Disease-Specific Trends of Comorbidity Coding and Implications for Risk Adjustment in Hospital Administrative Data

Ulrike Nimptsch

Objective. To investigate changes in comorbidity coding after the introduction of diagnosis related groups (DRGs) based prospective payment and whether trends differ regarding specific comorbidities.

Data Sources. Nationwide administrative data (DRG statistics) from German acute care hospitals from 2005 to 2012.

Study Design. Observational study to analyze trends in comorbidity coding in patients hospitalized for common primary diseases and the effects on comorbidity-related risk of in-hospital death.

Extraction Methods. Comorbidity coding was operationalized by Elixhauser diagnosis groups. The analyses focused on adult patients hospitalized for the primary diseases of heart failure, stroke, and pneumonia, as well as hip fracture.

Principal Findings. When focusing the total frequency of diagnosis groups per record, an increase in depth of coding was observed. Between-hospital variations in depth of coding were present throughout the observation period. Specific comorbidity increases were observed in 15 of the 31 diagnosis groups, and decreases in comorbidity were observed for 11 groups. In patients hospitalized for heart failure, shifts of comorbidity-related risk of in-hospital death occurred in nine diagnosis groups, in which eight groups were directed toward the null.

Conclusions. Comorbidity-adjusted outcomes in longitudinal administrative data analyses may be biased by nonconstant risk over time, changes in completeness of coding, and between-hospital variations in coding. Accounting for such issues is important when the respective observation period coincides with changes in the reimbursement system or other conditions that are likely to alter clinical coding practice.

Key Words. Risk adjustment, administrative data, comorbidity, Elixhauser diagnosis groups, coding practice

Prospective payment systems for hospital reimbursement, such as diagnosis related groups (DRG), require standardized collection of demographic and clinical inpatient information for billing purposes. Administrative hospital

data derived in this manner are commonly used for health services research. In addition, they are also used to support policy decisions, for example, hospital quality assessment. However, because of their administrative purpose, the accuracy of such information is sensitive to financial pressures or other incentives for coding.

Analyses of administrative hospital data often require risk adjustment, for example, to compare outcomes among providers or regions or over time. Important risk adjustment variables are comorbidities that are measured by secondary diagnoses. Because secondary diagnoses are relevant payment factors in DRG-based payment systems, their completeness and accuracy might change over time if incentives for coding change. From a longitudinal perspective, such temporal inconsistency can bias outcome measures, such as risk-adjusted mortality, when secondary diagnoses are used for risk adjustment (Nicholl 2007; Mohammed et al. 2009; Jaques et al. 2013).

Experiences from countries with newly introduced prospective payment systems have shown that coding practices subsequently change after the implementation of new systems. Hsia et al. (1988) found out that after introducing a DRG-based prospective payment system for patients covered by Medicare, coding errors occurred in one-fifth of inpatient cases resulting in a net overpayment for hospitals. In a later study, the rate of coding errors declined (Hsia et al. 1992), which points to a “learning curve” within the system, fostered by regulations for coding and educational activities. Studies from Canada (Preyra 2004) and Sweden (Serdén, Lindqvist, and Rosén 2003) found an increased reporting of secondary diagnoses following the introduction of DRG-based prospective payment systems. Such trends can be interpreted as increasing completeness of previously under-reported diagnoses, but they might also be caused by an over-reporting of secondary diagnoses that are relevant for the amount of payment (so-called upcoding or DRG creep).

In Germany, a prospective payment system for inpatient services based on DRGs was introduced in 2003 and became mandatory in 2004. This type of reimbursement is applied for all inpatient services (except psychiatry) in acute care hospitals and covers all patients and all payers. The German hospital reimbursement reform incentivized providers to pay more attention to coding quality (Ridder, Doege, and Martini 2007). In this context, the

Address correspondence to Ulrike Nimptsch, M.P.H., Department for Structural Advancement and Quality Management in Health Care, School Economics and Management, Technische Universität Berlin, ST 0-1, Steinplatz 2, 10623 Berlin, Germany; e-mail: ulrike.nimptsch@tu-berlin.de.

completeness of secondary diagnosis coding has been recognized as a measure to optimize remuneration (Wenke et al. 2012).

This study investigated whether secondary diagnosis coding in Germany changed during the years following the introduction of DRGs and, if so, whether different trends regarding specific comorbidities were observed. By studying administrative data of patients hospitalized for the common diseases of heart failure, stroke, pneumonia and hip fracture, temporal trends in secondary diagnosis coding and the effects on comorbidity-related risk of in-hospital death were analyzed. In addition, between-hospital variation in coding was evaluated.

METHODS

Data

The study is based on nationwide German administrative hospital data of the years 2005 to 2012. Within the German-refined DRG system (G-DRG), inpatient episodes are grouped into medically and economically homogenous groups based on coded principal and secondary diagnoses, procedures, demographics, and other information. The assignment to a DRG determines the amount of payment for the respective episode. Therefore, inpatient data are electronically recorded in a standardized format. Coding rules are regulated by mandatory coding guidelines that were introduced in 2002. According to these guidelines, a secondary diagnosis should only be coded if the management of the patient during the inpatient episode is affected by therapeutic, diagnostic, nursing, or monitoring efforts. Adherence to the coding guidelines is validated by insurers in suspected cases.

Since 2002, hospitals have been obliged to submit their data annually to a nationwide database of inpatient episodes, which is available for research purposes from 2005 onward. This database (Nationwide DRG Statistics provided by the Research Data Centres of the Federal Statistical Office and the statistical offices of the “Länder”) contains discharge information on all inpatient episodes from German acute care hospitals that are reimbursed via DRG. With respect to hospitalizations for the studied diseases, the data are virtually complete.

Principal and secondary diagnoses are coded via the German adaption of ICD-10 (ICD-10-GM). Information on gender, age, source of admission, procedures, discharge disposition, and length of stay are also included in the data. Based on an anonymized hospital identifier, episodes can be assigned to the respective treating hospital.

The data were evaluated via teleprocessing. Under this mode of access, statistical software scripts are programmed by the researcher but executed by staff members of the Federal Statistical Office. After execution of the scripts, the researcher receives the output files of the analysis results.

Patients

Based on principal diagnosis codes, patients hospitalized for heart failure (ICD-10 Codes I50, I11.0, I13.0, I13.2), stroke (I60, I61, I63, I64), pneumonia (A48.1, J10.0, J11.0, J12, J13, J14, J15, J16, J17, J18), and hip fracture (S72.0, S72.1) were identified. These primary diseases were chosen because they are common reasons for hospitalization accompanied by a relatively high risk of in-hospital death. Additionally, these diseases most commonly affect elderly persons in whom comorbidities are likely to be present.

To exclude hospitalizations that represent secondary sub-acute episodes, for example, for early rehabilitation, patients transferred from other acute care hospitals were identified via the source-of-admission code and excluded from the analysis. Thus, the analysis focuses on episodes in the respective first treating hospital.

Comorbidity Coding

Comorbidity coding was operationalized by the Elixhauser secondary diagnosis groups (Elixhauser et al. 1998) in the specification for ICD-10 (Quan et al. 2005; Turner and Burchill 2006). To identify the presence of diagnosis groups, all 89 secondary diagnosis fields available in the German database were included. The presence of a comorbidity according to the respective secondary diagnosis group was coded as a binary variable. Pairs of associated diagnosis groups (hypertension with and without complication, diabetes with and without complication, solid tumor without metastasis and metastatic cancer) were counted hierarchically to consider only the more severe condition. The secondary diagnosis group of congestive heart failure was not considered in patients hospitalized with a principal diagnosis of heart failure.

Analysis

Depth of coding is illustrated by the annual mean number of Elixhauser diagnosis groups per record. Variations among hospitals in depth of coding is assessed by the hospital-level distribution of the annual mean number of

diagnosis groups per record, whereby indirect standardization was used to control for demographic changes within the 8 years of observation and demographic variations among hospitals. Based on the distribution of each Elixhauser diagnosis group among sex and 5-year age groups of the index year 2005, the probability of the presence of the respective comorbidity was assigned to each inpatient episode. By summarizing these probabilities across all Elixhauser diagnosis groups, the expected mean number per hospital and year was calculated. The ratio of observed and expected means was then multiplied by the crude mean of the index year, thus deriving the age- and gender-adjusted mean.

Temporal changes in the prevalence of each Elixhauser diagnosis group are represented by the age-and-gender standardized morbidity ratio (SMR) and age-and-gender standardized proportions of patients with a coded comorbidity. The SMR is the ratio of the annual observed and expected proportion of a secondary diagnosis group. In the index year 2005, the value of the SMR is 1. A subsequent SMR above 1 indicates an increase over time, and an SMR below 1 indicates a decrease. The annual age-and-gender standardized proportion is the product of the SMR and the respective crude proportion of a secondary diagnosis group in the index year. These figures represent changes in proportions over time that are not attributable to demographic changes during the observation period.

Temporal trends are assessed by simple linear regression models, including least square weighting by number of observations. The p values derived from these models indicate the presence or absence of a linear trend during the years of observation.

To evaluate the impact of changes in secondary diagnosis coding on the respective comorbidity-related risk of in-hospital death, generalized estimating equation models with a logit link function were fitted for each primary disease and the respective first and last year of observation. The explanatory variables included gender, 5-year age groups, and Elixhauser diagnosis groups. The odds ratios for in-hospital death derived from the model of the first year of observation were compared to those derived from the model of the last year of observation. This comparison relies on the hypothesis that an over-reporting of secondary diagnoses would attenuate odds ratios toward the null value of 1.

Model performance is assessed by the c statistic, which is a measure of discrimination. It describes how well the model can differentiate between decedents and survivors based on the explanatory variables included in the model. A c statistic of 1 indicates perfect discrimination, and .5 indicates no

discrimination. However, differences in c statistics were not tested for statistical significance.

The level of statistical significance was set to .05. All data analyses were conducted using *SAS Version 9.3* (SAS Institute Inc., Cary, NC, USA).

RESULTS

Patient Characteristics

Table 1 displays characteristics of patients hospitalized for the studied diseases. From 2005 to 2012, approximately 2.85 million hospitalizations for heart failure, 1.94 million for stroke, 2.10 million for pneumonia, and 0.99 million hospitalizations for hip fracture were identified after the exclusion of patients who were transferred from other acute care hospitals. During the observation period, the annual absolute number of hospitalizations increased for every primary disease. For heart failure patients, an annual average growth of approximately 8,000 patients was observed. The annual average increases in hospitalizations for stroke (1,300), pneumonia (1,500), and hip fracture (750) were lower.

From 2005 to 2012, the mean age of patients increased slightly for heart failure (from 76.5 years to 77.6 years), stroke (from 72.5 to 72.8), and hip fracture (from 78.0 to 79.0). For patients hospitalized for pneumonia, the mean age increased from 56.1 to 59.9 years.

Trends in decreasing crude rates of in-hospital death from 2005 to 2012 were observed for heart failure (from 10.4 to 8.8 percent), stroke (from 11.8 to 9.1 percent), and hip fracture (from 6.0 to 5.5 percent), whereas the mortality for pneumonia showed no trend over time (2005 9.9 percent; 2012 9.6 percent).

The mean number of secondary diagnoses coded per record increased from 7.6 to 10.1 in patients hospitalized for heart failure, from 7.4 to 9.1 for stroke, from 5.5 to 7.7 for pneumonia and from 6.4 to 8.8 for hip fracture.

Trends in Depth of Coding and Variations among Hospitals

The mean number of Elixhauser diagnosis groups identified per record increased from 2005 to 2012 for each studied primary disease (see Table 2). In heart failure patients, the increase from 3.0 to 3.7 was higher than in patients hospitalized for pneumonia and femoral neck fracture (each from 2.0 to 2.5) as well as in patients hospitalized for stroke (from 2.9 to 3.2). Concomitantly, the

Table 1: Patient Characteristics

	2005	2006	2007	2008	2009	2010	2011	2012	<i>p for Trend*</i>
Principal diagnosis of heart failure									
No. of hospitalizations	320,761	328,073	342,319	355,558	364,865	370,475	378,827	385,251	<.01
Mean age	76.5	76.6	76.8	77.0	77.1	77.3	77.5	77.6	<.01
Female, %	54.8	54.4	53.9	53.8	53.4	53.0	52.5	52.1	<.01
In-hospital death, %	10.4	10.3	10.0	9.8	9.7	9.3	9.0	8.8	<.01
Mean length of stay, days	11.7	11.5	11.2	11.0	10.8	10.6	10.5	10.2	<.01
Mean no. of secondary diagnoses per record	7.6	7.9	8.1	8.5	8.9	9.5	9.9	10.1	<.01
Principal diagnosis of stroke									
No. of hospitalizations	235,868	242,703	241,375	243,888	241,194	243,523	243,951	246,579	.02
Mean age	72.5	72.4	72.6	72.8	72.8	72.9	72.9	72.8	<.01
Female, %	52.9	52.1	51.9	52.0	51.6	51.0	50.8	50.2	<.01
In-hospital death, %	11.8	10.9	10.5	10.2	9.8	9.5	9.2	9.1	<.01
Mean length of stay, days	12.5	12.0	11.8	11.6	11.3	11.0	10.8	10.6	<.01
Mean no. of secondary diagnoses per record	7.4	7.6	7.7	8.0	8.3	8.6	8.9	9.1	<.01

continued

Table 1. Continued

	2005	2006	2007	2008	2009	2010	2011	2012	<i>p</i> for <i>Trend</i> *
Principal diagnosis of pneumonia									
No. of hospitalizations	265,048	249,817	251,791	248,997	267,411	269,420	274,727	276,960	.04
Mean age	56.1	55.8	56.7	58.7	59.3	57.6	57.9	59.9	.02
Female, %	45.5	44.6	44.3	43.8	44.3	43.7	43.7	43.6	<.01
In-hospital death, %	9.9	10.1	9.9	10.6	10.5	9.9	9.4	9.6	.34
Mean length of stay, days	10.4	10.3	10.2	10.1	10.0	9.6	9.4	9.3	<.01
Mean no. of secondary diagnoses per record	5.5	5.9	6.1	6.6	6.9	7.1	7.3	7.7	<.01
Principal diagnosis of hip fracture									
No. of hospitalizations	121,595	121,682	119,986	123,817	123,927	128,612	127,669	127,613	<.01
Mean age	78.0	78.0	78.4	78.5	78.5	78.5	78.7	79.0	<.01
Female, %	73.6	73.0	73.5	72.5	71.6	71.0	70.7	70.6	<.01
In-hospital death, %	6.0	5.7	5.8	5.7	5.7	5.4	5.4	5.5	<.01
Mean length of stay, days	17.5	16.9	16.7	16.3	15.8	15.4	15.1	14.9	<.01
Mean no. of secondary diagnoses per record	6.4	6.7	7.0	7.4	7.7	8.0	8.4	8.8	<.01

*Two-sided *p*-value for linear trend. Direction of trend: + increase; - decrease; = no linear trend.

Table 2: Number of Elixhauser Diagnosis Groups Per Record

	2005	2006	2007	2008	2009	2010	2011	2012	<i>p for Trend*</i>	
Principal diagnosis of heart failure										
Mean no. per record	3.0	3.1	3.2	3.3	3.4	3.5	3.6	3.7	<.01	+
Standard deviation	1.6	1.7	1.7	1.7	1.7	1.8	1.8	1.8	<.01	+
No. of diagnosis groups, %										
0	14.7	15.0	15.5	16.0	16.4	17.1	17.5	17.8	<.01	+
1–4	78.1	76.8	75.5	74.1	72.3	69.8	67.7	66.9	<.01	–
5+	7.2	8.1	9.0	10.0	11.3	13.1	14.7	15.3	<.01	+
Principal diagnosis of stroke										
Mean no. per record	2.9	2.9	2.9	3.0	3.0	3.1	3.1	3.2	<.01	+
Standard deviation	1.7	1.7	1.7	1.7	1.8	1.8	1.8	1.8	<.01	+
No. of diagnosis groups, %										
0	15.6	15.7	15.8	15.8	16.0	16.2	16.3	16.3	<.01	+
1–4	77.1	76.8	76.5	76.1	75.4	74.3	73.6	73.1	<.01	–
5+	7.3	7.5	7.8	8.1	8.6	9.5	10.2	10.6	<.01	+
Principal diagnosis of pneumonia										
Mean no. per record	2.0	2.0	2.1	2.2	2.3	2.3	2.4	2.5	<.01	+
Standard deviation	1.8	1.8	1.8	1.9	1.9	2.0	2.0	2.1	<.01	+
No. of diagnosis groups, %										
0	30.6	30.9	29.9	29.0	28.4	30.2	30.1	29.0	.22	=
1–4	65.5	64.5	65.0	65.2	65.0	62.5	62.0	62.0	.01	–
5+	3.9	4.6	5.1	5.8	6.6	7.3	7.9	9.1	<.01	+
Principal diagnosis of hip fracture										
Mean no. per record	2.0	2.0	2.1	2.2	2.2	2.3	2.4	2.5	<.01	+
Standard deviation	1.6	1.7	1.7	1.7	1.8	1.8	1.8	1.9	<.01	+
No. of diagnosis groups, %										
0	24.8	24.6	24.1	23.3	22.9	22.3	21.9	21.2	<.01	–
1–4	72.0	71.9	72.0	72.4	72.2	72.2	72.0	71.7	.81	=
5+	3.2	3.6	4.0	4.4	4.9	5.5	6.1	7.0	<.01	+

*Two-sided *p*-value for linear trend. Direction of trend: + increase; – decrease; = no linear trend.

share of patients with more than five secondary diagnosis groups increased within the observation period. For heart failure, this proportion grew from 7.2 to 15.3 percent, for stroke from 7.3 to 10.6 percent, for pneumonia from 3.9 to 9.1 percent, and for hip fracture from 3.2 to 7.0 percent.

Figure 1 displays boxplots of the age- and gender-adjusted annual distributions of mean Elixhauser diagnosis groups per record at the hospital level, illustrating between-hospital variation in the depth of coding. Variations in coding were evident throughout the observation period. For example, in 2012, the 25th and 75th (5th and 95th) hospital quartile of mean Elixhauser diagnosis groups per record ranged from 3.2 to 4.1 (2.4–4.6) for heart failure,

Figure 1: Mean Age- and Gender-Adjusted Annual Number of Elixhauser Diagnosis Groups Per Record on the Hospital Level, by Principal Diagnosis

Note. The age-and-gender adjusted mean number of Elixhauser diagnosis groups per inpatient record was measured for each hospital. The distribution of this measure among hospitals is shown by box plots. Length of boxes display interquartile range. Symbols in the boxes represent the mean; horizontal lines represent the median. Whiskers extend to the 5th and 95th percentile.

*Adjusted by sex and 5-year age groups (reference = 2005).

from 2.7 to 3.5 (1.8–4.2) for stroke, from 2.1 to 2.7 (1.5–3.2) for pneumonia, and from 2.0 to 2.8 (0.9–3.5) for hip fracture.

Trends in Frequency of Coding of Specific Comorbidities

Regarding specific Elixhauser diagnosis groups, the analysis demonstrated different patterns of changes during the period of observation. Increases in coding frequency were observed in 15 of the 31 secondary diagnosis groups. Decreases were observed in 11 secondary diagnosis groups, whereas the remaining five diagnosis groups showed no or very minor changes in coding frequency.

Figure 2: Standardized Morbidity Ratios for Selected Elixhauser Diagnosis Groups, by Primary Disease

Note. Asterisks indicate a significant different SMR in 2012 compared to 2005. Test of significance relies on 95 percent confidence intervals for the SMR.

In Figure 2, the annual age- and gender-standardized morbidity ratios (SMR) for selected secondary diagnosis groups are displayed. For example, the frequency of coding of fluid and electrolyte disorders increased significantly during the observation period. These increases were higher in patients hospitalized for heart failure (SMR 2012 1.53; 95 percent CI 1.52–1.55) and

hip fracture (1.66; 1.64–1.69) than in patients hospitalized for stroke (1.29; 1.27–1.30) and pneumonia (1.33; 1.32–1.34). Furthermore, the frequency of renal failure increased from 2005 to 2012. This increase was higher for patients hospitalized for hip fracture (SMR 2012 1.86; 1.83–1.89) than for patients hospitalized for heart failure (1.39; 1.38–1.40), stroke (1.38; 1.36–1.40), or pneumonia (1.42; 1.41–1.44). Strong increases, following a linear pattern, were observed in the hypothyroidism diagnosis group. The SMR in 2012 was 2.54 (2.51–2.58) for heart failure patients, 2.60 (2.55–2.64) for stroke, 2.69 (2.65–2.74) for pneumonia, and 3.13 (3.06–3.20) for hip fracture patients. Similarly, the frequency of coagulopathy increased until 2012. This increase was rather moderate until 2008, but it accelerated from 2009 onward. The SMR in 2012 was 2.90 (2.85–2.95) for heart failure, 1.59 (1.55–1.63) for stroke, 2.14 (2.08–2.19) for pneumonia, and 2.85 (2.78–2.92) for hip fracture.

For each primary disease, increases were also observed for the secondary diagnosis groups of pulmonary circulation disorders and depression (see Figure 2), cardiac arrhythmia, uncomplicated hypertension, uncomplicated diabetes, rheumatoid arthritis or collagen vascular disease, and deficiency anemia. Increases in at least one primary disease occurred for valvular disease, peripheral vascular disorders, complicated hypertension, and psychoses (see Figure S2).

For obesity, a decline in the frequency of coding was observed, which was most distinct between 2007 and 2008. The SMR in 2012 was 0.88 (0.87–0.90) in patients hospitalized for heart failure, 0.60 (0.58–0.61) for stroke, 0.76 (0.74–0.78) for pneumonia, and 0.59 (0.57–0.61) for hip fracture patients. The frequency of peptic ulcer disease excluding bleeding decreased by half. In 2012, the SMR was 0.49 for heart failure (95 percent CI 0.45–0.54), 0.44 for stroke (0.38–0.50), 0.46 for pneumonia (0.40–0.52), and 0.38 for hip fracture (0.31–0.46) (see Figure 2).

Further diagnosis groups for which a decrease in coding was statistically significant for all primary diseases were lymphoma and alcohol abuse. Decreases in at least one primary disease were observed for the secondary diagnosis groups of congestive heart failure, complicated diabetes, liver disease, metastatic cancer, solid tumor without metastasis, weight loss, and blood loss anemia.

No significant or very minor changes were found for paralysis, other neurological disorders, chronic pulmonary disease, and drug abuse (see Figure S2).

Age- and gender-standardized proportions of coded secondary diagnosis groups for each primary disease are displayed in Tables S4A to S4D.

Figure 3: Odds Ratios of In-Hospital Death 2005 and 2012 in Patients Hospitalized for Heart Failure

Note. Whiskers indicate 95 percent confidence intervals. *In patients hospitalized for heart failure a secondary diagnosis of congestive heart failure was excluded.

Comorbidity-Related Risk of In-Hospital Death

In Figure 3 the odds ratios of in-hospital death for patients hospitalized for heart failure predicted from the model fitted on 2005 data are plotted against those predicted from the model fitted on 2012 data. Statistically significant shifts of the odds ratios between 2005 and 2012 were observed for nine secondary diagnosis groups. Eight of these shifts were directed toward the null value of 1; that is, the associated risk of in-hospital death was attenuated in the

2012 model compared to that in the 2005 model. For example, the risk associated with pulmonary circulation disorders demonstrated a shift to the null. While the odds ratio was 1.17 (95 percent CI 1.11–1.23) in 2005, the OR was attenuated to 0.97 (0.93–1.01) in 2012. Likewise, the risk associated with renal failure diminished from an OR of 1.19 (1.16–1.23) to 1.04 (1.01–1.07).

Not all Elixhauser diagnosis groups are associated with an increased risk of death. In the multivariate analysis, 12 of these groups were identified as “protective” factors in heart failure patients. One of those is hypothyroidism, for which the risk associated with in-hospital death was 0.59 (0.54–0.63) in the 2005 data. During the observational period, this risk shifted toward the null, resulting in an OR of 0.72 (0.68–0.75) in the 2012 model.

Other significant shifts towards the null were found for the secondary diagnosis groups of valvular disease, peripheral vascular disorders, uncomplicated diabetes, weight loss, and depression. For complicated diabetes, the associated risk of death increased from 2005 (0.98; 0.94–1.01) to 2012 (1.10; 1.06–1.14). For the remaining 22 secondary diagnosis groups, no significant changes in the associated risk of death were observed in patients hospitalized for heart failure. The models’ performance assessed by the c statistics decreased slightly from 2005 ($c = 0.73$) to 2012 ($c = 0.72$). However, the statistical significance of this difference was not tested.

Odds ratios derived from the multivariate analyses for the other primary diseases are displayed in Figure S3. In patients hospitalized for stroke, a shift of mortality risk toward the null occurred for the secondary diagnosis groups of peripheral vascular disorders and weight loss. Increased risks in 2012 compared to 2005 were found for other neurological disorders and solid tumor without metastasis. Model discrimination improved between 2005 ($c = 0.76$) and 2012 ($c = 0.79$), but it was not tested for statistical significance.

In patients hospitalized for pneumonia, all observed shifts of comorbidity-related risks of in-hospital death converged toward the null. These shifts concern the secondary diagnosis groups of valvular disease, pulmonary circulation disorders, hypothyroidism, renal failure, and weight loss. The discrimination of the 2012 model was slightly lower ($c = 0.79$) compared to the 2005 model ($c = 0.81$).

The models fitted on data of patients hospitalized for hip fracture demonstrated risk shifts toward the null regarding the secondary diagnosis groups of pulmonary circulation disorders, renal failure, coagulopathy, and weight loss. The model discrimination remained unchanged (2005, $c = 0.81$; 2012, $c = 0.81$).

DISCUSSION

In this study, approximately 7.88 million inpatient episodes were analyzed and reflected episodes that occurred during the first 8 years after the introduction of DRGs in Germany. One general finding is that the depth of comorbidity coding increased from 2005—the first year after introduction of DRGs—until 2012. The absolute number of coded secondary diagnoses as well as the mean number of Elixhauser diagnosis groups per record and the share of patients with more than five diagnosis groups increased significantly in patients hospitalized for the studied diseases. These general increases were most distinct in heart failure patients, moderate in patients hospitalized for pneumonia and hip fracture, and rather mild in stroke patients.

Variation among hospitals in depth of secondary diagnosis coding was observed throughout the whole observation period. This variation might be related to organizational factors, such as monitoring of coding, allocation of responsibilities for coding (medical doctors vs. nonphysician clinical coders), or training of coders. In any case, the observed variation indicates that comorbidity-adjusted comparisons between hospitals might be biased independently from temporal trends in coding. In extreme cases, risk adjustment using secondary diagnoses might rather control for coding behavior than for comorbidity-related case mix variation between hospitals.

The analysis of coding frequency of specific Elixhauser diagnosis groups showed different trends over time. Increased coding, as observed for 11 secondary diagnosis groups, is likely to be driven by financial incentives induced by the DRG-based prospective payment system. Of the 31 Elixhauser diagnosis groups, 28 groups contain ICD-10 codes that are defined as complications and comorbidities (CC) within the G-DRG system. Coding of CC-relevant secondary diagnoses in the administrative data record results in a higher patient clinical complexity level (PCCL). A high PCCL may raise the payment by assigning the inpatient episode to a higher-priced DRG. Given the principal diagnoses studied here, one example that would result in a PCCL of 3 is a combination of diagnosis codes of coagulopathy (D68.30 haemorrhagic disorder due to circulating anticoagulants) and fluid and electrolyte disorders (E87.6 hypokalaemia). A PCCL of 3 is the second highest clinical complexity level. The PCCL influences the amount of payment in the DRGs for heart failure, pneumonia and hip fracture, whereas the amount of payment for stroke rather depends on the type of stroke or therapeutic measures. This fact might explain why the observed general increases in secondary diagnosis coding

were less distinct in stroke patients compared to those for patients hospitalized for the other studied diseases. However, due to ongoing recalculation and adjustment of the German DRGs, the financial impact of the PCCL has been reduced during the recent years.

Counter-intuitively, decreases in coding were observed for some secondary diagnosis groups that are CC-relevant, such as liver disease or metastatic cancer. These decreases might be driven by ongoing examinations initiated by insurers. Such examinations focus on wrong or fraudulent billing in view of the German coding guidelines (Vetter et al. 2009) and might have induced some learning effects regarding guideline-compliant coding of secondary diagnoses. Other CC-relevant secondary diagnosis groups, for example, chronic pulmonary disease, paralysis, and other neurological disorders, showed no or very minor changes in coding during the observation period. However, the ICD-10 codes covered by three Elixhauser groups had no impact on DRG assignment at all. Those groups were obesity and peptic ulcer disease, excluding bleeding, for which a decreased frequency of coding was observed over time, and also uncomplicated hypertension, which demonstrated increases.

In addition to financial incentives, changes in the classification system might have determined changes in coding practice. For example, in the ICD-10-GM, the diagnosis of obesity was specified according to the body mass index in 2008, when a sudden decline in the frequency of coding was observed.

Although most changes are likely to be caused by altered coding practices, some other factors should be considered. Observed increases in secondary diagnosis coding could also be a consequence of increasing numbers of tests, leading to more health problems being discovered during hospitalization. For instance, the increased frequency of hypothyroidism might be related to a more frequent measurement of hormone markers (TSH and thyroid hormones) in blood serum. Correspondingly, the frequencies of hypothyroidism as a secondary diagnosis in patients hospitalized for the studied diseases (see Tables S4A to S4D) seem to approach successively the prevalence of hypothyroidism as estimated in the elderly population (Laurberg et al. 2005). Similarly, clinical awareness for depression as a comorbidity might also have increased during the studied time span.

Some of the observed changes in comorbidity coding might reflect true changes in epidemiology. This might be the case for peptic ulcer disease where the observed decrease of coding corresponds to the general decline in incidence and prevalence in western countries, attributed to the wider use of

proton pump inhibitors and the decrease of *Helicobacter pylori*-associated peptic ulcer disease (Sung, Kuipers, and El-Serag 2009).

The analysis of the risk of in-hospital death associated with specific Elixhauser diagnosis groups was carried out to detect nonconstant risk, possibly related to upcoding. The observed shifts of risk toward the null are likely to be caused by an increased coding of mild or uncertain manifestations of the respective comorbidity, consequentially attenuating the associated risk of death as observed for pulmonary circulation disorders, renal failure, and hypothyroidism. These indications of nonconstant risk over time are a possible source of bias in longitudinal analyses of comorbidity-adjusted outcomes (Nicholl 2007).

However, in the present study, for the majority of diagnosis groups rather stable associated risks of death were observed over time, even though the reporting of some of them increased within the observation period. For instance, the increased frequency of coding of uncomplicated hypertension as well as fluid and electrolyte disorders did not significantly change the associated risk of death over time. In these cases, the increase in coding might represent improved capturing of prevalent comorbidities rather than upcoding. These findings point to another possible source of bias, as increased completeness of coding over time means that patients with a prevalent comorbidity might be more often misclassified in early periods.

The multivariate analyses showed that not all of the Elixhauser diagnosis groups were associated with a higher risk of in-hospital death. Findings of protective effects of comorbidities have been reported before (Elixhauser et al. 1998). For example, hypertension in patients hospitalized for heart failure (Gheorghiade et al. 2006; Sasaki et al. 2013), hypothyroidism in stroke patients (Akhoundi et al. 2011), and obesity in pneumonia patients (Kahlon et al. 2013) were found to be associated with a lower risk of short-term mortality.

The performance of the multivariate risk-prediction models changed only marginally over time. The c statistics indicate a slightly deteriorated discrimination in patients hospitalized for heart failure and pneumonia, implying that the accuracy of risk prediction by Elixhauser diagnosis groups decreased regarding these primary diseases. However, improved model discrimination over time was found in stroke patients, whereas no changes were observed regarding patients hospitalized for hip fracture. However, observed differences of the c statistics were not tested for statistical significance.

The strength of this study is the nationwide database the analyses are based on. By covering all inpatient episodes in Germany, the data are

unaffected by selection bias. However, analyses of administrative data can only reflect what is being coded. Therefore, the true prevalence and the severity of comorbidities could not be assessed within this study. The generalizability of findings might be limited because features of the German administrative data (e.g., the number of secondary diagnosis fields), characteristics of the German DRG system, and the general framework of the German health care system must be taken into account. Because data of the years prior to the introduction of DRGs in Germany are not available, the observed temporal trends cannot reliably be attributed to the hospital reimbursement reform. However, the general increases in secondary diagnosis coding observed in this study are in line with findings from other countries. In the United States (Carter, Newhouse, and Relles 1990), Canada (Preyra 2004), or Sweden (Serdén, Lindqvist, and Rosén 2003), the number of secondary diagnoses subsequently increased after the introduction of DRG-based prospective payment. There is also evidence that rather minor changes in reimbursement, such as Medicare's nonpayment for preventable complications policy (Calderwood et al. 2014), or updates of treatment guidelines (Lindenauer et al. 2012) influence coding practice.

CONCLUSIONS

Comorbidity coding changed subsequently to the hospital reimbursement reform in Germany. With regard to the total number of diagnosis groups per inpatient record, an increase in depth of coding was observed. However, for the frequency of coding of specific comorbidities, varying trends were observed. While some indications for upcoding were found, most of the observed increases point rather to an improved accuracy in terms of capturing prevalent comorbidities, at least in cases where the coding is relevant for the amount of payment. The observed decreases might be related to learning effects regarding the German coding guidelines; that is, comorbidities should only be coded if they affect patient management during the hospitalization.

Comorbidity-adjusted outcomes in longitudinal administrative data analyses may be biased by nonconstant risk over time, changing completeness of coding and between-hospital variation. For most of the Elixhauser diagnosis groups, the association with the risk of in-hospital death did not change over time. Still, the increased frequency of coding as observed for several diagnosis groups implies that patients in early periods might be more often misclassified due to under-reporting of secondary diagnoses and thus be wrongly assigned

to the respective reference group in risk adjustment models. This issue might cause even more bias when analyses focus on the hospital level or other subsamples of data with different, yet unknown, probabilities of misclassification.

As recommended by Elixhauser et al. (1998), comorbidity measures should be applied in purposive modification, considering the studied disease and outcome and the specific research question. In addition, possible changes in coding of risk adjustment variables must be taken into account when outcomes are studied in a longitudinal perspective. This is particularly important when the observation period coincides with changes in the reimbursement system or in other conditions that are likely to alter incentives for clinical coding practice.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: The Department for Structural Advancement and Quality Management in Health Care, Technische Universität Berlin, receives ongoing funding from the Helios Hospital Group.

Disclosures: None.

Disclaimers: None.

REFERENCES

- Akhoundi, F. H., A. Ghorbani, A. Soltani, and A. Meysamie. 2011. "Favorable Functional Outcomes in Acute Ischemic Stroke Patients with Subclinical Hypothyroidism." *Neurology* 77 (4): 349–54.
- Calderwood, M. S., K. Kleinman, S. B. Soumerai, R. Jin, C. Gay, R. Platt, W. Kassler, D. A. Goldmann, A. K. Jha, and G. M. Lee. 2014. "Impact of Medicare's Payment Policy on Mediastinitis Following Coronary Artery Bypass Graft Surgery in US Hospitals." *Infection Control and Hospital Epidemiology* 35 (2): 144–51.
- Carter, G. M., J. P. Newhouse, and D. A. Relles. 1990. "How Much Change in the Case Mix Index Is DRG Creep?" *Journal of Health Economics* 9 (4): 411–28.
- Elixhauser, A., C. Steiner, D. R. Harris, and R. M. Coffey. 1998. "Comorbidity Measures for Use with Administrative Data." *Medical Care* 36 (1): 8–27.
- Gheorghiade, M., W. T. Abraham, N. M. Albert, B. H. Greenberg, C. M. O'Connor, L. She, W. G. Stough, C. W. Yancy, J. B. Young, C. G. Fonarow, and OPTIMIZE-HF Investigators and Coordinators. 2006. "Systolic Blood Pressure at Admission, Clinical Characteristics, and Outcomes in Patients Hospitalized with Acute Heart Failure." *JAMA* 296: 2217–26.

- Hsia, D. C., W. M. Krushat, A. B. Fagan, A. J. Tebbutt, and R. P. Kusserow. 1988. "Accuracy of Diagnostic Coding for Medicare Patients under the Prospective-Payment System." *New England Journal of Medicine* 318 (6): 352–5.
- Hsia, D. C., C. A. Ahern, B. P. Ritchie, L. M. Moscoe, and W. M. Krushat. 1992. "Medicare Reimbursement Accuracy under the Prospective Payment System, 1985 to 1988." *JAMA* 268 (7): 896–9.
- Jaques, R. M., J. Fotheringham, M. J. Campbell, and J. Nicholl. 2013. "Did Hospital Mortality in England Change from 2005 to 2010? A Retrospective Cohort Analysis." *BMC Health Services Research* 13 (1): 216.
- Kahlon, S., D. T. Eurich, R. S. Padwal, A. Malhotra, J. K. Minhas-Sandhu, T. J. Marrie, and S. R. Majumdar. 2013. "Obesity and Outcomes in Patients Hospitalized with Pneumonia." *Clinical Microbiology and Infection* 19 (8): 709–16.
- Laurberg, P., S. Andersen, I. Bülow Pedersen, and A. Carlé. 2005. "Hypothyroidism in the Elderly: Pathophysiology, Diagnosis and Treatment." *Drugs and Aging* 22 (1): 23–38.
- Lindenauer, P. K., T. Lagu, M.-S. Shieh, P. S. Pekow, and M. B. Rothberg. 2012. "Association of Diagnostic Coding with Trends in Hospitalizations and Mortality of Patients with Pneumonia, 2003-2009." *JAMA* 307 (13): 1405–13.
- Mohammed, M. A., J. J. Deeks, A. Girling, G. Rudge, M. Carmalt, A. J. Stevens, and R. J. Lilford. 2009. "Evidence of Methodological Bias in Hospital Standardised Mortality Ratios: Retrospective Database Study of English Hospitals." *BMJ* 338: b780.
- Nicholl, J. 2007. "Case-Mix Adjustment in Non-Randomised Observational Evaluations: The Constant Risk Fallacy." *Journal of Epidemiology and Community Health* 61 (11): 1010–3.
- Preyra, C. 2004. "Coding Response to a Case-Mix Measurement System Based on Multiple Diagnoses." *Health Services Research* 39 (4 Pt 1): 1027–45.
- Quan, H., V. Sundararajan, P. Halfon, A. Fong, B. Burnand, J. C. Luthi, L. D. Saunders, C. A. Beck, T. E. Feasby, and W. A. Ghali. 2005. "Coding Algorithms for Defining Comorbidities in ICD-9-CM and ICD-10 Administrative Data." *Medical Care* 43 (11): 1130–9.
- Ridder, H.-G., V. Doege, and S. Martini. 2007. "Differences in the Implementation of Diagnosis-Related Groups across Clinical Departments: A German Hospital Case Study." *Health Services Research* 42 (6 Pt 1): 2120–39.
- Sasaki, N., J. Lee, S. Park, T. Umegaki, S. Kunisawa, T. Otsubo, H. Ikai, and Y. Imanaka. 2013. "Development and Validation of an Acute Heart Failure-Specific Mortality Predictive Model Based on Administrative Data." *The Canadian Journal of Cardiology* 29 (9): 1055–61.
- Serdén, L., R. Lindqvist, and M. Rosén. 2003. "Have DRG-Based Prospective Payment Systems Influenced the Number of Secondary Diagnoses in Health Care Administrative Data?" *Health Policy* 65 (2): 101–7.
- Sung, J. J., E. J. Kuipers, and H. B. El-Serag. 2009. "Systematic Review: The Global Incidence and Prevalence of Peptic Ulcer Disease." *Alimentary Pharmacology & Therapeutics* 29 (9): 938–46.

- Turner, K., and C. Burchill. 2006. "Elixhauser Comorbidity Index Macro Code Using ICD-10" [accessed on October 9, 2012]. Available at http://mchp-appserv.cpe.umanitoba.ca/Upload/SAS/_ElixhauserICD10.sas.txt
- Vetter, S. Y., S. Studier-Fischer, A. Wentzensen, and C. Frank. 2009. "The Challenge of Auditing by Medical Health Insurance Inspectors: Development of Individual Case Inspections According to 275ff SGB V." *Unfallchirurg* 112 (8): 756–8.
- Wenke, A., A. Gaber, L. Hertle, N. Roeder, and G. Pühse. 2012. "Complexity Level Simulation in the German Diagnosis-Related Groups System: The Financial Effect of Coding of Comorbidity Diagnostics in Urology." *Urologe A* 51 (7): 975–81.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Figure S1: Standardized Morbidity Ratios for Selected Elixhauser Diagnosis Groups, by Primary Disease (color).

Figure S2: Standardized Morbidity Ratios for Elixhauser Diagnosis Groups, by Primary Disease.

Figure S3: Odds Ratios of In-Hospital Death 2005 and 2012 in Patients Hospitalized for Stroke, Pneumonia, and Hip Fracture.

Table S4A: Proportion of Elixhauser Diagnosis Groups Coded in Data Records of Patients Hospitalized for Heart Failure.

Table S4B: Proportion of Elixhauser Diagnosis Groups Coded in Data Records of Patients Hospitalized for Stroke.

Table S4C: Proportion of Elixhauser Diagnosis Groups Coded in Data Records of Patients Hospitalized for Pneumonia.

Table S4D: Proportion of Elixhauser Diagnosis Groups Coded in Data Records of Patients Hospitalized for Hip Fracture.