Research Report

JONATHAN S. LEVIN, DANIEL J. CRESPIN, JIN KIM, RACHEL O. REID, CHRISTOPHER M. WHALEY, MICHAEL DWORSKY

Methodologies to Measure Upcoding in Provider Settings

A Scoping Review

For more information on this publication, visit **www.rand.org/t/RRA2683-1**.

Published by the RAND Corporation, Santa Monica, Calif.

About This Report

Increases in payment for more-complicated patients can incentivize upcoding, a practice in which hospitals code more secondary diagnoses or complications to classify admissions or visits at higher complexity levels. However, the lack of standardized methodologies to measure upcoding hinders researchers and policymakers from fully understanding its prevalence, the medical conditions especially likely to be upcoded, the types of hospitals and geographic areas with high upcoding rates, and the overall impact of upcoding on health care spending. To address this gap, the authors of this report provide a review of methodologies to measure upcoding using literature from 2000 to 2023. Using seven electronic databases, the authors identified 39 articles that measure upcoding, which they classified into the following categories: variations in payments (n = 14 studies), variations in provider and patient attributes (n = 6), prediction algorithms (n = 6), self-reports (n = 3), and validations (n = 10). In most of these studies (n = 26), researchers examine upcoding at aggregated levels (e.g., frequency of upcoding at hospitals or in geographic areas) rather than identifying individually upcoded admissions (n = 13). Only three studies measure upcoding by comparing coding intensity with a gold standard, while the rest examine upcoding indirectly via measures of coding intensity. In their conclusion, the authors recommend that researchers and policymakers designing studies on provider-based upcoding consider outcome measures based on severity levels, payment changes that incentivize upcoding, provider and patient characteristics associated with upcoding, and indicators in administrative claims or electronic health record data.

This research was funded by the National Institute for Health Care Reform and carried out within the Payment, Cost, and Coverage Program in RAND Health Care.

RAND Health Care, a division of the RAND Corporation, promotes healthier societies by improving health care systems in the United States and other countries. We do this by providing health care decisionmakers, practitioners, and consumers with actionable, rigorous, objective evidence to support their most complex decisions. For more information, see www.rand.org/health-care, or contact

RAND Health Care Communications
1776 Main Street
P.O. Box 2138
Santa Monica, CA 90407-2138
(310) 393-0411, ext. 7775
RAND_Health-Care@rand.org

Acknowledgments

We would like to thank Jody Larkin for assistance with developing our search strategy. We also thank Nina Ryan and Jasmin Kelly for their administrative assistance. Finally, we would like to thank our quality assurance reviewers for their input on a draft of the report: Robert Berenson, Justin Timbie, Paul Koegel, and Christine Eibner.

Contents

Figure and Tables

Figure

Tables

Chapter 1

Introduction

Approximately $1.1 trillion was spent on hospital care in the United States in 2021, representing more than one-third of total health expenditures (Centers for Medicare & Medicaid Services, [CMS] 2023b). One of the major contributors of the high spending on hospital care is the costs of hospital inpatient admissions, which are, on average, $2,883 per patient per day (Kaiser Family Foundation, 2023). As a solution to control spending from overuse of services, Medicare Part A, which is the largest single health care payer, has used a diagnostic-based prospective payment system to compensate hospitals for inpatient care since 1983. Private insurers in the United States, as well as health care programs in other countries, have since adopted similar payment approaches (Massi, Ieva, and Lettieri, 2020; Chien, Lee, and Wang, 2019; Weber et al., 2021; Cooper et al., 2019).

Under most currently used diagnostic-based payment systems, hospitals receive a fixed payment for a given patient's admission based on their principal diagnosis and complexity level determined by secondary diagnoses at the time of admission or that develop during the stay (e.g., sepsis) (Office of Inspector General, 2001). The increase in payment for more-complicated patients can incentivize *upcoding*, in which hospitals code more secondary diagnoses or complications so that admissions or visits are coded at higher complexity levels. The scope of upcoding varies among policymakers and researchers. Much of the empirical research literature describes upcoding broadly, inclusive of both coding to accurately capture the patient's health status and fraudulent or abusive practices (Cook and Averett, 2020; Silverman and Skinner, 2004; Jürges and Köberlein, 2015; Dafny and Dranove, 2009). Some of the literature refers broadly to upcoding as increases in "coding intensity" (Geruso and Layton, 2020, p. 986; Cook and Averett, 2020, p. 12). Federal agencies, on the other hand, describe upcoding as solely a fraudulent or abusive practice, when admissions or visits are improperly coded to a higher complexity level (CMS, 2021; Office of Inspector General, undated). Across both definitions, upcoding contributes to higher health care spending, which translates to higher cost-sharing for patients; higher premiums for patients and employers; and, in the case of Medicare Part A, higher taxes (Coustasse et al., 2021). Because this report covers both academic and government studies, we used the broad definition of upcoding that is inclusive of accurate and improper increases in coding intensity.

Numerous studies have attempted to identify and quantify upcoding in the United States and other countries. However, despite the interest among researchers and policymakers, there is no standardized methodology used to measure upcoding. For this reason, researchers and policymakers have a limited understanding of the prevalence of upcoding, the medical conditions especially likely to be upcoded, the types of hospitals and geographic areas with high rates of upcoding, and the aggregate impact of upcoding on health care spending.

Policymakers need to provide appropriate oversight and regulation to ensure that upcoding does not financially harm patients, health plan enrollees, and taxpayers who fund the provision of health care. A 2021 report from the Office of Inspector General suggests that upcoding increased between 2014 and 2019 and recommends targeted audits of upcoding (Office of Inspector General, 2021). The report further recommends that these audits be based on stays that are billed at the highest severity level that also have characteristics that may suggest a lower intensity of care. However, more information is needed to identify the characteristics of stays, patients, providers, hospitals, and regions that may be indicative of upcoding. Researchers are especially in need of a centralized review of the methodologies that have been used to measure upcoding to better inform policymakers where it is occurring. In light of this gap, we sought to provide a scoping review of the methodologies used to measure upcoding in provider settings. We intend for our findings in this review to enhance researchers' and auditors' capacities to measure upcoding.

Chapter 2

Methods

Scope and Terminology

In this report, we primarily focus on reviewing papers that implemented methodologies used to measure upcoding of hospital inpatient admissions by measuring increases in patient complexity. We also consider methodologies regarding upcoding of outpatient and office-based visits if payment is determined by complexity or severity, as this may be applicable to determining how to measure upcoding of hospital admissions. Our scope of upcoding spans health care providers broadly, encompassing organizations, facilities, and individual clinicians. Excluded from our study are other types of improper coding by providers—for example, miscoding across patient care settings, such as classifying observation stays as inpatient admissions. It also excludes upcoding practices by health plans and Medicare Advantage Organizations, as this type of upcoding is based on an entire year of diagnoses rather than an individual admission or visit (Geruso and Layton, 2020).

We define *upcoding* as the coding of a patient to a higher complexity level than they would be if payment were unrelated to complexity. To identify the full set of studies on this topic, this definition covers both increases in coding intensity that accurately captures both a patient's complexity and fraudulent and abusive increases in coding intensity. In our summaries of the included studies, we consistently use the term *upcoding* to align with each respective study's definition of the term, even in cases in which the articles had a narrower scope of upcoding than our broader definition.

Much of our review concerns upcoding within Diagnosis Related Groups (DRGs). Under a DRG reimbursement system, hospitals assign patients to a base DRG, typically defined by a principal diagnosis (e.g., heart failure and shock), and then assign an individual DRG based on the presence (or lack) of complications or comorbidities. Hospitals receive higher payments for more-complex patients. Since 1983, Medicare Part A has paid for hospital admissions under a DRG system, which has undergone several major revisions. The most recent major change occurred in 2007, when Medicare restructured the system into the Medicare Severity Diagnosis Related Group (MS-DRG) system. The MS-DRG system uses secondary diagnoses, currently coded using the International Classification of Diseases, Tenth Revision (ICD-10), to determine severity levels, which are typically divided into three categories: with major complications and comorbidities (MCC), with complications and comorbidities (CC), or without CC/MCC. Each individual MS-DRG is assigned a relative weight that reflects its average resource use among Medicare patients, which is multiplied by the hospital's *base payment rate* (a factor that accounts for local costs of labor and nonlabor health care resources, among other aspects[1]) to determine payment (CMS, 2023a). DRGs are also frequently used as a basis for payment

[1] The base payment also considers whether the hospital is a teaching hospital, whether it is in a rural area, and whether it cares for a disproportionate share of the poor and uninsured population.

among private payers and other public health insurance programs, including from other countries, although the definitions and classifications of DRGs are not necessarily the same as those used by Medicare (Weber et al., 2021; Cooper et al., 2019).

Search Strategy

We searched for peer-reviewed articles, government reports, and gray literature from January 1, 2000, through February 27, 2023, that were written in English. The search terms used to identify articles within this scope are listed in Appendix A. We did not place any restrictions on country of publication. We included articles that identify individual admissions or visits that are suspected to be upcoded, as well as articles that estimate upcoding or code intensity at more-aggregate levels (e.g., frequency of upcoding at hospitals or in geographic areas).

We implemented our search strategy in the following databases: Academic Search Complete, Business Source Complete, Embase, Policy File Index, PubMed, Scopus, and Web of Science. Searches were not limited to specific fields in PubMed, but for other databases they were limited to the title, abstract, or subject and/or keyword fields. Our reporting of the search results follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidance. Two team members screened the title and abstract of each article. For situations in which an article's inclusion was indeterminable from the title and abstract, team members reviewed the full article. Articles deemed to meet inclusion criteria were marked for data abstraction. Studies meeting the following criteria were eligible for inclusion:

- **Article type and time frame:** peer-reviewed reports, government reports, or gray literature published between January 1, 2000, and February 27, 2023
- **Setting:** health care providers, including health systems, hospitals, and physician practices
- **Original investigation:** study implemented primary research and was not an editorial or review article.

After completion of article screening, the study team manually searched for government reports in the United States during the same time period from the U.S. Government Accountability Office, Office of Inspector General, and MedPAC. We did not undertake additional searches for government reports in other countries.

Data Abstraction and Synthesis

For each study, we abstracted specific information on the methodology used to measure upcoding. Initially, we determined each study's upcoding measure, which could vary from flagging individual claims as upcoded to population estimates of upcoding or coding intensity in a geographic area or at a specific hospital or physician practice. We also determined the specific technique used to calculate or estimate each upcoding measure, such as the identification of a certain indicator on a claim (e.g., a specific secondary diagnosis or an outlier length of stay [LOS]) or a statistical analysis (e.g., multivariate regression). We also abstracted the following study characteristics from each article in our sample: title, authors, publication year, research objective(s), therapeutic condition(s), geographic setting, provider setting, study period, study population, sample size, data source, study design, key findings, and notable limitations. After the first round of abstractions, we used commonalities among various studies to classify them into the following categories based on their methods:

- **Variations in payments:** Authors investigate coding intensity incentivized by external changes in policies and payment amounts—for example, the introduction of new severity ratings in the MS-DRG system and the annual recalibrating of payment weights for MS-DRGs. These changes can result in higher relative payments for MS-DRGs with a higher complexity compared with an MS-DRG with lower complexity, which incentivizes providers to upcode patients into the higher-complexity MS-DRG. These studies typically model coding intensity by estimating the change in the proportion of admissions among a given base MS-DRG assigned to the highest-severity MS-DRG. A statistically significant association between increased relative payment and the frequency of the highest severity MS-DRG would indicate that upcoding is occurring.
- **Variations in provider and patient attributes:** These studies examine how different characteristics of providers and their patient populations may be associated with upcoding. The authors of these studies hypothesize that certain provider characteristics, such as for-profit status or electronic health record (EHR) adoption, may incentivize upcoding. A statistically significant association between these provider characteristics and the frequency of codes indicating severity would indicate that upcoding is occurring.
- **Prediction algorithms:** Articles in this category implement algorithms that identify individual inpatient admission (or other types of visits) that have been upcoded or hospitals with a high rate of upcoding. These algorithms are based on a set of variables, usually from administrative claims data, that are then incorporated into a cluster analysis or another type of psychometric method to identify outliers in the distribution of variables.
- **Self-reports:** In these studies, authors identified upcoding through surveys or interviews with hospital staff or providers.
- **Validations:** In these studies, authors investigate whether measures of coding intensity are supported by indicators in claims, health records, or other datasets. A study of this type could estimate, for example, the association between coding intensity and LOS, readmission rates, or a comorbidity index. These studies determine whether trends in coding intensity are because of changes in health need versus changes in upcoding.

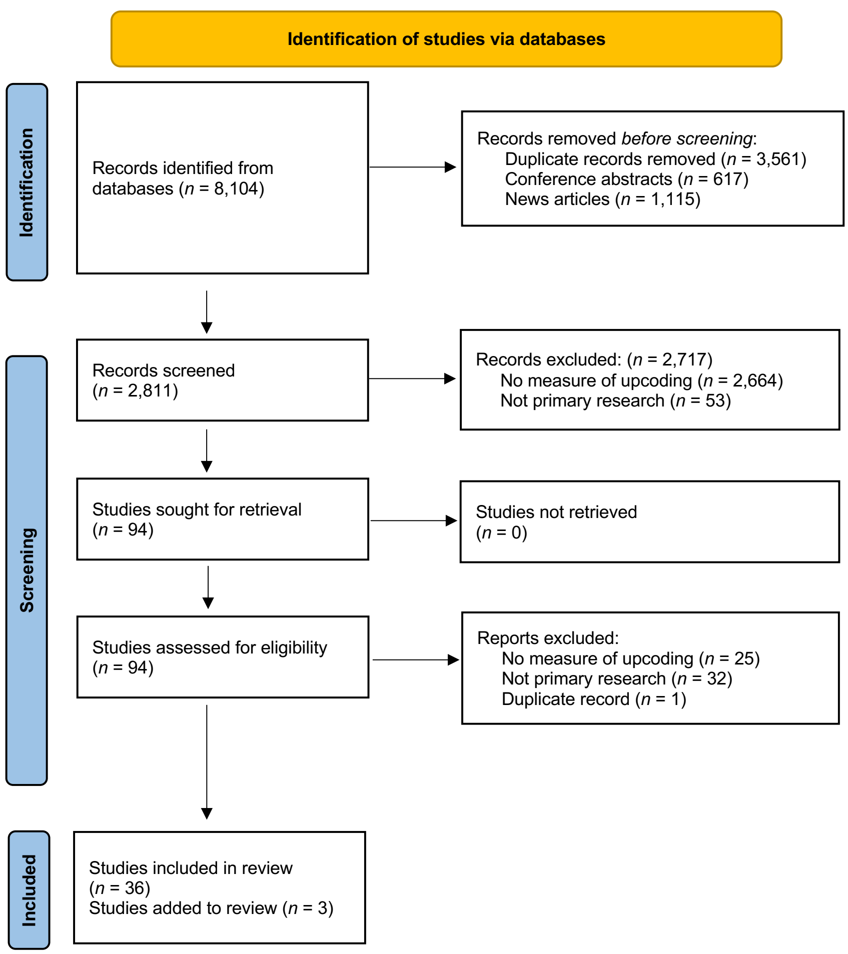
Classifications were not mutually exclusive, meaning some articles could contain multiple methods to measure upcoding. In those cases, we documented all types of upcoding methods specified in the article. We then summarized the methods, strengths, and limitations of the articles by category.

Chapter 3

Results

Of the 8,104 articles initially identified, 3,561 were duplicates, 617 were conference abstracts, and 1,115 were news articles. Upon review of titles and abstracts for the remaining 2,811 articles, 2,664 were excluded for not measuring upcoding and 53 for not involving primary research. Of the remaining 94 articles for which the full text was reviewed, 57 were excluded for the following reasons: 25 for not measuring provider upcoding, 32 for not being primary research, and one for being a duplicate. We then added three articles that the study team flagged as relevant despite not being captured by the initial search terms, amounting to 39 total articles. See Figure 3.1.

Figure 3.1. PRISMA Flow Diagram



Summary Statistics of Sample

Table 3.1 summarizes the characteristics of the included studies that measured upcoding in a provider setting. The majority of the articles included were published after 2015 (62 percent), and most articles were published in the United States (59 percent). The vast majority of articles used administrative claims as their main data sources; only two studies used health records. The full list of articles can be found in Appendix B.

Table 3.1. Characteristics of Included Studies

Characteristic	*N*
Total	39
Year of publication	
2000–2004	5
2005–2009	4
2010–2014	6
2015–2019	13
2020–2022	11
Location	
United States	23
Europe	14
Asia	2
Main data source	
Administrative claims data	25
Health records	2
Surveys/interviews	6
Other	6

NOTE: There were no studies from 2023.

In Table 3.2, we classify studies based on the method used and whether the study investigated upcoding at an aggregate level or for individual admissions or visits. Two-thirds of studies investigated upcoding by determining the level of coding intensity within a particular geographic area, hospital or health system, and/or provider. Of the 13 studies that investigated upcoding for individual admissions or visits, most (seven studies) did so by validating the accuracy of a coding intensity measure compared with specific indicators of severity of a patient's admission, usually from administrative claims data.

Table 3.2. Methodologies of Upcoding Studies

Upcoding Methodology	Aggregated Level (*N* = 26)	Individual Level (*N* = 13)	Total Articles (*N* = 39)
Variations in payments	12	2	**14**
Variations in provider and patient attributes	6	0	**6**
Validations	3	7	**10**
Prediction algorithms	2	4	**6**
Self-reports	3	0	**3**

Variations in Payments

Methods

Fourteen studies inferred that upcoding was occurring based on a statistical association between variations in payments and coding intensity rates. We consider *variation* to be the implementation of a new reimbursement policy, annual recalibrating of reimbursement weights, and/or adjustments to reimbursements based on geography. Nine of these studies examined upcoding for hospital inpatient care, three examined outpatient care, and two examined preterm birth newborns in Germany.

Of the nine studies that examined inpatient care, five examined the implementation of new reimbursement policies while the other four focused on upcoding resulting from annual changes in DRG payment weights. The earliest of the nine studies that examined inpatient care, Dafny in 2005, established the methodology that most subsequent articles followed. Dafny used changes in the relative payment weights between DRG pairs with the same main diagnosis, but differentiated by the presence of complications, to estimate the frequency with which hospitals coded a patient to the DRG with complications. The difference in payment weights between the DRG with complications and the DRG without complications is referred to as *spread*. The study was completed in the context of a policy change in the United States in 1988 that removed age categories in DRG weights, which effectively increased the spread of a given DRG pair, which Dafny hypothesized would increase the incentive to upcode. Dafny implemented this model at both the DRG-year level and the DRG-hospital-year level, using administrative claims data on inpatient admissions (Dafny, 2005). Six other studies, including all four that examined changes in DRG weights (or the equivalent), used the same overall modeling setup as the Dafny article (Barros and Braun, 2017; Januleviciute et al., 2016; Milcent, 2021; Anthun, Bjørngaard, and Magnussen, 2017; Anthun, 2022; Cook and Averett, 2020). All studies used repeated cross-sectional analyses, with consistent findings that increases in spread were associated with increases in the frequency of the DRG with the highest complexity.

Two other studies, out of the nine on inpatient care, examined the association of provider groups that had differences in spread and coding intensity. Thus, these studies estimated the impact of being part of a group with higher spread rather than the impact of spread on intensity. The first study—on increases in spread for birth delivery DRGs in Italy[2]—found that hospitals that experienced the largest increase in spread had a 2 percentage point increase in the probability of coding a vaginal delivery with complications (compared with deliveries with no complications) (Di Giacomo et al., 2017). The second study, by Bystrov and colleagues, compared the number of stroke admissions with complications among hospitals in Poland with stroke units, which received greater reimbursements for complications, to admissions in hospitals without stroke units. The greater reimbursements for hospitals with stroke units were established when Poland implemented its version of DRGs in mid-2008. By the end of the study period (2009–2013), the authors estimated that the diffusion of hospitals with stroke units was associated with a 37 percent increase in the number of stroke admissions classified into higher-severity levels (Bystrov et al., 2015).

Two studies examined differences in reimbursement rates in Germany based on various low-birth weight intervals. The first study used German birth statistics and the fee schedule from the Institute for the Hospital Remuneration System (InEK) from 1996 to 2010, finding that increases in payment differentials were associated with discontinuities of preterm babies assigned to lower birthweight categories, which increased reimbursements to providers (Jürges and Köberlein, 2015). The other study implemented a controlled simulation in which participants were randomized to different risks of audits and fines for reporting fraudulent birth weights that would increase reimbursements. In this case, participants were provided with information on a fictional newborn's birthweight and then had to report the weight interval of this weight to collect reimbursement. The study found that, coupled with a fine for fraudulent reporting, increases in audit probabilities of from 10 percent to 75 percent decreased the rate of detectable upcoding by 17 percentage points. For the audits in this simulation, the researchers considered *detectible upcoding* to be when there was a misreporting of at least 100 grams between the true birthweight and the reported weight interval, suggesting that this type of upcoding is inappropriate (Groß, Jürges, and Wiesen, 2021).

Finally, three studies modeled intensity of non-DRG codes, two of which examined variation in reimbursements over both time and geographic location. The first of these two studies examined variation in payments by the Medicare Physician Fee Schedule (MPFS) for Current Procedural Terminology (CPT) codes of general health assessments in Medicare Part B office-based visits, ranging from a problem-focused visit for a guideline time of ten minutes, reimbursed at the lowest level, to a comprehensive visit at 40 minutes and reimbursed at the highest level. Through analyzing Part B claims, Brunt (2011) estimated that a 5 percent increase in reimbursement differentials for a higher-intensity CPT code relative to a lower one was associated with 35 percent higher odds of billing to the higher-intensity CPT code. The second study on the MPFS, published in 2014, examined variation in average Medicare Part B reimbursements of relative value units (RVUs) among counties, finding that a decrease in average reimbursement of 10 percent in RVU payments was associated with increases of higher-intensity CPT codes between 1 and 6 percent and decreases in lower-intensity codes of between 6 and 10 percent (Brunt, 2015). The rationale for this finding is that providers may offset lower reimbursement rates by increasing the intensity of the CPT codes that were reported. The third study was a cross-sectional analysis, which estimated variation in Medicare reimbursements by locality and number of therapy minutes at skilled nursing facilities (SNFs). This

[2] Studies in countries outside the United States use the term *DRG*; however, the exact definitions, diagnosis categories, and severity ratings are not standardized across countries.

study found that SNFs in regions with greater payment differentials between the highest and lowest resource utilization groups billed for a higher proportion of services with a higher number of therapy minutes (Bowblis and Brunt, 2014).

Strengths and Limitations

The studies using this method have strong internal validity and are generalizable. By relying on exogenous changes to influence upcoding, these studies minimize concerns that the changes in the outcome could be attributed to unobserved variables associated with certain types of hospitals. Additionally, these exogenous changes typically affect all hospitals or health systems, thus allowing generalization of research results across geographic areas and hospitals. Given the annual recalibration of weights in the MS-DRG system, there are ample opportunities for researchers and auditors to continue to use this method. The main limitation of this method is that it does not directly identify individual admissions or visits that are inappropriately upcoded. However, by understanding differences in coding intensity, these methods could be leveraged to determine the types of providers likely to upcode and the types of medical conditions with a high probability of being upcoded. These methods also do not identify the level or total amount of upcoding; rather, they identify differences in coding intensity associated with changes in the financial rewards of upcoding.

Variations in Provider and Patient Attributes

Methods

Six studies examined differences in coding intensity between the characteristics of providers and patients. Two of these studies examined EHR adoption, with contrasting results. The first, by Adler-Milstein and Jha, compared hospitals by whether they had adopted EHRs between 2008 and 2010, finding that EHR adoption was not associated with changes in the proportion of DRGs billed to the highest severity rating (Adler-Milstein and Jha, 2014). By contrast, in a study by Webb and colleagues, the implementation of EHRs at a single academic medical center was associated with a 3.3 percent increase in CPT codes at the top intensity level for the general internal medicine department and a 44.7 percent increase in the top intensity level for the orthopedics department (Webb et al., 2017).

Two studies examined the association of hospital or health system characteristics and coding intensity based on for-profit status or level of specialization. In comparing for-profit with nonprofit hospitals, Pruitt and Pracht (2013) found that for-profit hospitals, which are hypothesized to have a greater incentive to upcode, had 10 percent higher odds of coding pediatric admissions for inconsequential injuries as an "emergency"[3] (which has the potential to increase revenues compared with coding as "urgent" or "elective"). In an evaluation of a policy in Taiwan that encouraged

[3] According to the study, an emergency classification helps avoid prior authorization requirements, which can increase revenues, and is exempt from a $1,500 cap associated with procedures in the emergency department. The authors note that the cap is only for adults, but they speculate that this may affect the hospitals' overall coding procedures when classifying an admission as an emergency.

controlling patient volume among tertiary centers, Liang and colleagues found that providers responded to this policy by decreasing their patient volume but increasing their coding of patients at the highest intensity level by 73 percent for individuals with hypertension and 46 percent for individuals with diabetes. By contrast, patients in clinics (their comparison group) experienced declines in intensity levels during the same study period (Liang et al., 2020).

Researchers have also examined how differences in payers influence the potential for upcoding. Private payers may create greater incentives to upcode because they have higher average reimbursement rates than public payers, as well as higher reimbursements for patients who are rated as high risk (versus low risk) of having complications when using outpatient care, such as by the American Society of Anesthesiologists (ASA) scores when providing anesthesia. One study, by Dragan and colleagues, found that hospitals with a larger share of private patients had a higher number of diagnoses per Medicaid claim, suggestive of upcoding behavior that has spilled over to Medicaid (Dragan et al., 2022), whereas another study, Schonberger, Dutton, and Dai (2016), found no difference in ASA scores between patients who were 65 years old (and therefore Medicare eligible based on age) versus age 64 (and just below Medicare eligibility based on age), suggesting that upcoding was not occurring for this population.

Strengths and Limitations

These studies offer important insights on the types of clinicians, provider groups, hospitals, and health systems that, according to the authors of these studies, may experience greater incentives to upcode, such as for-profit hospitals and hospitals with a larger share of privately insured patients. However, the statistical associations examined in these studies should not be interpreted causally; despite these studies controlling for observable characteristics of providers and using sensitivity analyses with different specifications, they are ultimately comparing providers across different settings that could be associated with the outcomes measured.

Prediction Algorithms

Methods

The authors of six studies implemented a prediction algorithm to identify or classify admissions or visits that were upcoded. Table 3.3 provides a description of the types of indicators that these studies used. Two of the six studies that used prediction algorithms did so at the hospital level, both using multivariable models. The first, by Aelvoet and colleagues in 2009, identified hospitals in Belgium suspected of fraud using a multitude of metrics from claims data, including LOS, frequencies of stays billed to a given DRG severity level, and a list of 21 diagnosis codes. Using these variables, the study team calculated the following measures for each hospital: (1) *distance* between observed and expected mean LOS,[4] with expected LOS based on DRG, severity of illness (SOI), and age; (2) odds ratio of the proportion of the lowest SOI class of DRGs billed for a hospital compared with the proportion of

[4] This discrepancy was calculated using the method of Mahalanobis distance (Brereton, 2015).

national average; (3) odds ratio of the proportion of the DRGs without a principal diagnosis for a hospital to the national average; (4) odds ratio of stays with a surgery unrelated to the principal diagnosis for a hospital to the national average; and (5) odds ratios of reimbursement outliers among given SOI classes to the national average.[5] These values (i.e., distance or odds ratios) were then summed into one score, where a score of greater than one was used as a threshold to determine suspicion of fraud. The authors then evaluated outcomes of audits using this algorithm compared with a random audit, finding that the fraud-directed audit increased the number of identified upcoded inpatient stays by 29 percent (Aelvoet et al., 2009).

Table 3.3. Variables in Claims Data from Predictive Algorithm Studies

Categories	Variables from Claims Data
Coding intensity	DRG, severity rating, principal diagnosis versus secondary diagnosis for a given condition
Procedures	Surgery unrelated to principal diagnosis, elective operations and/or admissions, procedures used during inpatient admission
Costs	Total costs of an admission
Patient characteristics	Age, race, ethnicity, sex, insurance plan, region, urbanicity
Length of stay	Number of days for a given admission
Outcome measures	30-day readmission rates, 30-day mortality rates, postoperative complications, discharge status
Comorbidities	Charlson Comorbidity Index (Charlson et al., 2022), hierarchical condition categories (HCCs), indicators for chronic conditions (e.g., hypertension, diabetes, end-stage renal disease), number of total diagnoses

SOURCE: The list of variables is from the six studies that used a predictive algorithm methodology (Aelvoet, 2009; Massi, Ieva, and Lettieri, 2020; Gani, Canner, and Pawlik, 2018; Chien, Lee, and Wang, 2019; Rosenberg, Fryback, and Katz, 2000; Souza et al., 2020).

The second study used a *k*-means cluster analysis to identify clusters of hospitals in Italy that may have higher rates of upcoding (Everitt et al., 2011). Using administrative claims data on hospitalizations for heart failure, Massi and colleagues generated six clusters of hospitals based on average and total comorbidities of their patient populations, average and total costs for heart failure hospitalizations, readmission rates, average age of patients, and indices that measured average coding intensity and comorbidities. The authors then identified hospitals as outliers if they were at the 95th percentile distribution of the distance to the centroid of their respective cluster. Using this method, the authors identified ten (out of 183) hospitals they considered to be outliers. In their exploration of

[5] In Belgium, the prospective payment system allows for exceptions to increase reimbursements for patients with exceptionally high costs. Thus, the study also considers *downcoding*, in which an admission may be assigned a lower severity level in order to be considered an outlier that would warrant an increase in reimbursements. The authors do not define the threshold they use for outliers in reimbursement amounts.

three of these hospitals, they concluded that one of them appeared to be justified in their measures of billing intensity, whereas the other two were flagged as warranting further investigation to determine if upcoding had occurred (Massi, Ieva, and Lettieri, 2020).

Researchers for four articles used algorithms to identify upcoding for individual admissions, of which two used multivariate models. The first, by Gani, Canner, and Pawlik in 2018, was the only study to do so based on a single variable, which was the number of diagnosis codes per admission for gastrointestinal surgery. Admissions with a greater-than-average number of codes (in this case, nine or more) represented upcoded admissions. The authors found that upcoded admissions increased by 133 percent from 2001 to 2011 and that upcoded admissions were more likely to be emergent and/or urgent (versus elective) and among those with a higher Charlson Comorbidity Index. To identify potentially inappropriate cases of upcoding, the authors designed a cohort of "low-risk" patients based on those without preexisting comorbidities, those undergoing an elective operation, and those who did not develop a postoperative complication during their inpatient admission. Among this cohort, upcoded admissions increased by 65 percent during the study period (Gani, Canner, and Pawlik, 2018).

In a 2019 study, Chien, Lee, and Wang applied a Rasch model to a set of 17 continuous variables, each representing spending for prescription drugs, medical materials, or examinations.[6] The threshold for outliers was based on whether a given spending variable was two standard deviations above the mean in a random sample of 300 cases from 20 DRGs. The authors then compared these outlier thresholds with a set of 194 cases that Taiwan's Bureau of National Health Insurance (BNHI) had previously deducted medical fees from, implying that these cases were inappropriately upcoded. Using this threshold, the authors would have identified 82 percent of the cases that the BNHI had deduced (Chien, Lee, and Wang, 2019).

The third study, by Rosenberg, Fryback, and Katz in 2000 designed a hierarchical Bayesian model to predict the probability that an inpatient stay was incorrectly coded. Using private insurance claims data, the authors predicted an admission as upcoded based on the coding intensity of the inpatient stay, the major diagnostic category, patient age, sex, LOS, the log of total payments, an indicator of whether this service was "medical" versus "surgical" or "maternity," and diagnoses beyond the major diagnostic category. In applying this method to claims already identified with inaccurate DRG codes, the authors identified 89 percent of these claims using their model (Rosenberg, Fryback, and Katz, 2000).

The fourth study, by Souza and colleagues in 2020, used a support vector machine—a classification method using machine learning—to predict assignment of a DRG and a severity level for a given admission for a respiratory disease based on the following variables in claims data: diagnoses, inpatient procedures, discharge status, sex, and age. The authors then performed sensitivity analyses by removing each variable from the models to determine whether that would change the classification. Where dropping a single variable moved the admission to a lower severity level, the authors flagged this admission as potentially upcoded. Using this method, they estimated approximately 3 percent of admissions were upcoded (Souza et al., 2020).

[6] For information on the Rasch model, see Bond, Yan, and Heene, 2021.

Strengths and Limitations

Studies using algorithms hold promise to directly estimate the extent of upcoding rather than the change in upcoding or coding intensity to which other methods are limited. They also are encouraging in identifying individual admissions that may be upcoded, especially given the high predictive value of the algorithms in Chien, Lee, and Wang (2019) and in Rosenberg, Fryback, and Katz (2000). However, unlike other methods for measuring upcoding, there has yet to be an established or consistent algorithm used across these studies, which makes it challenging for future researchers who are interested in using this methodology. Many of these studies also do not have identification strategies, as they are instead relying on outliers in a distribution. Moreover, these algorithms may require advanced computational skills in machine learning, which may present a barrier to feasibly implementing this method. Finally, all the algorithms in this sample of articles used variables from claims data and thus are subject to the same limitations of possible unobserved characteristics of admissions or hospitals that may be predictive of upcoding.

Self-Reports

Methods

Three studies identified upcoding based on self-reported information, of which two were from hospital staff. In a mixed methods study of a hospital in Indonesia, Palutturi and colleagues (2019) interviewed hospital staff, who then reported descriptions of billing activities that are consistent with upcoding behavior. Likewise, in a 1999 survey among hospital information managers in the United States, approximately 44 percent of respondents reported that they "often" or "sometimes" promote coding optimization, suggesting that they are engaging in upcoding behavior (Lorence and Spink, 2002).

The third study reported upcoding rates among individual providers. In a survey administered to physicians in the United States in 1998, Wynia and colleagues asked respondents for the frequency with which they implemented billing practices that were synonymous with upcoding. They found that 39 percent of respondents reported that they at least "sometimes" exaggerated the severity of their patients' conditions, manipulated their patients' billing diagnoses, or reported conditions that the patient did not have to help their patients receive coverage for their care. The authors also implemented analyses comparing physicians who "sometimes" use these upcoding practices with physicians who "never" or "rarely" used them to identify differences in attitudes on and pressures to achieve quality care that may explain why some providers upcode (Wynia et al., 2000).

Strengths and Limitations

A notable strength of these studies is that surveys of providers can estimate the full extent of upcoding rather than upcoding at the margin resulting from an exogenous change. However, honest reporting and self-awareness of providers with regard to their coding behavior is a major concern with this method. Given that these concerns would bias the estimates downward, self-reported data may offer a lower-bound estimate of upcoding rates among providers. Furthermore, even precise estimates of the prevalence of upcoding do not provide evidence about the associated costs to payers.

Validations

Methods

Of the ten studies using validation methods, three of them did so at an aggregated level. The first study, by Silverman and Skinner (2004), estimated statistically significant increases in the frequencies of DRGs billed at the highest reimbursement levels between 1989 and 1996 in the United States; the greatest increases occurred among hospitals that converted from nonprofit to for-profit status (37 percent), and the smallest increases occurred among stable nonprofit and government hospitals (approximately 10 percent). The authors then compared this trend with annual measures of mortality and the Charlson index of comorbidities. They observed declines in mortality rates and increases in the Charlson index, although these increases did not differ by the profit status of hospitals. Given these trends in measures used as proxies for SOI, the authors concluded that the increases in severity levels billed among DRGs in for-profit hospitals are likely attributed to upcoding of inpatient admissions (Silverman and Skinner, 2004). Another study has since replicated this approach of comparing annual trends in coding intensity with trends in LOS (Office of Inspector General, 2021). The third study, by Chernew and colleagues (2021), compared trends in annual risk score growth of HCCs, generated through Medicare Parts A and B claims data, with trends in risk score growth measured by the Consumer Assessment of Healthcare Providers and Systems (CAHPS) survey data, which are self-reported by patients.[7] Between 2013 and 2016, HCC scores increased by 6.3 percent, while CAHPS risk scores increased by 0.8 percent. The discrepancy between these two estimates suggests that upcoding of diagnosis codes may have been occurring (Chernew et al., 2021). Although HCC scores are not directly tied to provider payments from Medicare, this increase in HCC scores may reflect provider upcoding of diagnosis codes and severity levels.

Seven studies used a validation method to determine whether upcoding was occurring among individual admissions or visits. Four of these studies did so by modeling the predicted probabilities of being rated with the highest-severity DRG, the highest-intensity evaluation and management (E&M) CPT code, or a high risk score for anesthesia based on a set of covariates of patient characteristics from claims data (Burke et al., 2018; Seiber, 2007; Willer et al., 2021; Nie et al., 2016). The objectives of these studies were largely to examine changes in these severity ratings over time; the authors concluded that increases in the predicted severity ratings after controlling for these covariates were the result of upcoding. These covariates include a broad variety of patient characteristics, such as patient demographics (i.e., age, race, ethnicity, sex), comorbidities (both as binary and count variables), LOS, whether the patient was readmitted within 30 days, whether the patient died within 30 days, insurance plan, region, urbanicity, and provider setting.

[7] These risk scores were generated by the study team using linear regression models based on either claims data or CAHPS survey data. The variables for the claims-based risk score consisted of age, sex, race/ethnicity, dual-eligible status, and diagnoses, whereas variables for the CAHPS-based risk score were age, sex, race/ethnicity, education level, self-reported mental health, self-reported general health, and functional limitations.

Three other studies assessed criterion validity by a comparison against a gold standard. The first study, by Allen and colleagues (2003), reviewed medical records to determine the appropriate E&M visit billing codes based on the patient history, body exam areas, diagnoses and treatment options, and amount and complexity of data. After comparing their results with those of the clinicians who completed these records and ordered the E&M billing codes, the authors determined that four of the ten health records reviewed should have been billed to lower-intensity E&M codes, while two should have been billed to higher-intensity E&M codes (Allen et al., 2003). The second study, a 2014 report by the Office of Inspector General, used similar methods of comparing medical records with E&M visits in Medicare Part B claims in 2010, finding that 26 percent of claims were upcoded and 15 percent of claims were downcoded (Office of Inspector General, 2014). The third study, by Hurst, compared two sets of codes submitted by nurses for the same patients, one of which was assumed by the authors to be more accurate than the other (Hurst, 2009).

Strengths and Limitations

Similar to algorithms and self-reported methods, studies using "validation" methods may allow for identifying a broader extent of upcoding than upcoding at the margin. They also help rule out alternative explanations for studies identifying upcoding associated with provider characteristics. The main downside of the majority of these studies is that they assume that all types of patient acuity and severity are based on observable characteristics in claims data. Only three studies compared coding in claims data with a gold standard, making it challenging to draw implications.

Chapter 4

Discussion

Overview

A lack of a standardized methodology to measure upcoding of hospital admissions constrains researchers' and policymakers' capacity to understand the prevalence of upcoding. This, in turn, limits their ability to quantify the amount by which upcoding may be increasing health care spending, particularly for patients, health plan enrollees, and taxpayers. Our scoping review of methodologies used to measure provider upcoding found that, because of the limitations of administrative data, nearly all studies measured upcoding indirectly rather than identifying specific admissions or visits that were upcoded. By contrast, only three studies measured upcoding directly, by comparing coding intensity with a gold standard. The studies that measured upcoding indirectly were generally at aggregated levels and restricted to identifying differences in coding intensity among providers driven by differences in payment or hospital attributes, such as for-profit ownership. These studies are often statistically strong (i.e., unbiased) and helpful in identifying the aggregate impact of a policy change on upcoding and types of providers likely to upcode, but they rely on measures that may be too broad to inform stakeholders in need of precise information about how to recognize upcoding at the claim level. A smaller, second set of studies developed methods that identify individually upcoded admissions, using algorithms, or that validate coding intensity measures against specific indicators of diagnoses or utilization; however, not all researchers may have access to the types of data used in these analyses. By considering the strengths and weaknesses of both sets of studies, researchers can enhance measures of upcoding and coding intensity, particularly for researchers restricted to administrative data.

Most studies in our review measured upcoding at an aggregated level, using measures of coding intensity and estimated differences in coding intensity associated with changes in payment and/or provider and patient characteristics. Several studies replicated the same analytic approach that was initially designed by Dafny (2005), which identifies increases in upcoding driven by changes in the relative weights of DRGs. Compared with studies that estimated associations between coding intensity and provider or patient characteristics, which are arguably the weakest form of evidence, studies relying on payment changes are less biased because they can more accurately attribute changes in upcoding to a specific policy change that affected all hospitals. A further weakness of studies that indirectly measure upcoding using these methods is that they are generally limited to estimating increases (or decreases) in coding intensity over time rather than the level of upcoding (e.g., the exact number of hospital admissions or visits upcoding in a specific year).

Studies that identify upcoding of individual admissions or visits, or that validate measures of upcoding, incorporate additional information from claims and medical records. Although researchers may lack access to some of these data, they may be able to incorporate some of the knowledge and lessons learned from these studies to either enhance the measures of upcoding used in their analyses or provide supporting information pertaining to why the measures they selected are valid. Studies that used validation methods mostly compared coding intensity in claims data with other variables in claims data that were indicative of severity, such as LOS and readmissions rates. The most common analytic approach to validating coding intensity rates was modeling whether differences in coding intensities could be explained by differences in variables indicative of severity. A less common approach, though more internally valid, was to compare coding intensity with a gold standard, such as medical records. Studies using predictive algorithms each had a different analytic approach, suggesting that there is a need for additional research that can either validate these current algorithms or design new ones that have validity and feasibility for use by other researchers. The limitations of these studies may explain, in part, why they are less common in the literature.

Incorporating Guidance from the Literature on Measuring Upcoding

Below, we present some guidance on how researchers and policymakers who are examining provider-based upcoding can incorporate the information learned in this review.

1. **Outcome measures based on severity levels.** The vast majority of studies measured outcomes based on frequencies of codes billed at the highest severity level of a patient's admission or visit. Severity levels that are strictly based on an evaluation of the patient's health condition, such as the MS-DRG system, rather than severity levels that are tied to procedures or duration of care, such as CPT codes that capture different visit lengths, are more appropriate for measuring whether upcoding is occurring. When examining health services in settings that do not have severity-adjusted prospective payments, researchers should consider whether there are other outcome variables in administrative claims data that may influence reimbursements. These ratings may be associated with certain differences in service use but could be identified based on whether there are breaks in the distribution of these ratings. Some of these include assessments of anesthetic risk scores in outpatient services, number of therapy minutes by SNFs, and number of diagnoses.
2. **Payment changes that incentivize upcoding.** Studies that identified statistical associations between exogenous changes and outcome measures offer greater validity in detecting upcoding. These exogenous changes take the form of policy changes that influence reimbursement rates to providers, such as the introduction or removal of certain DRGs and severity levels, as well as annual recalibrating of weights that influence "spread." Researchers evaluating these exogenous changes should consider whether their sample of providers are all subjected to the policy changes. If there is a potential comparison group of providers unaffected by the policy change, then researchers should consider whether the unaffected providers have characteristics that may be systematically different from those who were affected by the policy. However, payment changes may not exist in all settings where upcoding may be occurring, warranting researchers to use other methodologies.
3. **Provider and patient characteristics associated with upcoding.** Should there be a lack of exogenous changes for a given population or scope of health services, researchers could

consider identifying characteristics among providers, their patient populations, and their regions that may encourage them to upcode. The identification of these characteristics should be based on documented differences in reimbursement rates among various providers. Given the multitude of other variables that may explain potential associations between provider characteristics and outcome measures, researchers pursuing this approach should adjust for patient and provider demographics and implement numerous sensitivity analyses to ensure the robustness of their results.

4. **Indicators in administrative claims or EHR data.** Claims data offer rich information on patient health and sociodemographic information that could explain some of the variation in outcome measures that are indicative of upcoding. Some of the variables to consider are comorbidity indexes, binary indicators for specific diagnoses, elective versus urgent operations, development of postoperative complications, LOS, readmissions, and death within 30 days of admission. Researchers should control for these variables when designing predictive algorithms or validating variation or trends in coding intensity. To the extent that there is unexplained variation in coding intensity after controlling for these other variables, then upcoding may be a plausible explanation.

   Researchers attempting to identify individual admissions that may be upcoded could identify outliers that have high severity ratings, or a high number of diagnosis codes, but low ratings for these other variables that are indicative of severity of condition (e.g., below-average LOS). Researchers should consider using a multitude of variables that are indicative of severity for a given condition in consultation with a clinician. Predictive algorithms, such as cluster analyses that classify admissions into categories based on divergence of coding intensity with these other variables, hold promise in identifying upcoded visits; however, there has yet to be a standard method replicated by other researchers on this topic. When policymakers are selecting variables in claims data to identify upcoded admissions, they should also consider whether some of the variables indicative of a lower severity of condition, such as a low LOS, may also be reflective of a higher quality of care.

   EHRs offer a complementary set of information to claims data that may further suggest where upcoding is occurring. As implemented by Allen and colleagues (2003) and the Office of Inspector General (2014), researchers could compare severity ratings in administrative claims data with the details of diagnoses, procedures, and screenings ordered for patients. To the extent that records are complete in their documentation, objective determinations of whether the information in EHR data is consistent with the reported severity ratings in claims data may offer the ability to estimate upcoding that avoids the internal validity concerns of many claims-based studies. Policymakers may be in a unique position to access such data through interagency data-sharing agreements; otherwise, access to and costs of linked EHR-administrative claims data may be a considerable barrier to implementing these studies, especially among academic researchers.

Limitations

There are several limitations in this review. Articles published prior to 2000 and in languages other than English were not included. Although we captured more than 8,000 records, it is possible that we were unsuccessful in identifying other relevant studies, especially in the gray literature and government reports. We attempted to account for this by additionally searching for government reports among several U.S. agencies during the same time period, but we did not extend this supplemental search to other countries. Our sample of articles also disproportionately draws on methodologies that examined DRGs for inpatient stays (far fewer studies examined upcoding in office-based settings) and does not include miscoding of observations as inpatient stays. Although we offer our own assessment of the general strengths and limitations of articles across each methods category, we do not assign individual assessments to each article. Finally, given the risk of publication bias, it is possible that other methods for upcoding detection have been used but were not published because they did not generate statistically significant results.

Conclusions

This scoping review is the first to examine the methods used to measure provider upcoding. The majority of studies detected upcoding indirectly via measures of coding intensity rather than by comparing coding intensity with a gold standard. Most studies also examined upcoding at aggregated levels and based on changes in upcoding resulting from changes in payments rather than the full extent of upcoding within a given locality. Future studies should consider the outcome measures of coding intensity that are based strictly on patient health, payment changes that incentivize upcoding, provider and patient characteristics associated with upcoding, and indicators in datasets suggestive of upcoding.

Appendix A

Literature Review Search Parameters

The following are the search parameters for our literature review:

- **Search executed:** February 27, 2023
- **Databases**
  - Academic Search Complete
  - Business Source Complete
  - Embase
  - Policy File Index
  - PubMed
  - Scopus
  - Web of Science

- **Language:** English
- **Publication date:** 2000 to February 27, 2023
- **Search terms** (not limited to specific fields in PubMed; limited to title, abstract, or subject and/or keywords fields in the other databases):

  "up-cod*" OR "upcod*" OR "unbundl*" OR ("fraud*" AND "health*") OR ("fraud*" AND "provider*") OR "code creep" OR "coding practice*"

  AND

  "charg*" OR "bill*" OR "ICD" OR "CPT" OR "DRG" OR "MS-DRG*" OR "diagnos*" OR "RAF" OR "HCC" OR "risk-adjustment factor*" OR "maxim*" OR "optim*" OR "intens*"

Table A.1 lists the databases that we used in our search and the number of relevant results pre-deduplication and post-deduplication.

Table A.1. Search Databases and Number of Relevant Results

Database	Number of Results Pre-Deduplication	Number of Results Post-Deduplication
Academic Search Complete	1,330	585
Business Source Complete	1,742	1,617
Embase	834	269
Policy File Index	100	82
PubMed	927	921
Scopus	2,052	954
Web of Science	1,119	115

Appendix B

Studies Included in the Scoping Review

Table B.1 lists the articles used in our scoping review that measured upcoding in a provider setting.

Table B.1. Study Articles and Characteristics

Author	Title	Classification of Upcoding Measure	Granularity of Identification	Geographic Setting	Study Period	Study Population	Sample Size	Data Source
Adler-Milstein and Jha, 2014	"No Evidence Found That Hospitals Are Using New Electronic Health Records to Increase Medicare Reimbursements"	Variations in provider and patient attributes	Hospitals or health systems	United States	2008–2010	Medicare enrollees with inpatient admissions to hospitals in the highest quartile of proportion of Medicare admissions	1,175 hospitals	CMS impact files; Medicare Provider Analysis and Review files
Aelvoet et al., 2009	"Miscoding: A Threat to the Hospital Care System. How to Detect It?"	Prediction algorithms	Hospitals or health systems	Belgium	2001–2003	Entire population	11,743,945 discharges	Belgian Minimal Clinical Database
Allen et al., 2003	"Nurse Practitioner Coding Practices in Primary Care: A Retrospective Chart Review"	Validations	Admissions or visits	United States	1999	Primary care patients	10 charts	Medical records

Author	Title	Classification of Upcoding Measure	Granularity of Identification	Geographic Setting	Study Period	Study Population	Sample Size	Data Source
Anthun, 2022	"Predicting Diagnostic Coding in Hospitals: Individual Level Effects of Price Incentives"	Variations in payments	Admissions or visits	Norway	1999–2012	Entire population	11,065,330 discharges	Norwegian Patient Registry
Anthun, Bjørngaard, and Magnussen, 2017	"Economic Incentives and Diagnostic Coding in a Public Health Care System"	Variations in payments	Hospitals or health systems	Norway	1999–2008	Entire population	3,180,578 discharges	Norwegian Patient Registry
Barros and Braun, 2017	"Upcoding in a National Health Service: The Evidence from Portugal"	Variations in payments	Hospitals or health systems	Portugal	2001–2008	Entire population	54,593 discharges	Hospital discharge data from Central Administration of the Health System, Portugal
Bowblis and Brunt, 2014	"Medicare Skilled Nursing Facility Reimbursement and Upcoding"	Variations in payments	Hospitals or health systems	United States	2005	Medicare enrollees in SNFs with hip fractures and strokes	95,975 hip fracture patients; 147,858 stroke patients	Minimum Data Set
Brunt, 2011	"CPT Fee Differentials and Visit Upcoding Under Medicare Part B"	Variations in payments	County	United States	2001–2003	Medicare enrollees	16,000 Medicare enrollees	Medicare Current Beneficiary Survey
Brunt, 2015	"Medicare Part B Intensity and Volume Offset"	Variations in payments	County	United States	2002–2005	Medicare enrollees	6,336 outpatient visits	Medicare Current Beneficiary Survey

Author	Title	Classification of Upcoding Measure	Granularity of Identification	Geographic Setting	Study Period	Study Population	Sample Size	Data Source
Burke et al., 2018	“Are Trends in Billing for High-Intensity Emergency Care Explained by Changes in Services Provided in the Emergency Department? An Observational Study Among US Medicare Beneficiaries”	Validations	Admissions or visits	United States	2006, 2009, 2012	Medicare enrollees	1,883,650 emergency department visits	Medicare fee-for-service claims
Bystrov et al., 2015	“Effects of DRG-Based Hospital Payment in Poland on Treatment of Patients with Stroke”	Variations in payments	Hospitals or health systems	Poland	2010	Stroke patients	45,776 admissions	Narodowy Fundusz Zdrowia
Chernew et al., 2021	“Coding-Driven Changes in Measured Risk in Accountable Care Organizations”	Validations	National	United States	2013–2016	Medicare enrollees	76,801 Medicare enrollees during the baseline year	Medicare CAHPS
Chien, Lee, and Wang, 2019	“Detecting Hospital Behaviors of Up-Coding on DRGs Using Rasch Model of Continuous Variables and online Cloud Computing in Taiwan”	Prediction algorithms	Admissions or visits	Taiwan	2015–2016	Entire population	300 discharges	Taiwan’s BNHI

Author	Title	Classification of Upcoding Measure	Granularity of Identification	Geographic Setting	Study Period	Study Population	Sample Size	Data Source
Cook and Averett, 2020	"Do Hospitals Respond to Changing Incentive Structures? Evidence from Medicare's 2007 DRG Restructuring"	Variations in payments	Hospitals or health systems	United States	2005–2010	Entire population	Approximately 7 million discharges	National Inpatient Survey data from Healthcare Cost and Utilization Project
Dafny, 2005	"How Do Hospitals Respond to Price Changes?"	Variations in payments	Hospitals or health systems	United States	1985–1991	Medicare enrollees	650 DRG pair years; 36,651 hospital years	Medicare Provider Analysis and Review files
Di Giacomo et al., 2017	"Do Public Hospitals Respond to Changes in DRG Price Regulation? The Case of Birth Deliveries in the Italian NHS"	Variations in payment	Hospitals or health systems	Italy	2007–2012	Pregnant women	216,760 deliveries	Certificate of Delivery Assistance registry
Dragan et al., 2022	"Association of Insurance Mix and Diagnostic Coding Practices in New York State Hospitals"	Variations in provider and patient attributes	Hospitals or health systems	United States	2010–2014	Medicaid enrollees with 2 or more hospitalizations across 2 or more facilities	1,614,630 hospitalizations	Statewide Planning and Research Cooperative System and New York State Medicaid
Gani, Canner, and Pawlik, 2018	"Assessing Coding Practices for Gastrointestinal Surgery over Time in the United States"	Prediction algorithms	Admissions or visits	United States	2001–2011	Adult patients undergoing a gastrointestinal operation	1,344,152 admissions	National Inpatient Sample

Author	Title	Classification of Upcoding Measure	Granularity of Identification	Geographic Setting	Study Period	Study Population	Sample Size	Data Source
Groß, Jürges, and Wiesen, 2021	“The Effects of Audits and Fines on Upcoding in Neonatology”	Variations in payments	Individual providers	Germany	2019	Infants	197 participants	Primary data collection
Hurst, 2009	“Gaming and Up-Coding”	Validations	Admissions or visits	United Kingdom	2003-2004	Entire population	12,800 patients (2,500 English adults; 10,300 Scottish children)	English and Scottish hospital wards
Januleviciute et al., 2016	“How Do Hospitals Respond to Price Changes? Evidence from Norway”	Variations in payments	Hospitals or health systems	Norway	2003–2007	Entire population	63 hospitals	Norwegian Patient Register
Jürges and Köberlein, 2015	“What Explains DRG Upcoding in Neonatology? The Roles of Financial Incentives and Infant Health”	Variations in payments	National	Germany	1996–2011	Preterm infants with low birth weight	748,000 birth weights	German birth statistics; fee schedule from the InEK
Liang et al., 2020	“Physician Code Creep After the Initiation of Outpatient Volume Control Program and Implications for Appropriate ICD-10-CM Coding”	Variations in provider and patient attributes	Individual providers	Taiwan	2016–2017	Individuals with hypertension and diabetes	1,946,347 individuals with hypertension; 957,333 individuals with diabetes	Outpatient claims data
Lorence and Spink, 2002	“Regional Variation in Medical Systems Data: Influences on Upcoding”	Self-reports	Hospitals or health systems	United States	1998–1999	Health information managers	16,591 health information managers	Self-reported surveys sent out by the Foundation of Record Education

Author	Title	Classification of Upcoding Measure	Granularity of Identification	Geographic Setting	Study Period	Study Population	Sample Size	Data Source
Massi, Ieva, and Lettieri, 2020	"Data Mining Application to Healthcare Fraud Detection: A Two-Step Unsupervised Clustering Method for Outlier Detection with Administrative Databases"	Prediction algorithms	Hospitals or health systems	Italy	2013–2015	Entire population	396,246 discharges	Hospital discharge chart
Milcent, 2021	"From Downcoding to Upcoding: DRG Based Payment in Hospitals"	Variations in payments	Hospitals or health systems	France	2006–2011	Entire population	145 million discharges	French hospital discharge database
Nie et al., 2016	"Upcoding and Anesthesia Risk in Outpatient Gastrointestinal Endoscopy Procedures"	Validations	Admissions or visits	United States	2005–2013	Gastrointestinal anesthesia service claims of patients between 18 and 65 years old	1,001,841 claims	Truven MarketScan
Office of Inspector General, 2014	*Improper Payments for Evaluation and Management Services Cost Medicare Billions in 2010*	Validations	Admissions or visits	United States	2010	Medicare Part B enrollees	657 claims	Medicare Part B claims linked to medical records
Office of Inspector General, 2021	"Trend Toward More Expensive Inpatient Hospital Stays in Medicare Emerged Before COVID-19 and Warrants Further Scrutiny"	Validations	National	United States	2014–2019	Medicare Part A enrollees	8.7 million discharges in 2019	Medicare Part A claims

Author	Title	Classification of Upcoding Measure	Granularity of Identification	Geographic Setting	Study Period	Study Population	Sample Size	Data Source
Palutturi et al., 2019	"Potential for Fraud of Health Service Claims to BPJS Health at Tenriawaru Public Hospital, Bone Regency, Indonesia"	Self-reports	Hospitals or health systems	Indonesia	2016	Entire population (patients at Tenriawaru Public Hospital of Bone Regency)	15 interviewees	Interviews from Tenriawaru Public Hospital
Pruitt and Pracht, 2013	"Upcoding Emergency Admissions for Non-Life-Threatening Injuries to Children"	Variations in provider and patient attributes	Hospitals or health systems	Florida	2001–2010	Children 15 years and younger, excluding newborns and life-threatening injuries	8,694 patients	Florida inpatient discharge data
Rosenberg, Fryback, and Katz, 2000	"A Statistical Model to Detect DRG Upcoding"	Prediction algorithms	Admissions or visits	United States	1989–1998	Claims audited by Blue Cross Blue Shield of Michigan	23,248 claims	Blue Cross Blue Shield of Michigan
Schonberger, Dutton, and Dai, 2016	"Is There Evidence for Systematic Upcoding of ASA Physical Status Coincident with Payer Incentives? A Regression Discontinuity Analysis of the National Anesthesia Clinical Outcomes Registry"	Variations in provider and patient attributes	National	United States	2014	Hip, femur, or lower-leg fracture cases requiring anesthesia	49,850 claims	National Anesthesia Clinical Outcomes Registry
Seiber, 2007	"Physician Code Creep: Evidence in Medicaid and State Employee Health Insurance Billing"	Validations	Admissions or visits	United States	2001–2003	Medicaid enrollees	204,945 office visits	South Carolina Medicaid and the State Employees Health Plan

Author	Title	Classification of Upcoding Measure	Granularity of Identification	Geographic Setting	Study Period	Study Population	Sample Size	Data Source
Silverman and Skinner, 2004	"Medicare Upcoding and Hospital Ownership"	Validations	Hospitals or health systems	United States	1989–1997	Medicare enrollees with pneumonia or respiratory complications diagnosis	5,000 hospitals	20% sample of Medicare inpatient claims
Souza et al., 2020	"Measuring Data Credibility and Medical Coding: A Case Study Using a Nationwide Portuguese Inpatient Database"	Prediction algorithms	Admissions or visits	Portugal	2015	Entire population	487,156 discharges	Portuguese National Hospital Morbidity Database
Webb et al., 2017	"Electronic Health Record Implementation Is Associated with a Negligible Change in Outpatient Volume and Billing"	Variations in provider and patient attributes	Hospitals or health systems	United States	2011–2012	General internal medicine and orthopedics patients	14,395 discharges	Hospital billing data
Willer et al., 2021	"Trends in Hospital Costs and Levels of Services Provided for Children With Bronchiolitis Treated in Children's Hospitals"	Validations	Admissions or visits	United States	2010–2019	Infants	385,883 infants	Pediatric Health Information System
Wynia et al., 2000	"Physician Manipulation of Reimbursement Rules for Patients: Between a Rock and a Hard Place"	Self-reports	Individual providers	United States	1998	Practicing physicians	720 physicians	Self-reported surveys sent out by the National Opinion Research Center

Abbreviations

ASA	American Society of Anesthesiologists
BNHI	Bureau of National Health Insurance
CAHPS	Consumer Assessment of Healthcare Providers and Systems
CMS	Centers for Medicare & Medicaid Services
CPT	Current Procedural Terminology
DRG	Diagnosis Related Group
E&M	evaluation and management
EHR	electronic health record
HCC	hierarchical condition category
ICD-10	International Classification of Diseases, Tenth Revision
InEK	Institute for the Hospital Remuneration System
LOS	length of stay
MS-DRG	Medicare Severity Diagnosis Related Group
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SNF	skilled nursing facility
SOI	severity of illness

References

Adler-Milstein, Julia, and Ashish K. Jha, "No Evidence Found That Hospitals Are Using New Electronic Health Records to Increase Medicare Reimbursements," *Health Affairs (Millwood)*, Vol. 33, No. 7, 2014.

Aelvoet, Willem H., Nathalie Terryn, F. Windey, Martina Redivo, Marc van Sprundel, and Christel Faes, "Miscoding: A Threat to the Hospital Care System. How to Detect It?" *Revue d'Épidémiologie et de Santé*, Vol. 57, No. 3, 2009.

Allen, Kymberly R., Caroline B. Reinke, Joanne M. Pohl, Kristy K. Martyn, and Elaine P. McIntosh, "Nurse Practitioner Coding Practices in Primary Care: A Retrospective Chart Review," *Journal of the American Academy of Nurse Practitioners*, Vol. 15, No. 5, 2003.

Anthun, Kjartan Sarheim, "Predicting Diagnostic Coding in Hospitals: Individual Level Effects of Price Incentives," *International Journal of Health Economics and Management*, Vol. 22, No. 2, 2022.

Anthun, Kjartan Sarheim, Johan H. Bjørngaard, and Jon Magnussen, "Economic Incentives and Diagnostic Coding in a Public Health Care System," *International Journal of Health Economics and Management*, Vol. 17, No. 1, 2017.

Barros, Pedro, and Gisele Braun, "Upcoding in a National Health Service: The Evidence from Portugal," *Health Economics*, Vol. 26, No. 5, 2017.

Bond, Trevor, Zi Yan, and Moritz Heene, *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*, Routledge, 2021.

Bowblis, John R., and Christopher S. Brunt, "Medicare Skilled Nursing Facility Reimbursement and Upcoding," *Health Economics*, Vol. 23, No. 7, 2014.

Brereton, Richard G., "The Mahalanobis Distance and Its Relationship to Principal Component Scores," *Journal of Chemometrics*, Vol. 29, No. 3, 2015.

Brunt, Christopher S., "CPT Fee Differentials and Visit Upcoding Under Medicare Part B," *Health Economics*, Vol. 20, No. 7, 2011.

Brunt, Christopher S., "Medicare Part B Intensity and Volume Offset," *Health Economics*, Vol. 24, No. 8, 2015.

Burke, Laura G., Robert C. Wild, E. John Orav, and Renee Y. Hsia, "Are Trends in Billing for High-Intensity Emergency Care Explained by Changes in Services Provided in the Emergency Department? An Observational Study Among US Medicare Beneficiaries," *BMJ Open*, Vol. 8, No. 1, 2018.

Bystrov, Victor, Anna Staszewska-Bystrova, Daniel Rutkowski, and Tomasz Hermanowski, "Effects of DRG-Based Hospital Payment in Poland on Treatment of Patients with Stroke," *Health Policy*, Vol. 119, No. 8, 2015.

Centers for Medicare & Medicaid Services, *Medicare Fraud & Abuse: Prevent, Detect, Report*, January 2021. As of January 7, 2024:
https://www.cms.gov/Outreach-and-Education/Medicare-Learning-Network-MLN/MLNProducts/Downloads/Fraud-Abuse-MLN4649244.pdf

Centers for Medicare & Medicaid Services, "Acute Inpatient PPS," webpage, last modified September 6, 2023a. As of January 7, 2024:
https://www.cms.gov/medicare/medicare-fee-for-service-payment/acuteinpatientpps

Centers for Medicare & Medicaid Services, "NHE Fact Sheet," webpage, last modified December 13, 2023b. As of January 7, 2024:
https://www.cms.gov/research-statistics-data-and-systems/statistics-trends-and-reports/nationalhealthexpenddata/nhe-fact-sheet

Charlson, Mary E., Danilo Carrozzino, Jenny Guidi, and Chiara Patierno, "Charlson Comorbidity Index: A Critical Review of Clinimetric Properties," *Psychotherapy and Psychosomatics*, Vol. 91, No. 1, 2022.

Chernew, Michael E., Jessica Carichner, Jeron Impreso, J. Michael McWilliams, Thomas G. McGuire, Sartaj Alam, Bruce E. Landon, and Mary Beth Landrum, "Coding-Driven Changes in Measured Risk in Accountable Care Organizations," *Health Affairs (Millwood)*, Vol. 40, No. 12, 2021.

Chien, Tsair-Wei, Yi-Lien Lee, and Hsien-Yi Wang, "Detecting Hospital Behaviors of Up-Coding on DRGs Using Rasch Model of Continuous Variables and Online Cloud Computing in Taiwan," *BMC Health Services Research*, Vol. 19, No. 1, 2019.

CMS—*See* Centers for Medicare & Medicaid Services.

Cook, Amanda, and Susan Averett, "Do Hospitals Respond to Changing Incentive Structures? Evidence from Medicare's 2007 DRG Restructuring," *Journal of Health Economics*, Vol. 73, September 2020.

Cooper, Zack, Stuart V. Craig, Martin Gaynor, and John Van Reenen, "The Price Ain't Right? Hospital Prices and Health Spending on the Privately Insured," *Quarterly Journal of Economics*, Vol. 134, No. 1, 2019.

Coustasse, Alberto, Whitney Layton, Laykin Nelson, and Victoria Walker, "Upcoding Medicare: Is Healthcare Fraud and Abuse Increasing?" *Perspectives in Health Information Management*, Vol. 18, No. 4, 2021.

Dafny, Leemore S., "How Do Hospitals Respond to Price Changes?" *American Economic Review*, Vol. 95, No. 5, 2005.

Dafny, Leemore S., and David Dranove, "Regulatory Exploitation and Management Changes: Upcoding in the Hospital Industry," *Journal of Law & Economics*, Vol. 52, No. 2, 2009.

Di Giacomo, Marina, Massimiliano Piacenza, Luigi Siciliani, and Gilberto Turati, "Do Public Hospitals Respond to Changes in DRG Price Regulation? The Case of Birth Deliveries in the Italian NHS," *Health Economics*, Vol. 26, Suppl. 2, September 2017.

Dragan, Kacie L., Sunita M. Desai, John Billings, and Sherry A. Glied, "Association of Insurance Mix and Diagnostic Coding Practices in New York State Hospitals," *JAMA Health Forum*, Vol. 3, No. 9, 2022.

Everitt, Brian S., Sabine Landau, Morven Leese, and Daniel Stahl, *Cluster Analysis*, 5th ed., Wiley, 2011.

Gani, Faiz, Joseph K. Canner, and Timothy M. Pawlik, "Assessing Coding Practices for Gastrointestinal Surgery over Time in the United States," *Surgery*, Vol. 164, No. 3, 2018.

Geruso, Michael, and Timothy Layton, "Upcoding: Evidence from Medicare on Squishy Risk Adjustment," *Journal of Political Economy*, Vol. 12, No. 3, 2020.

Groß, Mona, Hendrik Jürges, and Daniel Wiesen, "The Effects of Audits and Fines on Upcoding in Neonatology," *Health Economics*, Vol. 30, No. 8, 2021.

Hurst, Keith, "Gaming and Up-Coding," *Nursing Management (Harrow)*, Vol. 15, No. 9, 2009.

Januleviciute, Jurgita, Jan Erik Askildsen, Oddvar Kaarboe, Luigi Siciliani, and Matt Sutton, "How Do Hospitals Respond to Price Changes? Evidence from Norway," *Health Economics*, Vol. 25, No. 5, 2016.

Jürges, Hendrik, and Juliane Köberlein, "What Explains DRG Upcoding in Neonatology? The Roles of Financial Incentives and Infant Health," *Journal of Health Economics*, Vol. 43, September 2015.

Kaiser Family Foundation, "Hospital Adjusted Expenses per Inpatient Day," webpage, circa 2023. As of January 7, 2024:
https://www.kff.org/health-costs/state-indicator/expenses-per-inpatient-day/?currentTimeframe=0&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D

Liang, Fu-Wen, Liang-Yi Wang, Lin-Yi Liu, Chung Yi Li, and Tsung-Hsueh Lu, "Physician Code Creep After the Initiation of Outpatient Volume Control Program and Implications for Appropriate ICD-10-CM Coding," *BMC Health Services Research*, Vol. 20, No. 1, 2020.

Lorence, Daniel P., and Amanda Spink, "Regional Variation in Medical Systems Data: Influences on Upcoding," *Journal of Medical Systems*, Vol. 26, No. 5, 2002.

Massi, Michela C., Francesca Ieva, and Emanuele Lettieri, "Data Mining Application to Healthcare Fraud Detection: A Two-Step Unsupervised Clustering Method for Outlier Detection with Administrative Databases," *BMC Medical Informatics and Decision Making*, Vol. 20, No. 1, 2020.

Milcent, Carine, "From Downcoding to Upcoding: DRG Based Payment in Hospitals," *International Journal of Health Economics Management*, Vol. 21, No. 1, 2021.

Nie, Xiaoyu, Soeren Mattke, Zachary Predmore, and Hangsheng Liu, "Upcoding and Anesthesia Risk in Outpatient Gastrointestinal Endoscopy Procedures," *JAMA Internal Medicine*, Vol. 176, No. 6, 2016.

Office of Inspector General, "Physician Relationships with Payers," webpage, undated.
https://oig.hhs.gov/compliance/physician-education/i-physician-relationships-with-payers/

Office of Inspector General, *Medicare Hospital Prospective Payment System: How DRG Rates Are Calculated and Updated*, August 2001.

Office of Inspector General, *Improper Payments for Evaluation and Management Services Cost Medicare Billions in 2010*, May 2014.

Office of Inspector General, "Trend Toward More Expensive Inpatient Hospital Stays in Medicare Emerged Before COVID-19 and Warrants Further Scrutiny," data brief, February 2021.

Palutturi, Sukri, Siti Rahmawati Makkurade, Reza Aril Ahri, and Ade Suzana Eka Putri, "Potential for Fraud of Health Service Claims to BPJS Health at Tenriawaru Public Hospital, Bone Regency, Indonesia," *International Journal of Innovation, Creativity and Change*, Vol. 8, No. 5, 2019.

Pruitt, Zachary, and Etienne Pracht, "Upcoding Emergency Admissions for Non-Life-Threatening Injuries to Children," *American Journal of Managed Care*, Vol. 19, No. 11, 2013.

Rosenberg, Marjorie A., Dennis G. Fryback, and David A. Katz, "A Statistical Model to Detect DRG Upcoding," *Health Services and Outcomes Research Methodology*, Vol. 1, December 2000.

Schonberger, Robert B., Richard P. Dutton, and Feng Dai, "Is There Evidence for Systematic Upcoding of ASA Physical Status Coincident with Payer Incentives? A Regression Discontinuity Analysis of the National Anesthesia Clinical Outcomes Registry," *Anesthesia and Analgesia*, Vol. 122, No. 1, 2016.

Seiber, Eric E., "Physician Code Creep: Evidence in Medicaid and State Employee Health Insurance Billing," *Health Care Financing Review*, Vol. 28, No. 4, 2007.

Silverman, Elaine, and Jonathan Skinner, "Medicare Upcoding and Hospital Ownership," *Journal of Health Economics*, Vol. 23, No. 2, 2004.

Souza, Julio, Diana Pimenta, Ismael Caballero, and Alberto Freitas, "Measuring Data Credibility and Medical Coding: A Case Study Using a Nationwide Portuguese Inpatient Database," *Software Quality Journal*, Vol. 28, June 2020.

Webb, Matthew L., Daniel D. Bohl, Jennifer M. Fischer, Andre M. Samuel, Adam M. Lukasiewicz, Bryce A. Basques, and Jonathan N. Grauer, "Electronic Health Record Implementation Is Associated with a Negligible Change in Outpatient Volume and Billing," *American Journal of Orthopedics (Belle Mead, N.J.)*, Vol. 46, No. 3, 2017.

Weber, Ellerie, Eric Floyd, Youngran Kim, and Chapin White, "Peering Behind the Veil: Trends in Types of Contracts Between Private Health Plans and Hospitals," *Medical Care Research and Review*, Vol. 78, No. 3, 2021.

Willer, Robert J., Eric R. Coon, Wade N. Harrison, and Shawn L. Ralston, "Trends in Hospital Costs and Levels of Services Provided for Children with Bronchiolitis Treated in Children's Hospitals," *JAMA Network Open*, Vol. 4, No. 10, 2021.

Wynia, Matthew K., Deborah S. Cummins, Jonathan B. VanGeest, and Ira B. Wilson, "Physician Manipulation of Reimbursement Rules for Patients: Between a Rock and a Hard Place," *JAMA*, Vol. 283, No. 14, 2000.