Journal of Forensic & Investigative Accounting
Volume 10: Issue 3, July–December 2018

Healthcare Data Sources and Fraud Research Opportunities

Renee Flasher
*Melvin A. Lamboy-Ruiz**

I. Introduction

> The National Health Care Anti-Fraud Association estimates conservatively that health care fraud costs the nation *about $68 billion* annually—about three percent of the nation's *$2.26 trillion* in health care spending. Other estimates range as high as ten percent of annual health care expenditure, or *$230 billion* [emphasis added].
>
> —The National Health Care Anti-Fraud Association (NHCAA, 2017)

The healthcare sector has significant expenditures and structural issues that motivate fraud and provide research opportunities that could impact policymaking. An estimate of annual healthcare expenditures in the U.S. economy is currently greater than two trillion dollars, with projected increases greater than five percent annually.[1] The trend of growing expenditures places pressure on the largest federal government healthcare programs, Medicare and Medicaid, to control their spending. They consume an ever-increasing portion of mandatory spending each fiscal year, competing with the amount of discretionary funding allowable for other federal government programs. The Congressional Budget Office (2017) projects that Medicare spending, combined with Medicaid and Social Security, will account for half of all federal spending in 2047 after excluding interest payments on the federal debt. However, fraud constitutes a significant portion of the expenditures in federal government health-related programs (Krause, 2003). The magnitude of this issue provides an opportunity for impactful academic research.

The federal government has increased its financial data releases from the Centers for Medicare and Medicaid Services (CMS) along with other federal government entities with the goal to "promote better care, smarter spending, and healthier people" (CMS, 2015). For example, in 2014, the U.S. Congress passed the Digital Accountability and Transparency Act directing the federal government to transform all spending information into open data. Any changes that the CMS ultimately makes to its healthcare offerings, based partly on research done using these data, may have a repercussive impact because such data-driven changes may influence the actions taken by private insurers.

While researchers need contextual, industry-specific knowledge to execute any healthcare research project, academic researchers could make significant contributions to the policy discussion around healthcare fraud in particular. This potential for contribution is what motivates our survey of the data sources available for healthcare fraud research and our review of such research with an accounting focus. In addition, we describe some other data sources that provide information needed to normalize or benchmark entities' financial activities, which in turn allows researchers to identify potential outliers that are common targets for fraud examination.

The remainder of the paper is organized as follows. Section II describes the pervasiveness of fraud in the U.S. healthcare industry. It also reviews healthcare fraud research and the data sources used to conduct it. Section III introduces two key criminology theories that could be tested using healthcare related data sources. This highlights the potential ability to learn about perpetrators in a white-collar crime environment. Section IV discusses additional data sources. Section V warns of some limitations with using the data sources surveyed, and Section VI concludes.

[1] The CMS (n.d.b) details a current breakdown of national health expenditures for the United States.

*The authors are, respectively, an assistant professor at Penn State Harrisburg and an assistant professor at Iowa State University. They would like to acknowledge the research assistance of Anthony Tuel.

II. Fraud in Healthcare

Prevalence of Healthcare Fraud

The U.S. Government Accountability Office (GAO) recognized in 2016 that there are no reliable estimates of the magnitude of fraud in federal healthcare programs or across the healthcare industry, even after having convictions for multimillion-dollar schemes that defrauded this industry (GAO, 2016). A significant factor leading to the unreliability of estimates is that healthcare fraud in the U.S. exists in a complex environment composed of multiple fraudulent schemes, perpetrated by multiple players, affected by multiple payers, and investigated by many agencies at various levels of government.

Schemes. In its report, the GAO documents types of healthcare fraud that include fraudulent billing of services, kickbacks, and mismanagement of prescription drugs. The catalogued abuses of these types are many: (i) billing for services not provided; (ii) billing for services provided that were not medically necessary; (iii) "upcoding," or intentionally billing for services at a higher level than appropriate for the services that were provided; (iv) provision of compensation to healthcare providers, beneficiaries, or other parties for participating in the fraud scheme; (v) submission of false claims for prescription drugs that have been improperly marketed for non-FDA-approved uses; and (vi) illicit diversion of prescription drugs for profit or abuse (GAO, 2016).

Players. In addition to multiple types of fraud schemes, multiple players are involved.[2] Medical facilities and durable medical equipment suppliers are the most common provider or supplier types that were found or plead guilty to criminal fraud charges. Hospitals and medical facilities are the most common defendants in civil fraud investigations resulting in judgments or settlements. As physicians are the principal gatekeepers in the provision of healthcare, they are frequently associated with fraud investigations and convictions as the sole perpetrator or in partnership with other players (Taitsman, 2011).

Enforcement. Several federal, state, and local agencies are involved in investigating and prosecuting healthcare fraud cases, including the CMS; the Department of Health and Human Services Office of Inspector General (HHS-OIG); the Department of Justice's Offices of the US Attorneys, Civil and Criminal Divisions, and the FBI. At the state level, Medicaid Fraud Control Units (MFCUs) and State Fraud Bureaus also fight healthcare fraud. The National Health Care Anti-Fraud Association (NHCAA) is a nonprofit private organization dedicated to fraud control. The association's website lists additional resources to fight fraud including organizations at the state level and local level like the State Fraud Bureaus Directory and Utah Health Care Anti-Fraud Alliance.

Fraud waste. Fraudulent billing schemes deviate funds meant for reimbursing legitimate claims away from the intended parties and into "private hands," thereby generating healthcare waste: that is, funds used improperly without producing any benefits to the payer or intended recipient of the healthcare service. The mix of multiple fraud investigators and enforcement groups has not provided a consistent or reliable estimate of healthcare fraud waste in the United States. The NHCAA (2017) estimates conservatively that healthcare fraud costs the nation many tens of billions of dollars annually. In 2016 the CMS estimated that Medicare and Medicaid made ninety-seven billion dollars in improper payments, although the level of fraudulent payments has not been provided (NHCAA, 2017).[3]

Although it is not possible to provide a dollar estimate of fraud waste, we provide frequencies and trends in cases associated with fraud and waste as reported in academic literature in the next subsections.

Research in Healthcare Fraud

Reviews of literature. Clemente et al. (2017) performed a literature review of legislative reforms, related databases, and research studies to examine the effects of the Patient Protection and Accountable Care Act (ACA) of 2010 in combating Medicare fraud. They find that legislative reforms seem to be incompatible with the ACA's new reimbursement reforms. Joudaki et al. (2015) reviewed studies that performed data-mining techniques for detecting healthcare fraud and abuse using supervised and unsupervised data-mining approaches. They find that most of the studies examined have focused on algorithmic data mining without an emphasis on or application to fraud detection efforts in the context of health service provision or health insurance policy and so are disconnected from sound and evidence-based diagnosis and treatment

---

[2] The GAO (2016) based its report in court cases that were subject to civil and/or criminal fraud investigations in 2010.
[3] The NHCAA (2017) lists other estimates of healthcare fraud and the sources used to make the estimates.

approaches to curbing fraudulent or abusive behaviors. Travaille et al. (2011) conducted a systematic literature review to analyze the applicability of existing electronic fraud detection techniques in the insurance, telecommunications, and financial industries to the U.S. Medicaid program. These and other reviews of the healthcare fraud literature listed in Table 1, Panel A underscore the need for data customized to the type of fraud under investigation. [see Table 1, pg 389]

Fraud convictions. Flasher and Lamboy-Ruiz (2017) studied yearly frequencies of individuals publicly excluded by the HHS-OIG from 2007 to 2014 and report some trends in the volume of fraud perpetrators. They document an almost monotonic increase in the number of convictions for program-related crimes associated with fraudulent billings, representing an increase of 222 percent of cases in eight years.[4] Felony convictions related to healthcare fraud increased from 219 cases in 2010 to 351 cases in 2014, showing a sixty percent increase in four years. Exclusions related to physicians who were not properly licensed for the services they provided decreased thirty-five percent, on average, for years 2009–2011 but later increased forty-seven percent, on average, for years 2012–2014.

Upcoding. In the current body of healthcare fraud detection research, upcoding fraud studies are scarce. Bauder, Khoshgoftaar, and Seliya (2017) provide a review of upcoding in healthcare and of the current data-mining techniques used therein. Recent evidence for the prevalence of upcoding—which amounts to overbilling—by for-profit U.S. hospitals has been documented by Heese (2017). He finds that overbilling allows hospitals to increase revenues without altering operations, affecting costs, or having to reverse such behavior in the future.

In general, academic studies support the argument that healthcare fraud is difficult to estimate and is affected by changes in technology and the reimbursement systems adopted by payers of healthcare services.

Trends in the Healthcare Industry Impacting the Prevalence of Fraud

Current trends in the healthcare industry such as changes in the systems and procedures for reimbursing providers and increases in the coverage of healthcare services for the U.S. population are impacting the healthcare fraud environment (e.g., the mix of patients, healthcare providers, and payers).

Frequent changes in reimbursement. Changes to the reimbursement formula used by the CMS to compensate healthcare providers clearly affect billing procedures, making them more complex and therefore a target of continuing fraud. The major payer of healthcare services in United States, the CMS introduced the hospital outpatient prospective payment system in 2000, which changed reimbursement for outpatient services. The Children's Health Insurance Program (CHIP) Reauthorization Act of 2009 established new incentives for covering children through changes in reimbursement from Medicaid.[5] The ACA's new initiative called the Hospital Value-Based Purchasing Program adjusts payments to service providers based on the quality of care and resulted in 778 hospitals losing more than 0.2 percent of their Medicare reimbursement in 2014 (Becker's Hospital CFO Report, 2014).

The Medicare Access and CHIP Reauthorization Act (MACRA) of 2015 transformed the way physicians are reimbursed by Medicare. Considering that the main and direct provider of services is still the physician, the new value-based payment methodology keyed to multiple factors may provide more incentives to commit fraud. There has been a comprehensive adoption of new technology in the provision of healthcare, including electronic medical records (Wang, Wang, and McLeod, 2018). The use of this technology may have added more incentives to commit fraud after the CMS started reimbursing hospitals partially for their investment in hospital information systems and technology.

Big Data. The "Big Data" era in accounting and healthcare, driven by advances in technology and informatics, makes feasible the public availability of data sources that may facilitate research into the avenues of fraud. Data transparency initiatives at the federal and state levels are leading to public access to online healthcare information through such means as Guidestar's IRS Form 990, Hospital Compare, the Medical Expenditure Panel Survey (MEPS), Risk Sharing, and HHS-OIG's list of exclusions and breach data (Becker's Hospital CFO Report, 2014). The proliferation of databases publicly

[4] Table 1 in Flasher and Lamboy-Ruiz (2017) shows the convictions for program-related crimes associated with fraudulent billings, which are classified by the HHS-OIG as type 1128a1. All exclusions examined in their study are cases at the individual level (e.g., physicians and nurses) and do not include cases at the institutional level (e.g., hospitals and pharmaceutical companies).
[5] CHIP provides healthcare coverage to eligible children, through both Medicaid and separate CHIP programs. CHIP is administered by states, according to federal requirements, and funded jointly by states and the federal government.

available led to a stream of new data-mining techniques needed to extract information used in fraud research (Bauder and Khoshgoftaar, 2017; Bauder et al., 2017).

In the next section, we discuss how the availability of existing and newly created healthcare-related databases, as a result of the current trends in healthcare, creates a unique opportunity to expand research on the topics just discussed.

III. Potential Applications of Databases to Fraud Research

Fraud, waste, and abuse are not easily separated within the healthcare industry; however, all detract from patient care resources. Thus, we leverage two different theories to frame the fraud discovery process while acknowledging that we cannot separate fraud from waste/abuse and that only a court of law can effectively adjudicate fraud. The two theories we explore for their application to the healthcare industry are general strain theory (Agnew, 1992) and routine activity theory (Cohen and Felson, 1979).

General Strain Theory

General strain theory posits that external circumstances may result in a negative emotional reaction by a perpetrator who, when legitimate means do not alleviate the strain, turns to illegal behavior to redress the situation (Agnew, 1992). Langton and Piquero (2007) provide evidence that individuals may turn to white-collar crime out of external negative pressures, but they do not find the same trend with corporate offenses. The potential for negative external pressure on individuals employed in healthcare is evident from continual revenue maintenance and maximization pressures. These pressures are acute within the healthcare sector when regulations continually change and reimbursement pressures are consistently downward (e.g., Axelrod, Millman, and Abecassis, 2010; Carrigan and Kujawa, 2006). Particularly with the focus on federal reimbursements trending to value-based care instead of fee-for-service, researchers can study the consequences of the paradigm shift empirically.

As individuals feel economic pressure to maintain their income as the payment regimes shift, data exists that allows researchers to baseline amounts and study outliers under the new regime. For example, to baseline under the fee-for-service regime, physicians' Medicare reimbursements from fee-for-service data can be accessed through the CMS's Medicare Provider Utilization and Payment Data (PUP files) at https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/medicare-provider-charge-data/index.html. These files detail payments made by Medicare, the federal government program covering seniors, for procedures and drugs for individuals on traditional Medicare (as opposed to Medicare Advantage). Separate files exist for Physician and Other Supplier; Inpatient; Outpatient; Part D Prescriber; Referring Durable Medical Equipment, Prosthetics, Orthotics, and Supplies; Home Health Agencies; Skilled Nursing Facilities; and Hospice Providers. Specifically, for the Physician and Other Supplier files, there are years of data covering over 900,000 healthcare providers at a procedure code level with the average amount billed to Medicare and the average amount paid by Medicare when the provider serviced at least ten fee-for-service beneficiaries. Data releases have occurred every April/May for the past few years; thus data covers calendar years 2012–2016.[6]

With its introduction in April 2015, MACRA has implications for healthcare spending in outpatient locations, among others (CMS, n.d.a), as this is a further step toward value-based-care payments. Payment incentives under MACRA comprise metrics along four dimensions—cost, quality, information technology, and improvements. The CMS is planning on releasing these incentive payment data at a physician level (https://data.medicare.gov/data/physician-compare) (SA Ignite Inc., 2016).

Therefore, having these data available, a researcher can identify individuals with substantial practice revenue related to fee-for-service and then examine whether they "win" or "lose" under the new payment methodology. This could potentially highlight individual practices that might be susceptible to fraudulent behavior as an individual feels pressure to maintain his/her lifestyle under the new payment regime.

Another example of a way to examine incentives for healthcare stakeholders would be to leverage the open payments data that reflect payments made to physicians or hospitals by pharmaceutical or medical device companies. Open payments

---

[6] Many of the Medicare files are large and require a statistical package like SAS to process because spreadsheet software cannot handle the record counts.

include consulting fees, gifts, and entertainment, among other items. For example, a pharmaceutical company may sponsor a dinner where physicians also hear a presentation concerning a new drug. These physicians should be included in the open payments file. Thus, these data can highlight potential conflicts of interest or incentives for a physician or a hospital. The open payment files, accessed through https://www.cms.gov/openpayments/, covers program years 2013–2017, with 11.9 million records from 1,456 companies showing aggregated payments of $7.5 billion for 2015 alone. Combining the open payments data with the PUP files discussed above might identify those physicians and others with financial incentives beyond patient service revenue.[7] Empirically applying cluster analysis to these data may reveal physicians taking advantage of patients according to significantly different patterns of procedures relative to fellow physicians.

One other avenue for incentive research involves leveraging Part D prescriber data from the PUP files. Combined with the Part B data (the Physician and Other Supplier PUP file), analysis may provide answers to questions about pressures to maintain revenue from patients when a physician practice joins an approved alternative payment structure. Under some of these alternative payment structures, there is an opportunity for risk sharing with the payer. Risk-sharing relationship data can be obtained from the CMS: https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/sharedsavingsprogram/program-data.html. Do patient care procedures significantly differ between the risk-sharing and the non-risk-sharing payment regimes for physicians in the same geographic locations? If they do differ, the results might speak to the perception that risk of payment is a significant influencer of physician behavior, providing an opportunity to examine an incentive contributing to potentially fraudulent behavior.

Any time that payment methodologies change, people find new ways to commit fraud, especially when part of their rationalization framework is taking money away from the government or an insurance company. Tennyson (2008) discusses this rationalization as one in which fraudsters excuse their crimes because they perceive them to be "victimless" (p. 1194). Exclusion data reveal organizations and individuals that are effectively barred from billing a federal government health program for reasons including committing healthcare fraud. The HHS-OIG publishes exclusion data monthly (https://exclusions.oig.hhs.gov/). At the state level, exclusion lists may also identify healthcare providers that can no longer bill Medicaid (e.g., for the state of Texas, see https://oig.hhsc.state.tx.us/oigportal/EXCLUSIONS.aspx). The excluded individuals have successfully rationalized unacceptable behaviors, including bad billing practices. Using these exclusion lists to identify sanctioned individuals along with an examination of the PUP or MACRA files could reveal if billing patterns for certain patient populations were consistently in line with those of fellow healthcare providers or statistically different prior to exclusion. This analysis might provide fraud insights that would allow patients to be more vocal when they see suspicious items on their explanation of benefits.

Routine Activity Theory

Routine activity theory has three key elements, but we focus on the element where the target organization provides an opportunity for the fraud to occur (Clarke and Felson, 1993). A perpetrator's ability to execute a scheme is partially dependent on organizational opportunities presented to the individual. The organizational opportunities arise due to changes in how healthcare is delivered and will continue to evolve with the expansion of technologies within healthcare and medicine.

One major shift that researchers can exploit is the adoption of electronic health records (EHR). The Health Information Technology for Economic and Clinical Health (HITECH) Act of 2009, combined with payment incentives from the CMS, spurred significant adoption of EHR applications throughout the industry (Tripathi, 2012). EHR adoption began with hospitals and continues to proliferate throughout other entities in the sector. The adoption of EHR systems can result in more or less integration between patient care and accounting/billing functions and among practice locations. Maintained by the Office of the National Coordinator for Health Information Technology, data accessed through HealthIT.gov includes the HealthIT dashboard (http://dashboard.healthit.gov/datadashboard/data.php). Specific datasets address the adoption of EHR including a list of certified vendors and products. For example, the Non-federal Acute Care Hospital Health IT Adoption and Use data, which summarize patient engagement statistics, could be combined with state exclusion data to see if patient access to information could be an effective control in reducing negative professional behavioral outcomes.

[7] PUP files data have been used to examine trends in payments made to physician subspecialties, such as Chang (2015) did for ophthalmologists.

Also, HITECH requires breach reporting for this single industry across the country, as opposed to the state laws governing breaches for most other industries. This speaks to opportunities for fraud, as less secure systems can be breached more easily than more secure ones, all else being equal. The Office for Civil Rights maintains data on breaches of healthcare information (https://ocrportal.hhs.gov/ocr/breach/breach_report.jsf). This dataset is updated for any breach affecting more than 500 individuals' health information and covers entities beyond healthcare providers and insurance plans. For example, if a public company's employee health information was compromised at a sufficient level, the breach should be reported. Moreover, these breaches are defined as disclosures of information involving any media type (e.g., from paper records to electronic records stored on a server or a portable device). These breaches can result in subsequent identity theft and theft of healthcare resources.

Breach data can be used to examine the impact of firms' internal controls. Nonprofit healthcare providers' IRS Form 990 data could be examined for financial impacts after a data breach with contribution changes or expense shifts.[8] Form 990 information can be accessed through Guidestar (http://www.guidestar.org/Home.aspx), albeit challenges exist to using the Form 990 data (Feng et al., 2014).

Other Theories and Published Studies

While we focus on two specific theories in this paper, there are additional criminology and fraud theories and techniques that can be applied. For example, discourse analysis can take the qualitative information within healthcare, for example, patient charts, and examine the language to highlight false or misleading documentation (Busch, 2012). In addition to testing theories, there are additional methodological and technical advances that can be applied to healthcare data. For example, Joudaki et al. (2015) summarize the application of knowledge discovery databases, a data-mining application, to the healthcare industry as undertaken in studies from around the globe. Moreover, our Table 1, Panel B summarizes recently published healthcare fraud research using the data sources described in this paper and adds our own suggestions for extending this research. As these studies demonstrate, archival healthcare research depends on the existence and availability of data from many academic disciplines, so we discuss other data sources in the next section.

IV. Additional Data Sources

Data Sources Used in Published Accounting and Fraud Research

Extant research focusing on healthcare fraud does exist across a variety of disciplines and geographic locations as reflected in the variety of publication outlets (e.g., Yang and Hwang, 2006; Thornton et al., 2013; Rudman et al., 2009; Kang et al., 2010; Krause, 2004; Gee, Button, and Brooks, 2009; Hannigan, 2006; Liu and Vasarhelyi, 2013). Wang and Yang (2009) specifically discuss a data-mining example for healthcare fraud detection using insurance claim data. From an empirical research viewpoint, Liou, Tang, and Chen (2008) use data from the National Health Research Institute in Taiwan to examine the classification accuracy of several methods for detecting fraudulent claims. However, there remain significant amounts of unexplored territory.

In a manner similar to instances of fraud in healthcare being tangentially addressed in general fraud articles (e.g., Lindberg and Seifert, 2016; Lynch, Bryant, and Reck, 2011; Burnaby, Howe, and Muehlmann, 2011), other data sources for healthcare fraud research can be identified through examining past healthcare accounting articles.[9] These accounting data sources can provide opportunities to baseline levels of activities or develop expectations for normal behavior. This allows a fraud researcher to leverage unusual or different behaviors as potentially interesting to pursue through further investigation. These accounting data sources include publicly available state resources (e.g., Eldenburg, et al., 2004), requests for data (e.g., Eldenburg et al., 2010), a professional organization's survey data (e.g., Ittner, Larcker, and Pizzini, 2007; Kim, 1988; Pizzini, 2006, 2010), Medicare cost reports (e.g., Eldenburg and Kallapur, 1997; Hwang and Kirby, 1994), and combinations of these or other sources (e.g., Eldenburg and Vines, 2004; Evans et al., 2010; Krishnan and Yetman, 2011). The variety of sources, both public and private, reflect the disaggregate nature and immense amount of data that exist. Research

[8] The IRS Form 990 "Return of Organization Exempt from Income Tax" is an informational tax form that most tax-exempt organizations must file annually. The form gives the IRS an overview of the organization's activities, governance, and detailed financial information. It also includes a section for the organization to outline its accomplishments in the previous year to justify maintaining its tax-exempt status.

[9] See Eldenburg and Krishnan (2007) for a review of data sources used in managerial accounting and control research in healthcare.

opportunities exist to link the medically related data with overall financial performance data captured by traditional accounting metrics for further study of the organizational influences on fraud.

Other Healthcare-Related Data Sources

Paid or restricted access. Accessibility to some healthcare data sources is restricted by on-demand access fees or through subscription memberships to healthcare-related associations or organizations. Some examples include (1) the American Hospital Association Annual Survey Database, now accessible through Wharton Research Data Services; (2) costs of medical group practices and productivity, collected by the Medical Group Management Association; and (3) the Massachusetts All-Payer Claims Database, which has an application process to access data that includes a potential fee waiver under certain circumstances. There are also data sources with free-of-charge access but that require users to have technical knowledge of the data structure in order to prepare the data for final use (e.g., the Healthcare Provider Cost Reporting Information System, operated by the CMS).

Federal sources. HSS, the federal government agency with the largest budget (Greer, 2016), is responsible for providing many of the data sources discussed in this paper. Figure 1 charts the offices and divisions under the HHS umbrella that produce, collect, compile, or disclose various data items discussed in this paper. [see Figure 1, pg 388]

Beyond the other HHS sources named in this paper, the CMS maintains a Research Statistics Data and Systems overview page with links to free and subscription-based resources. Another HHS source is HealthData.gov (https://www.healthdata.gov/), which showcases resources for healthcare research by aggregating data from a multitude of local, state, and federal sources. In a similar manner, HHS also feeds information into the Data.gov healthcare section. Both of these websites aggregate data from other sources, so there is some overlap in content between them.

The Agency for Healthcare Research and Quality, another HHS agency, conducts annual surveys at the national level. One extensively used survey is MEPS https://meps.ahrq.gov/mepsweb/data_stats/download_data_files.jsp). The MEPS annually surveys patients and providers along with employers to collect insurance information. The survey provides information on how patients pay for healthcare, including their out-of-pocket spending.

State-level resources. States require healthcare entities to report information annually and make some items publicly available. A cautionary note should be taken, as Lamboy-Ruiz, No and Watanabe (2018) document discrepancies between financial items reported in the Medicare cost reports and the ones reported in hospital financial statements filed to state data repositories. Appendix A lists state data repositories. Additionally, Appendix B lists websites of state repositories that compile community benefits, such as charity care, provided by hospitals.

For a quick reference list of all healthcare databases discussed above, and other additional sources, see Table 2. The resources are listed by topic, along with the managers of the data sources and their respective websites. [see Table 2, pg 391]

V. Data Quality Issues

Researchers encounter challenges when performing healthcare fraud research and must be aware of these potential roadblocks. First, the importance of understanding the data cannot be understated in this complicated area of research. For example, if the data assume a copay by the patient who pays only eighty percent of a medical charge, this may be a critical piece of information if one is examining the revenues received by a particular medical practice. Another example of the specialized knowledge needed to interpret healthcare data properly comes from Bauder et al.'s (2017) survey paper involving one particular healthcare practice—upcoding. Upcoding occurs when a provider bills for a different procedure code than the correct procedure code for the service performed, and this alternative procedure code results in a higher payment than the correct code. Researchers would have to know the proper code to determine what constitutes upcoding and would also have to make some allowances for simple human error on the part of the original coders.

Second, data cleansing may need to occur, as the compilation processes behind these datasets are rarely standardized or entirely automated. Spelling errors abound. Sometimes, the name of a healthcare provider system is used while, at other times, a component piece of the same system is used (e.g., a hospital in a larger hospital system). If multiple states submit information to the same federal data source, this source may contain duplicate entries. For example, if an individual is sanctioned in one state, that person may move to another and also be sanctioned there. Thus, the total number of sanctions may overstate the number of individuals who actually commit violations.

Finally, another limitation is that rarely do researchers have the full picture of revenues or expenses for a particular entity, nor does the literature agree on a single source of information for these financial numbers. For example, Sing et al. (2006) attempt to reconcile MEPS expenditure-level data with National Health Expenditure Accounts, another source used by healthcare researchers. Kane and Magnus (2001) specifically compare Medicare cost reports with audited financial amounts and discuss differences in amounts. These articles highlight the challenges that the lack of having a single source of financial information for even an element, like expenditures or profit, poses.

Also, the data sources may help to direct researchers to the appropriate publication outlet for a healthcare fraud article, as there is not an established home for fraud research within a particular discipline. Due to this limitation, this paper mostly surveys federal websites, as these have the largest record sets available and, arguably, may have data with broader generalizability than a single state dataset.[10] However, an empirical researcher may exploit state variations in program implementations for answering different questions.

VI. Conclusion

Policy debates around healthcare access and cost continue to involve all levels of government and elements of society throughout the U.S. At the same time, about $3.3 billion in healthcare fraud judgments and settlements were collected in a single year, 2014, as a result of the HHS-OIG's and Department of Justice's investigations and prosecutions (GAO, 2016). If researchers can effectively leverage nontraditional data sources, sometimes in combination with more traditional accounting data sources, we may bring additional transparency to healthcare spending within the U.S. With transparency, discussions about common attitudes and realistic expectations might happen that also benchmark normative practices for professions. These discussions assist in ensuring that individuals know where the bright lines are drawn and can assist in deterring fraud.

This paper focuses on archival data sources that could be used directly for empirical research or used to motive behavioral research to address healthcare fraud. Although we have not attempted a complete, exhaustive detailing of data sources, we hope that our survey of data sources promotes discussions and ideas for combating healthcare fraud throughout the system, from home health to skilled nursing facilities to hospitals. For every dollar in fraud that is taken out of the system, there is a reduction in healthcare spending that is not patient centered. Moreover, a focus on fraud expands the examination of behaviors beyond financial statement impacts related to organizations, to include asset misappropriation schemes featuring cash and pharmaceutical drugs, which may be more salient to an individual healthcare provider.

While we believe this paper contributes most directly to answering questions in a single economic sector, the findings from the wide data availability within this sector can provide opportunities for generalization to the broader economy. At a minimum, healthcare fraud research provides researchers with another opportunity to make an important contribution to healthcare policy and local economy discussions, since the provision of healthcare services remains an essentially local venture. Current initiatives such as MITRE's Healthcare Anti-Fraud Academic Competition may benefit from the databases and the proposed research questions developed in our study.[11]

[10] If healthcare insurance is critical to the question being investigated, a single state focus may be most appropriate due to the fact that variation in state laws and regulations exists.

[11] MITRE's Healthcare Anti-Fraud Academic Competition seeks to discover talented individuals with innovative solutions to assist government and private healthcare payers in reducing dollars lost to healthcare fraud. MITRE will provide the tools and data to be examined in their initiative. More information on the MITRE initiative can be found at https://www.nhcaa.org/news/mitres-healthcare-anti-fraud-academic-competition.aspx

Appendix A: State Data Repositories with Hospital Financial Statement Data

State	Name of the reports	Website	IS	BS	SE	SCF
Arizona	AFS The Hospital Uniform Accounting Report	http://www.azdhs.gov/preparedness/public-health-statistics/health-facility-cost-reporting/#compiled-health-facility-financial-reporting	X	X	X	X
California	Hospital Annual Financial Disclosure Report	http://www.oshpd.ca.gov/hid/Products/Hospitals/AnnFinanData/DsclsureRpts/index.html	X	X	X	X
Connecticut	AFS HRS Annual Reporting HRS Twelve Months Actual Filing Report	http://www.portal.ct.gov/DPH/Office-of-Health-Care-Access/Health-Data/Hospital-Fillings	X	X	X	
Florida	AFS Financial Data—Hospital	http://www.fha.org/reports-and-resources/templates/health-care-issue/content/reports-and-data/340.aspx	X	X		
Indiana	AFS Hospital Fiscal Reports	http://www.in.gov/isdh/20123.htm	X	X	X	X
Iowa	American Hospital Association Annual Survey	http://www.iowahospitalfacts.com/	X	X*		
Maine	Hospital Financial Information Part III IRS Form 990 Schedule H	https://mhdo.maine.gov/hospital_financials.htm	X	X	X	
Maryland	AFS Annual Report of Revenue, Expenses and Volume	http://www.hscrc.maryland.gov/Pages/hsp_Data2.aspx	X	X	X	X
Minnesota	AFS Hospital Fiscal Year End Data	http://www.health.state.mn.us/divs/hpsc/dap/hccis/setsofdata.htm	X			
Nevada	NHQR Financial Reports	http://nevadacomparecare.net/static-nhqr.php	X	X		
New Jersey	AFS Annual ACH Cost Reports	http://www.nj.gov/health/hcf/reports-data/	X	X	X	
North Carolina	Selected Financial Information from AFS	http://www.ncdhhs.gov/dhsr/ncmcc/	X*		X*	X*
Oregon	AFS Patient Revenue and Unreimbursed Care Form	http://www.oregon.gov/OHA/OHPR/RSCH/Pages/Hospital_Reporting.aspx	X			
Pennsylvania	AFS Financial Analysis Report	http://www.phc4.org/reports/fin/	X			
Rhode Island	Hospital and Hospital Health Care Complex Cost Report	http://health.ri.gov/programs/detail.php?pgm_id=138/	X			
Vermont	Act 53 Hospital Report Card—Financial	http://www.healthvermont.gov/health-statistics-vital-records/health-care-systems-reporting/hospital-report-cards	X			
Washington	Year End Reports	http://www.doh.wa.gov/DataandStatisticalReports/HealthcareinWashington/HospitalandPatientData/HospitalFinancialData.aspx	X	X		

Note: ACH – Acute Care Hospital. AFS – Audited Financial Statements. BS – Balance Sheet availability. HRS – Hospital Reporting System. IS – Income Statement availability. NHQR – National Healthcare Quality Report. SCF – Statement of Cash Flows availability. SE – Statement of Stockholders' Equity availability. X – Data are available for the respective financial statement. * – Only partial data from the statement are available.

Appendix B: State Data Repositories with Hospital Community Benefits Data

State	Name of the reports	Website
California	Hospital Community Benefit Plans	http://www.oshpd.ca.gov/HID/SubmitData/CommunityBenefit/
Connecticut	HRS Annual Reporting HRS Twelve Months Actual Filing Report	http://www.portal.ct.gov/DPH/Office-of-Health-Care-Access/Health-Data/Hospital-Fillings
Florida	Florida Hospitals' Community Benefit Report	http://www.fha.org/reports-and-resources/templates/health-care-issue/content/reports-and-data/340.aspx
Georgia	Hospital Financial Survey	http://www.gamap2care.info/HFS.php
Indiana	Hospital Community Benefit Reports IRS Form 990 Schedule H	http://www.in.gov/isdh/25324.htm
Iowa	Iowa Hospital Association Community Benefit Report	http://www.iowahospitalfacts.com/CommunityBenefits/community_benefits.html
Maine	Hospital Financial Information Part III IRS Form 990 Schedule H	https://mhdo.maine.gov/hospital_financials.htm
Maryland	Community Benefits Data Report	http://hscrc.maryland.gov/Pages/init_cb.aspx
Minnesota	Uncompensated Care Data	http://www.health.state.mn.us/divs/hpsc/dap/hccis/stndrdrpts.htm
Nevada	Nevada Annual Hospital Reporting	https://www.nevadacomparecare.net/nv-reports.php
North Carolina	Hospital Community Benefits Report	https://www.ncha.org/issues/community-benefit
Oregon	Community Benefit Report	http://www.oregon.gov/oha/HPA/ANALYTICS/Pages/Hospital-Reporting.aspx
Rhode Island	Hospital Community Benefits	http://health.ri.gov/programs/detail.php?pgm_id=138/
Vermont	Act 53 Hospital Report Card—Financial	http://gmcboard.vermont.gov/publications/ACT53-HRC-Financial

Note: For a detailed description of mandated reporting requirements for community benefits, visit the Hospital Community Benefit Program website operated by the Hilltop Institute (http://www.hilltopinstitute.org/hcbp.cfm).

References

Agnew, R. (1992). Foundation for a general strain theory of crime and delinquency. *Criminology, 30*(1), 47–88.

Axelrod, D. A., Millman, D., and Abecassis, M. M. (2010). U.S. health care reform and transplantation. Part I: Overview and impact on access and reimbursement in the private sector. *American Journal of Transplantation, 10*(10), 2197–2202.

Bai, G., Jiang, J. X., and Flasher, R. (2017). Hospital risk of data breaches. *JAMA Internal Medicine, 177*(6), 878–880.

Bauder, R. A., and Khoshgoftaar, T. M. (2017). Multivariate outlier detection in Medicare claims payments applying probabilistic programming methods. *Health Services and Outcomes Research Methodology, 17*(3–4), 256–289.

Bauder, R., Khoshgoftaar, T. M., and Seliya, N. (2017). A survey on the state of healthcare upcoding fraud analysis and detection. *Health Services and Outcomes Research Methodology*, *17*(1), 31–55.

Becker's Hospital CFO Report (2014, January 21). 25 hospitals with the highest value-based purchasing scores. Becker's Hospital Review.com. Retrieved from: https://www.beckershospitalreview.com/finance/25-hospitals-with-the-highest-value-based-purchasing-scores.html

Burnaby, P., Howe, M., and Muehlmann, B. W. (2011). Detecting fraud in the organization: An internal audit perspective. *Journal of Forensic & Investigative Accounting*, *3*(1), 195–233.

Busch, R. S. (2012). *Healthcare fraud: Auditing and detection guide*. Hoboken, NJ: John Wiley & Sons.

Carrigan, M. D., and Kujawa, D. (2006). Six Sigma in health care management and strategy. *Health Care Manager, 25*(2), 133–141.

Centers for Medicare and Medicaid Services (CMS). (2015). *New Medicare data available to increase transparency on hospital and physician utilization* [Press release]. Retrieved from: https://www.cms.gov/Newsroom/MediaReleaseDatabase/Press-releases/2015-Press-releases-items/2015-06-01.html

Centers for Medicare and Medicaid Services (CMS). (n.d.a). *The Medicare Access & CHIP Reauthorization Act of 2015: Path to value*. Retrieved from: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/MACRA-MIPS-and-APMs/MACRA-LAN-PPT.pdf

Centers for Medicare and Medicaid Services (CMS). (n.d.b). National health expenditure data: NHE fact sheet. Retrieved from: https://www.cms.gov/research-statistics-data-and-systems/statistics-trends-and-reports/nationalhealthexpenddata/nhe-fact-sheet.html

Chang, J. S. (2015). The Physician Payments Sunshine Act: Data evaluation regarding payments to ophthalmologists. *Ophthalmology*, *122*(4), 656–661.

Clarke, R. V. G., and Felson, M. (Eds.). (1993). *Routine activity and rational choice* (Vol. 5). Piscataway, NJ: Transaction Publishers.

Clemente, S., McGrady, R., Repass, R., Paul, D. P., III, and Coustasse, A. (2017). Medicare and the Affordable Care Act: Fraud control efforts and results. *International Journal of Healthcare Management*. doi:10.1080/20479700.2017.1401277

Cohen, L. E., and Felson, M. (1979). Social change and crime rate trends: A routine activity approach. *American Sociological Review*, *44*(4), 588–608.

Congressional Budget Office. (2017). *The 2017 long-term budget outlook*. Retrieved from: https://www.cbo.gov/system/files/115th-congress-2017-2018/reports/52480-ltbo.pdf

Eldenburg, L., Hermalin, B. E., Weisbach, M. S., and Wosinska, M. (2004). Governance, performance objectives and organizational form: Evidence from hospitals. *Journal of Corporate Finance*, *10*(4), 527–548.

Eldenburg, L., and Kallapur, S. (1997). Changes in hospital service mix and cost allocations in response to changes in Medicare reimbursement schemes. *Journal of Accounting and Economics*, *23*(1), 31–51.

Eldenburg, L., and Krishnan, R. (2007). Management accounting and control in health care: An economics perspective. In C. S. Chapman, A. G. Hopwood, and M. D. Shields (Vol. Eds.) *Handbooks of Management Accounting Research: Vol. 2* (pp. 859–883). Amsterdam, The Netherlands: Elsevier.

Eldenburg, L., Soderstrom, N., Willis, V., and Wu, A. (2010). Behavioral changes following the collaborative development of an accounting information system. *Accounting, Organizations and Society, 35*(2), 222–237.

Eldenburg, L., and Vines, C. C. (2004). Nonprofit classification decisions in response to a change in accounting rules. *Journal of Accounting and Public Policy, 23*(1), 1–22.

Evans J. H., III, Kim, K., Nagarajan, N. J., and Patro, S. (2010). Nonfinancial performance measures and physician compensation. *Journal of Management Accounting Research, 22*(1), 31–56.

Feng, N. C., Ling, Q., Neely, D., and Roberts, A. A. (2014). Using archival data sources to conduct nonprofit accounting research. *Journal of Public Budgeting, Accounting & Financial Management, 26*(3), 458–493.

Flasher, R., and Lamboy-Ruiz, M. A. (2017). Impact of enforcement on healthcare billing fraud: Evidence from the USA. *Journal of Business Ethics.* https://doi.org/10.1007/s10551-017-3650-z

Gee, J., Button, M., and Brooks, G. (2010). *The financial cost of healthcare fraud: What data from around the world shows* [Commissioned report]. Milton Keynes, England: MacIntyre Hudson. Retrieved from: https://researchportal.port.ac.uk/portal/en/publications/the-financial-cost-of-fraud(dbc84015-7b10-4b5b-9265-bc38c21790b8).html

Government Accountability Office (GAO). (2016). *Health care fraud: Information on most common schemes and the likely effect of smart cards*. Report GAO-16-216. Retrieved from: https://www.gao.gov/assets/680/674771.pdf

Greer, A. (2016, May 12). These government agencies receive the most money. Graphiq.com. Retrieved from: http://federal-budget.insidegov.com/stories/13371/government-agencies-most-money

Hannigan, N. S. (2006). Blowing the whistle on healthcare fraud: Should I? *Journal of the American Association of Nurse Practitioners, 18*(11), 512–517.

Heese, J. (2017). The role of overbilling in hospitals' earnings management decisions. *European Accounting Review*. doi:10.1080/09638180.2017.1383168

Hwang, Y., and Kirby, A. J. (1994). Distorted Medicare reimbursements: The effect of cost accounting choices. *Journal of Management Accounting Research, 6*, 128–140.

Ittner, C. D., Larcker, D. F., and Pizzini, M. (2007). Performance-based compensation in member-owned firms: An examination of medical group practices. *Journal of Accounting and Economics, 44*(3), 300–327.

Joudaki, H., Rashidian, A., Minaei-Bidgoli, B., Mahmoodi, M., Geraili, B., Nasiri, M., et al. (2015). Using data mining to detect health care fraud and abuse: A review of literature. *Global Journal of Health Science, 7*(1), 194–202.

Kane, N. M., and Magnus, S. A. (2001). The Medicare cost report and the limits of hospital accountability: Improving financial accounting data. *Journal of Health Politics, Policy and Law, 26*(1), 81–106.

Kang, H., Hong, J., Lee, K., and Kim, S. (2010). The effects of the fraud and abuse enforcement program under the National Health Insurance Program in Korea. *Health Policy, 95*(1), 41–49.

Kim, K. K. (1988). Organizational coordination and performance in hospital accounting information systems: An empirical investigation. *Accounting Review, 63*(3), 472–489.

Krause, J. H. (2003). A conceptual model of health care fraud enforcement. *Journal of Law and Policy, 12*, 55–147.

Krause, J. H. (2004). Healthcare fraud and quality of care: A patient-centered approach. *Journal of Health Law, 37*(2), 161–195.

Krishnan, R., and Yetman, M. H. (2011). Institutional drivers of reporting decisions in nonprofit hospitals. *Journal of Accounting Research, 49*(4), 1001–1039.

Lamboy Ruiz, M. A., No, W. G., & Watanabe, O. V. (2018). Discrepancies in Hospital Financial Information: Comparison of Financial Data in State Data Repositories and the Healthcare Cost Reporting Information System. *Journal of Information Systems (in-press)*. https://doi.org/10.2308/isys-52149

Langton, L., and Piquero, N. L. (2007). Can general strain theory explain white-collar crime? A preliminary investigation of the relationship between strain and select white-collar offenses. *Journal of Criminal Justice*, *35*(1), 1–15.

Lindberg, D. L., and Seifert, D. L. (2016). Microinsurance, fraud, and fraud controls. *Journal of Forensic & Investigative Accounting, 8*(1), 29–35.

Liou, F.-M., Tang, Y.-C., and Chen, J.-Y. (2008). Detecting hospital fraud and claim abuse through diabetic outpatient services. *Health Care Management Science*, *11*(4), 353–358.

Liu, Q., and Vasarhelyi, M. (2013, November). *Healthcare fraud detection: A survey and a clustering model incorporating geo-location information.* Paper presented at the 29th World Continuous Auditing and Reporting Symposium (29WCARS), Brisbane, Australia.

Lynch, A., Bryant, S., and Reck, J. (2011). Fraudulent financial reporting: An update on SEC investigations. *Journal of Forensic & Investigative Accounting, 3*(1), 116–148.

National Health Care Anti-Fraud Association (NHCAA). (2017). The challenge of health care fraud. Retrieved from: https://www.nhcaa.org/resources/health-care-anti-fraud-resources/the-challenge-of-health-care-fraud.aspx

Pizzini, M. J. (2006). The relation between cost-system design, managers' evaluations of the relevance and usefulness of cost data, and financial performance: An empirical study of US hospitals. *Accounting*, *Organizations and Society*, *31*(2), 179–210.

Pizzini, M. (2010). Group-based compensation in professional service firms: An empirical analysis of medical group practices. *The Accounting Review*, *85*(1), 343–380.

Rudman, W. J., Eberhardt, J. S., Pierce, W., and Hart-Hester, S. (2009). Healthcare fraud and abuse. *Perspectives in Health Information Management*, *6*(Fall). Retrieved from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2804462/

SA Ignite Inc. (2016). 10 FAQs about the merit-based incentive payment system (MIPS). Retrieved from: http://www.saignite.com/resources/faq-about-merit-based-incentive-payment-mips

Sing, M., Banthin, J. S., Selden, T. M., Cowan, C. A., and Keehan, S. P. (2006). Reconciling medical expenditure estimates from the MEPS and NHEA, 2002. *Health Care Financing Review*, *28*(1), 25–40.

Taitsman, J. K. (2011). Educating physicians to prevent fraud, waste, and abuse. *New England Journal of Medicine, 364*(2), 102–103.

Tennyson, S. (2008). Moral, social, and economic dimensions of insurance claims fraud. *Social Research, 75*(4), 1181–1204.

Thornton, D., Mueller, R. M., Schoutsen, P., and van Hillegersberg, J. (2013). Predicting healthcare fraud in Medicaid: A multidimensional data model and analysis techniques for fraud detection. *Procedia Technology*, *9*, 1252–1264.

Travaille, P., Mueller, R., Thornton, D., and van Hillegersberg, J. (2011). Electronic fraud detection in the US Medicaid healthcare program: Lessons learned from other industries. In *Proceedings of the 17th Americas Conference on Information Systems, AMCIS*. Atlanta, GA: Association for Information Systems. Retrieved from: http://aisel.aisnet.org/amcis2011_submissions/328

Tripathi, M. (2012). EHR evolution: Policy and legislation forces changing the EHR. *Journal of AHIMA, 83*(10), 24–29.

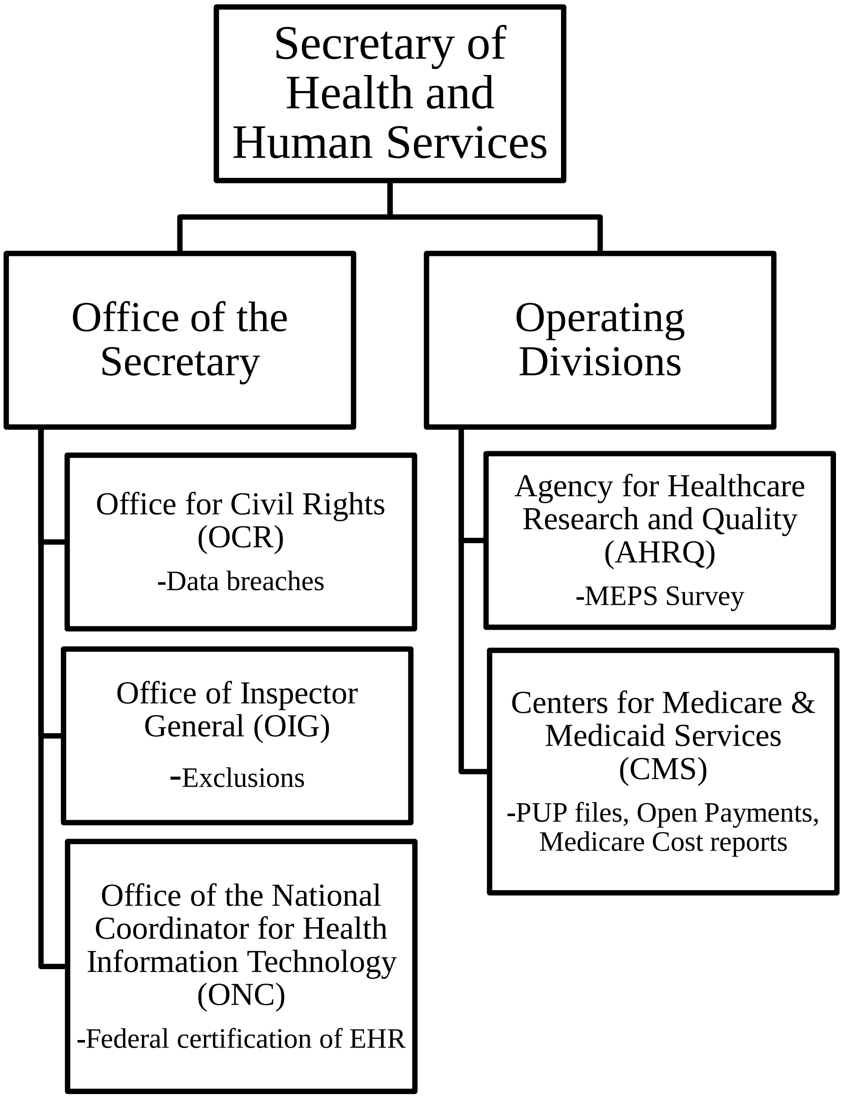
van Capelleveen, G., Poel, M., Mueller, R. M., Thornton, D., and van Hillegersberg, J. (2016). Outlier detection in healthcare fraud: A case study in the Medicaid dental domain. *International Journal of Accounting Information Systems*, *21*, 18–31.

Wang, J., and Yang, J. G. (2009). Data mining techniques for auditing attest function and fraud detection. *Journal of Forensic & Investigative Accounting*, *1*(1), 1–24.

Wang T., Wang, Y., and McLeod, A. (2018) Do health information technology investments impact hospital financial performance and productivity? *International Journal of Accounting Information Systems*, *18*, 1–13.

Yang, W.-S., and Hwang, S.-Y. (2006). A process-mining framework for the detection of healthcare fraud and abuse. *Expert Systems with Applications, 31*(1), 56–68.

Figure 1: U.S. Department of Health and Human Services Organizational Chart



Note: In this abbreviated version, we include only the offices and divisions that oversee the data sources discussed in this paper. A full organizational chart of HHS is available at https://www.hhs.gov/about/agencies/orgchart/index.html#

Table 1: Recent Healthcare Fraud Research Using U.S. Data

Panel A. Reviews of Literature in U.S. Healthcare Fraud

First Author and Year	*Fraud Topic*
Clemente et al. (2017)	Fraud control efforts
Joudaki et al. (2015)	Data-mining techniques
Travaille et al. (2011)	Electronic fraud detection

Panel B. Recent Archival U.S. Healthcare Fraud Research

First Author and Year	*Data Source(s) Used*	*Summary of Fraud Findings*	*Proposed Extension*
van Capelleveen (2016)	Medicaid Management Information System; Medicaid Statistical Information System; reference files from several states, the CMS, and other governmental agencies	Seventeen out of 369 dental providers were flagged from the analyses for further investigation, with twelve of the seventeen being confirmed as worthy of investigation by experts	Other providers beyond dentists
Bauder [...Seliya] (2017)	Medicare Provider Utilization and Payment Data: Physician and Other Supplier Public Use File CY 2012, 2013, 2014; HHS-OIG Exclusion List	Discussed a new data-mining technique application to healthcare and compared results to prior research for Florida-based providers within seven select specializations	Other geographic areas or areas of specialization
Flasher (2017)	HHS-OIG Exclusion List	MFCUs in each state can be associated with a detection effect on excluded individuals associated with program related offenses	Other types of offenses beyond billing fraud; clusters or connections among the excluded individuals
Bai (2017)	HHS Breach Data	Large teaching hospitals are experiencing breaches more so than smaller nonteaching hospitals	Determinants and consequences of breaches for healthcare entities beyond hospitals; reporting and detection challenges for all entities handling healthcare data
Heese (2017)	California Office of Statewide Health Planning and Development	Findings suggest that overbilling is an important alternative manipulation tool of earnings in for-profit hospitals	Examining not-for-profit hospitals, which cover at least sixty percent of the U.S. population of hospitals

Note: Panel A lists recent reviews of the literature on healthcare fraud. Panel B summarizes recent empirical archival healthcare fraud studies that used U.S. data sources described or alluded to in this article. This table is not meant to be comprehensive but is meant instead to introduce the reader to some of the recent literature, with suggestions for how that research could be extended.

Table 2: Healthcare Databases by Topic

Topic	Level	Agency	Website
Community Benefits ($)	State	Various state departments of health	Links listed in Appendix B
Community Benefits Reporting Laws	State	The Hilltop Institute	http://www.hilltopinstitute.org/hcbp.cfm
Quality Indicators Reporting	Federal	CMS	http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetHomepage&cid=1120143435363
Cost Accounting	Federal	CMS	https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/Cost-Reports/Cost-Reports-by-Fiscal-Year.html
Electronic Health Records	Federal	ONC	http://dashboard.healthit.gov/datadashboard/data.php
Exclusions from HHS-OIG	Federal	OIG	https://oig.hhs.gov/exclusions/exclusions_list.asp
Financial Statements	Federal	CMS	https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/Cost-Reports/Cost-Reports-by-Fiscal-Year.html
Healthcare Information Breaches	Federal	OCR	https://ocrportal.hhs.gov/ocr/breach/breach_report.jsf
Healthcare Providers (characteristics)	Federal	CMS	https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/Provider-of-Services/
Patient Payments for Healthcare	Federal	AHRQ	https://meps.ahrq.gov/mepsweb/data_stats/download_data_files.jsp
Payments to Physicians	Federal	CMS	https://www.cms.gov/openpayments/
Physician Quality Indicators	Federal	CMS	https://data.medicare.gov/data/physician-compare
Utilization of Services	Federal	CMS	https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/medicare-provider-charge-data/index.html

Note: AHRQ – Agency for Healthcare Research and Quality. CMS – Centers for Medicare and Medicaid Services. OCR – Office for Civil Rights. OIG – Office of Inspector General. ONC – Office of the National Coordinator for Health Information Technology.