF is not perfect. It requires the assumption that Medicare beneficiaries at SNFs constitute the entirety of nursing care visits of fewer than 100 days. CMS-SNF tracks charges and not spending, and the assumption that they are equal is known not to be true[4]. Additionally, CMS-SNF requires an extra step of processing, in which younger aggregate age bins are split into the five year bins used in this study.

Long-term stays at nursing facilities

MCBS and NNHS were used to estimate long-term stays at nursing facilities. Medicaid and out-of-pocket spending make up the majority of spending in long-term nursing home visits[24]. Any Medicare beneficiary spending out-of-pocket or through Medicaid is tracked in MCBS. However, those not eligible for Medicare are out-of-scope of MCBS. NNHS, on the other hand, is nationally representative of all nursing home visits. However, NNHS was only run in 1997, 1999, and 2004. Consequently, long-term stays in NNHS were regressed on long-term stays in MCBS to estimate all long-term nursing care spending and volume for the entire period of this study.

NNHS was processed similarly to all other data sources, except that it was not smoothed across time (see section four). It was not smoothed across time because the only years of the study for which it exists are 1997, 1999, and 2004, which meant there were not enough data available for the smoothing model to make valid predictions.

MCBS did not require the same processing steps as the other data sources. MCBS was obtained from the Bureau of Economic Analysis in tabulated form[20]. It was tabulated by age, sex, year, and cause. Ages were aggregated to the five-year age bins used in this study. However, the causes were coded as Clinical Classification Software (CCS) codes rather than GBD causes. Similarly to GBD causes, CCS codes are an aggregated coding of ICD-9 diagnoses. There are 260 mutually exclusive CCS codes. The tabulated MCBS spending and volume estimates were put through the same smoothing machinery as the other data sources and as described in section four. However, in order to use MCBS as a time trend for NNHS, NNHS's estimates, stratified by GBD cause, had to be mapped to MCBS estimates stratified by CCS code. GBD causes do not perfectly align with CCS codes. A CCS code might be made up of ICD-9 diagnoses that map to multiple GBD causes. Similarly, a GBD cause might be made up of ICD-9 diagnoses that map to multiple CCS codes. For each GBD cause present in NNHS, each CCS code that shared a common ICD-9 code was found. Then each GBD cause and sex combination in NNHS was analyzed individually. First, spending for a given GBD cause and sex was compared to spending for each CCS cause and sex mapped to it. If the spending between the two was positively correlated across time, the CCS code was considered to be appropriately mapped to the GBD cause. If the spending between the two was negatively correlated across time, the CCS code was considered to be poorly mapped to the GBD cause, and this CCS code time trend was not used. For example, if a GBD cause and CCS code only shared one ICD-9 diagnosis that appears rarely in the nursing care setting, these time trends would not necessarily be correlated, and the CCS code would be dropped from the analysis.

A regression was run for each CCS code that shared an ICD-9 diagnosis with a GBD cause and was positively correlated with the time trend for a given GBD cause and sex. Specifically, a sex- and GBD cause-specific mixed effects regression was run on NNHS to estimate nationally representative spending and volume for long-term nursing care visits across the entire time period of interest. The regression was given by:

$$NNHS_{age,sex,i} = \beta_0 + \beta_1 MCBS_{sex,j} + \mu_{age} MCBS_{sex,j} + \varepsilon_i$$

where *i* is a GBD cause and *j* is a CCS code that maps to it. If the regression did not converge after 200 iterations, a linear regression was run. In this linear regression, cause- and sex-specific NNHS spending was regressed on MCBS spending and fixed effects on age. If no CCS codes were associated with a given cause, an average was taken across time, with random intercepts on age. If multiple CCS codes were associated with a GBD cause, multiple regressions were run. The root-mean-square error for each regression was calculated. Then a weighted average of the different outputs was calculated, with the weights being the normalized inverse of the root-mean-square error. In this way, the most weight was given to the CCS codes that best predict the NNHS data.

eAppendix 5. Uncertainty

Bootstrapping

To obtain uncertainty, all data sources were bootstrapped 1,000 times at the beginning of the analysis. Encounters was bootstrapped stratified by year and data source, creating 1,000 individual samples on which to run analysis. Complex survey design was taken into consideration for bootstrapping by using the user-written *bsweights* command in Stata 13.131. This command ensured that the bootstrapped data resembled the original sampling scheme by resampling the whole primary sampling unis with each strata.

All statistical analyses were performed at the bootstrap draw level. This includes redistribution of garbage code, the three-digit ICD9 -codes in the MEPS adjustment, the comorbidity regression, the charges-to-payment regression, the Bayesian hierarchical model, the long-term adjustment, and scaling to the National Health Expenditure Account envelopes.

Final estimates and uncertainty intervals

After the data were fully adjusted, final estimates and uncertainty intervals were calculated across the one thousand draws. Final estimates were the mean of spending or volume for each age, sex, condition, year, and type combination. Uncertainty intervals were taken to be the 2.5th and 97.5th percentiles.

eReferences.

1. Bui AL, Dieleman JL, Hamavid H, et al. Spending on children's personal health care in the United States, 1996-2013. JAMA Pediatr. Published online December 27, 2016. doi:10.1001/jamapediatrics.2016.4086.

2. Center for Medicare and Medicaid Services. *National Health Expenditure Accounts: Methodology Paper, 2013*. Washington D.C.; 2014. http://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/Downloads/DSM-13.pdf. Accessed May 12, 2015.

3. Tompkins CP, Altman SH, Eilat E. The Precarious Pricing System For Hospital Services. *Health Aff*. 2006;25(1):45-56. doi:10.1377/hlthaff.25.1.45.

4. Reinhardt UE. The Pricing Of U.S. Hospital Services: Chaos Behind A Veil Of Secrecy. *Health Aff*. 2006;25(1):57-69. doi:10.1377/hlthaff.25.1.57.

5. Hamavid H, Birger M, Dieleman J, et al. Assessing the complex and evolving relationship between charges and payments in US hospitals: 1996-2012. *Accepted by PLOS ONE*. August 2015.

6. International Monetary Fund. *World Economic Outlook*. Washington D.C.; 2015. http://www.imf.org/external/pubs/ft/weo/2015/01/weodata/index.aspx. Accessed April 12, 2015.

7. Global, regional, and national age–sex specific all-cause and cause-specific mortality for 240 causes of death, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *The Lancet*. 2015;385(9963):117-171. doi:10.1016/S0140-6736(14)61682-2.

8. Organisation for Economic Co-operation and Development. *OECD.StatExtracts*. http://stats.oecd.org/. Accessed October 7, 2014.

9. Agency for Healthcare Research and Quality. *United States Medical Expenditure Panel Survey*. Rockville, United States: Agency for Health Care Research and Quality

10. Substance Abuse and Mental Health Services Administration. *National Expenditures for Mental Health Services & Substance Abuse Treatment: 1986 - 2009*. Rockville, MD: Substance Abuse and Mental Health Services Administration; 2013.

11. Lefit K, Kassed C, Coffey R, et al. *Projections of National Expenditures for Mental Health Services and Substance Abuse Treatment, 2004–- 2014-*. Rockville, MD: Substance Abuse and Mental Health Services Administration; 2008.

12. MEPS HC - 135D: 2010 Hospital Inpatient Stays. July 2012. http://meps.ahrq.gov/mepsweb/data_stats/download_data/pufs/h135d/h135ddoc.shtml.

13. Schneider A, Hommel G, Blettner M. Linear Regression Analysis. *Dtsch Arztebl Int*. 2010;107(44):776-782. doi:10.3238/arztebl.2010.0776.

14. Green SB. How Many Subjects Does It Take To Do A Regression Analysis. *Multivariate Behavioral Research*. 1991;26(3):499-510. doi:10.1207/s15327906mbr2603_7.

15. HCUP NIS Description of Data Elements. ptember 2008. http://www.hcup-us.ahrq.gov/db/vars/totchg/nisnote.jsp.

16. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. *Bayesian Data Analysis*. 3rd Edition. Boca Raton, FL: Chapman and Hall/CRC Press; 2014.

17. Patil A, Fonnesbeck C, Huard D, Salvatier J. *PyMC2*. (Version 2.3.6); 2005. https://github.com/pymc-devs/pymc.

18. Lang S, Brezger A. Bayesian P-Splines. *Journal of Computational and Graphical Statistics*. 13(1):183-212. doi:10.1198/1061860043010.

19. Foster JT. *Bspline*. (Version 1.0); 2015. https://github.com/johntfoster/bspline.

20. Tina Highfill, David Johnson. Measuring Nursing Home Price Growth between 2000-2009. http://www.bea.gov/papers/pdf/Highfill%20Johnson_Nursing%20Homes%20MCE%20index.pdf. Accessed May 12, 2016.

21. Thomas Day. About Nursing Homes. https://www.longtermcarelink.net/eldercare/nursing_home.htm. Accessed May 12, 2016.

22. *National Nursing Home Survey: 2004 Overview*. Washington, DC: Center for Disease Control and Prevention; 2009. http://www.cdc.gov/nchs/data/series/sr_13/sr13_167.pdf. Accessed May 12, 2016.

23. Skilled nursing facility (SNF) care | Medicare.gov. Medicare.gov The Official U.S. Government site for Medicare. https://www.medicare.gov/coverage/skilled-nursing-facility-care.html. Accessed May 12, 2016.