

Online Appendix to “Operational Implications of Diagnostic Discrepancies: The Weigh-In of Focal and Related Physician Experience”

Seokjun Youn¹, Yingchao Lan²

¹Eller College of Management, The University of Arizona

²College of Business, University of Nebraska-Lincoln

Appendix A: Methodology for Diagnostic Discrepancy Classification and Retrospective Audit Framework

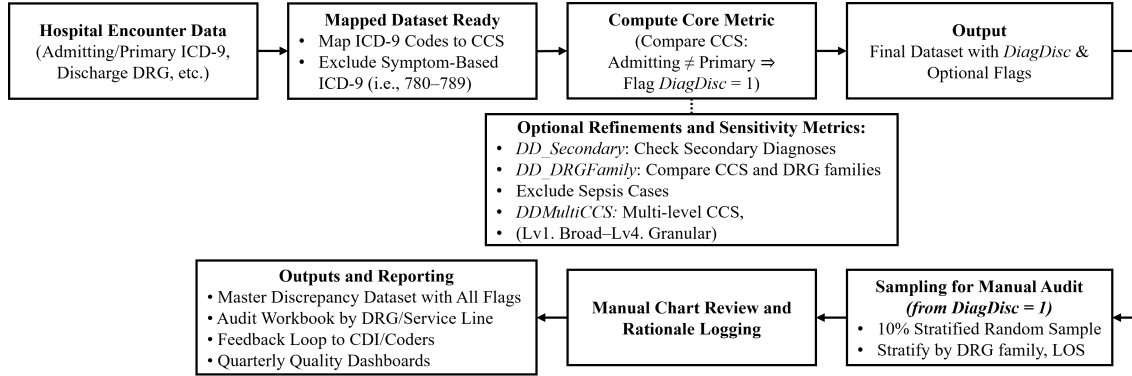


Figure A.1 Diagnostic Discrepancy Classification and Audit Framework

We provide an audit-ready framework, summarized in Figure A.1, for identifying admitting-to-principal diagnostic misalignment, which has potential applicability to retrospective operational monitoring, benchmarking, and case review for hospital operations and continuous improvement efforts (e.g., Clinical Documentation Improvement (CDI) programs, physician training, and retrospective decision support). This classification framework is based on the Clinical Classifications Software (CCS) to flag larger diagnostic shifts that may warrant coordination or documentation review, without presuming diagnostic error. We further supplement our main CCS-based approach by conducting extensive alternative operationalizations of diagnostic discrepancies, such as 1) examining secondary diagnoses, 2) comparing DRG families to CCS families, 3) incorporating multi-level CCS categories, 4) multi-scheme discordance across alternative diagnosis-classification systems and 5) excluding certain clinically expected discrepancy groups such as sepsis.

1. **Data Source, Scope, and Structure:** We developed a methodology to identify and classify diagnostic discrepancies between the initial admitting diagnosis and the principal diagnosis for each hospital encounter. We begin by understanding the types of diagnostic discrepancies summarized in Appendix B. This approach serves as the foundation of an audit framework that highlights potential diagnostic discrepancies or documentation issues. The methodology balances clinical nuance with policy relevance: it provides a systematic, data-driven metric that is detailed enough for clinicians and data scientists, yet straightforward and interpretable for administrators and compliance reviewers.
2. **Primary Objective and Core Metric (*DiagDisc*):** The core measure of our framework is a diagnostic discrepancy indicator called *DiagDisc*. For each admission, *DiagDisc* evaluates whether the admitting diagnosis (the provisional diagnosis recorded at entry) falls under a different diagnostic category than the principal diagnosis after diagnostic investigations. To make this comparison robust, we leverage

the single-level Clinical Classifications Software (CCS) from AHRQ.¹ The single-level CCS aggregates 14,567 ICD-9-CM diagnosis codes into 285 mutually exclusive, diagnostically meaningful categories. Appendix C provides examples of mapping ICD-9-CM diagnosis codes to CCS categories. By mapping both the admitting and principal diagnoses to their respective CCS categories, we can determine whether they fall in the same or different CCS categories. A discrepancy is flagged (i.e., *DiagDisc* = 1) if and only if the admitting diagnosis’s CCS category differs from the principal diagnosis’s CCS category.

3. **Conceptual Foundation: What Constitutes a Discrepancy?** In practice, using CCS groupings helps filter out superficial coding differences and focus on a scalable, reproducible marker of diagnostic distance. For example, if an admission’s initial ICD-9 code and principal ICD-9 code refer to variants of the same condition (say, both being types of myocardial infarction), they will map to the same CCS category (e.g., CCS category 100, “Acute myocardial infarction”), and *DiagDisc* will treat this as no discrepancy. In this way, CCS-based comparison mitigates false-positive flags that could arise from granular coding variations or updates (Hong et al. 2024), providing a consistent basis for analysis across large datasets.

DiagDisc is designed to ignore certain expected diagnosis changes. Also, some patients are admitted with non-specific, symptom-based diagnoses when the exact cause of illness is not yet known. Therefore, if the admitting diagnosis falls in the ICD-9 780–789 range (the chapter of “Symptoms, Signs, and Ill-defined Conditions” Buck (2015)), we do not count it as a discrepancy even if the final diagnosis is in a different category, and they are excluded. This rule covers common scenarios such as a patient admitted for “chest pain” (ICD-9 786.50, a general symptom) who is later diagnosed with an acute myocardial infarction. Here the admitting CCS category (symptoms) naturally differs from the principal CCS category (e.g., acute ischemic heart disease), yet that change reflects a proper diagnostic workup rather than an error.

In contrast, if a patient is admitted with a more definitive initial diagnosis that differs from the principal diagnosis, say an admission labeled as pneumonia (a respiratory infection CCS category) but an ultimate principal diagnosis of congestive heart failure (a cardiovascular CCS category), then *DiagDisc* would flag a discrepancy. Such a case represents cross-category admitting-to-principal diagnostic misalignment, potentially reflecting diagnostic reclassification, a significant evolution in the patient’s condition, or documentation/coding dynamics. We ensure that *DiagDisc* highlights larger diagnostic shifts that may warrant further coordination, documentation review, or audit follow-up.

Once computed for each encounter, the *DiagDisc* metric can be aggregated or stratified to inform institutional audits. For example, a hospital’s overall discrepancy rate (the percentage of admissions flagged by *DiagDisc*) can be calculated and compared across departments or time periods. Outliers can be detected, such as a particular service line with an unusually high frequency of diagnostic discrepancies, prompting targeted review. The audit framework uses these metrics in conjunction with traditional quality assurance processes: a high *DiagDisc* rate does not by itself confirm improper practice, but it flags cases for further clinical investigation and coding compliance checks. Through this combination of quantitative flags and qualitative follow-up, the framework provides a balanced approach to oversight, ensuring that potential issues in diagnostic coding or care delivery are systematically brought to light.

4. Optional Refinements and Sensitivity Metrics for Robustness Checks:

In addition to the primary *DiagDisc* measure, we developed several variant metrics as analytical extensions to test the robustness of our approach (these are reported only as supplementary analyses rather than as main results).

- One such variant, *DD.Secondary*, checks whether the admitting diagnosis’s CCS category matches any of the diagnoses recorded (including up to 31 secondary diagnoses). This variant addresses cases where the initial diagnostic suspicion was not the main cause of admission but was still present as a comorbidity or secondary condition. If the admitting diagnosis appears at least among the secondary diagnoses, then *DD.Secondary* does not count it as a discrepancy, effectively recognizing that the admitting diagnosis remained clinically documented during the hospitalization. The results are consistent with those in Table G.1.

¹ Clinical Classifications Software (CCS) for ICD-9-CM Fact Sheet, <https://hcup-us.ahrq.gov/>, Accessed: 6/1/2025.

- Another extension, *DD_DRGFamily*, evaluates consistency at the level of Diagnosis-Related Group (DRG) families. DRG families group clinically related diagnoses for billing and case-mix purposes; thus, it flags a discrepancy only if the admitting and principal diagnoses fall into different DRG families. See Table A.1 for specific DRG and CCS codes considered to classify into circulatory, respiratory, nervous, infectious, or digestive family. This provides a higher-level check, asking whether an apparent discrepancy still lies within the same broad clinical category or represents a larger cross-category diagnostic shift. The results are consistent with those in Table G.1.

Table A.1 DRG Family Categories and Corresponding CCS Family Codes

Family	DRG Range	CCS Codes (admit/principal)
Circulatory	280–316	100, 101, 102
Respiratory	163–208	122, 127
Nervous	20–35	109, 95
Infectious	870–897	2
Digestive	329–393	140, 141

- We implemented multi-level CCS versions of the discrepancy metric (denoted *DDMultiCCSlv1_Pri* to *DDMultiCCSlv4_Pri*), which compare diagnoses at broader CCS hierarchy levels. Table A.2 summarizes single-level and multi-level CCS counts and granularity. For example, instead of requiring an exact single-level CCS match, a multi-level variant might only require that the admitting and principal diagnoses share a major diagnostic category (such as both being diseases of the circulatory system). This tests whether our findings change when using a coarser-grained definition of “same diagnosis.” We further consider if the admitting diagnosis appears at least among the secondary diagnoses (denoted *DDMultiCCSlv1_Sec* to *DDMultiCCSlv4_Sec*). These optional analyses yielded patterns similar to the primary analysis (See Table G.2), reinforcing that the single-level *DiagDisc* metric captures the salient information needed for the audit. We therefore focus our discussion on the main *DiagDisc* results, using the extensions only to provide additional context where relevant.

Table A.2 Summary of Single-Level and Multi-Level CCS Counts and Granularity

CCS Type	Granularity	Approx. Count
Single-Level CCS	Clinically grouped	285
Multi-Level CCS		
Level 1	Body system (broad)	18
Level 2	Major condition groups	259
Level 3	Sub-condition groups	558
Level 4	Highly specific groups	1229

- Finally, we conducted a sensitivity analysis excluding sepsis cases from the *DiagDisc* calculation. Sepsis is a complex, systemic condition often arising secondary to localized infections, and it can complicate diagnostic coding (for instance, a patient might be admitted for pneumonia but develop sepsis, shifting the principal diagnosis). By removing all encounters involving sepsis as either principal or secondary diagnoses from consideration, we ensured that our discrepancy rates were not unduly influenced by such cases. In our data, about 0.31M encounters include sepsis, which is about 5.76% out of over 5.7M encounters. The results are consistent with those in Table G.5.
5. **Algorithmic Flow** The computational pipeline is as follows: convert all ICD-9 codes to CCS; calculate *DiagDisc* by comparing single-level CCS categories while excluding 780–789 admits; iterate through optional rules (*DD_Secondary*, *DD_DRGFamily*, sepsis exclusion) as overlays rather than replacements; and, finally, recompute discrepancy indicators under each multi-level CCS resolution. Each step is applied to a copy of the data set and retained as a distinct field so that analysts can toggle between definitions without rerunning mappings.
 6. **Sampling for Manual Audit** Because even a refined automated flag cannot judge documentation sufficiency, a 10% stratified random sample of cases with *DiagDisc* = 1, and, where desired, the comparable subsets under the optional metrics, could be drawn for human review. Stratification may occur on DRG family and length of stay to ensure coverage of diverse clinical contexts. Reviewers assess whether progress notes, imaging, and laboratory results reveal a defensible diagnostic pathway. Whenever they

reclassify a cross-category flag as explainable based on documentation and clinical context, the rationale is captured to improve rule logic.

7. **Outputs and Reporting** The framework produces (i) a master data set with every discrepancy flag, (ii) a workbook summarizing discrepancy rates by service line, DRG family, and physician, and (iii) a log of audited cases with reviewer determinations. Recurrent coding or documentation errors uncovered during audit are fed back to the Clinical Documentation Integrity (CDI) team for education and to coders for follow-up. Hospital quality committees receive quarterly dashboards showing both the raw *DiagDisc* rate and the proportion requiring documentation or clinical-review follow-up after review.
8. **Summary Statistics of Metrics Operationalized based on Our Data** Tables A.3 and A.4 provide summary statistics and a correlation matrix of alternative diagnostic discrepancy variables following our framework. *DiagDisc_ICD9* and *DD_Secondary_ICD9*, based on principal and secondary diagnoses, respectively, are discrepancy metrics solely based on granular ICD-9-CM codes, acting as benchmarks to our CCS-based discrepancy framework. Relative to traditional operationalizations based solely on raw ICD-9-CM codes or their first three digits, our CCS-based framework reclassifies 10.77% (663,363) and 3.62% (222,939) of cases, respectively, as within-category diagnostic updates.

Table A.3 Summary Statistics of Alternative Diagnostic Discrepancy Variables

Variable	Mean	Std. dev.	Min	Max	Obs	Missing (%)
Main Discrepancy Metric Based on the Proposed Classification Framework						
<i>DiagDisc</i>	0.5244	0.4994	0	1	6,158,529	0.00
Alternative Metrics Incorporating Secondary Diagnoses and DRG Family						
<i>DD_Secondary</i>	0.3479	0.4763	0	1	6,158,529	0.00
<i>DD_DRGFamily</i>	0.4222	0.4939	0	1	6,158,529	0.00
<i>DiagDisc</i> (Excluding Sepsis)	0.5123	0.4998	0	1	5,803,512	0.00
Alternative Metrics Utilizing Multi-Level CCS and Principal Diagnosis						
<i>DDMultiCCSlv1_Pri</i>	0.3716	0.4832	0	1	6,158,529	0.00
<i>DDMultiCCSlv2_Pri</i>	0.4588	0.4983	0	1	5,903,062	4.15
<i>DDMultiCCSlv3_Pri</i>	0.5342	0.4986	0	1	5,350,944	13.11
<i>DDMultiCCSlv4_Pri</i>	0.3994	0.4898	0	1	1,717,366	72.11
Alternative Metrics Utilizing Multi-Level CCS and Secondary Diagnoses						
<i>DDMultiCCSlv1_Sec</i>	0.1467	0.3538	0	1	6,158,529	0.00
<i>DDMultiCCSlv2_Sec</i>	0.2662	0.4420	0	1	5,903,062	4.15
<i>DDMultiCCSlv3_Sec</i>	0.3798	0.4853	0	1	5,350,944	13.11
<i>DDMultiCCSlv4_Sec</i>	0.2205	0.4146	0	1	1,717,366	72.11
Alternative Metrics Based on First 3-Digit of ICD-9-CM Codes						
<i>DD_ICD9_3Digit_Pri</i>	0.5606	0.4963	0	1	6,158,529	0.00
<i>DD_ICD9_3Digit_Sec</i>	0.4063	0.4911	0	1	6,158,529	0.00
Discrepancy Metrics Based Directly on ICD-9-CM Codes						
<i>DiagDisc_ICD9</i>	0.6321	0.4822	0	1	6,158,539	0.00
<i>DD_Secondary_ICD9</i>	0.5147	0.4998	0	1	6,158,539	0.00
Multi-scheme Diagnostic Discrepancy Metric						
<i>DiagDisc_Robust</i>	0.1926	0.3943	0	1	5,903,062	4.15

Notes. Reports alternative binary discrepancy measures in the full pre-analysis sample; each mean is the discrepancy rate. “Pri” compares admitting and principal diagnoses, whereas “Sec” credits secondary-diagnosis matches. Analysis samples vary slightly across tables because of missing values in outcome-specific variables and the removal of fixed-effect singletons. *DiagDisc_Robust* equals one when the admitting and principal diagnoses are classified into different diagnostic groups across major alternative coding schemes considered, including single-level CCS, first-three-digit ICD-9 grouping, DRG family, and multi-level CCS groupings up to level 2, along with consideration of secondary diagnoses (a total of nine coding variants were considered). The summary statistics reported are based on the full pre-analysis sample before listwise deletion.

Table A.4 Correlation Table of Alternative CCS-Based Discrepancy Variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(1) <i>DiagDisc</i>	1.0000										
(2) <i>DD_Secondary</i>	0.7077	1.0000									
(3) <i>DD_DRGFamily</i>	0.8019	0.5605	1.0000								
(4) <i>DDMultiCCSlv1_Pri</i>	0.7218	0.5260	0.6852	1.0000							
(5) <i>DDMultiCCSlv2_Pri</i>	0.8990	0.6223	0.7578	0.8029	1.0000						
(6) <i>DDMultiCCSlv3_Pri</i>	0.9514	0.6733	0.7630	0.6867	0.8553	1.0000					
(7) <i>DDMultiCCSlv1_Sec</i>	0.3802	0.5372	0.3945	0.5267	0.4229	0.3617	1.0000				
(8) <i>DDMultiCCSlv2_Sec</i>	0.5825	0.8231	0.4923	0.5149	0.6479	0.5542	0.6527	1.0000			
(9) <i>DDMultiCCSlv3_Sec</i>	0.6762	0.9215	0.5387	0.4907	0.6010	0.7307	0.4951	0.7585	1.0000		
(10) <i>DD_ICD9_3Digit_Pri</i>	0.9029	0.6387	0.7205	0.6548	0.8176	0.9280	0.3454	0.5298	0.6744	1.0000	
(11) <i>DD_ICD9_3Digit_Sec</i>	0.6355	0.8028	0.5051	0.4184	0.5635	0.6699	0.4350	0.6722	0.8427	0.7346	1.0000

Notes. Reports Pearson correlations in the common complete-case sample (5,350,944 encounters); Level 4 CCS measures are omitted for high missingness.

Appendix B: Types of Diagnostic Discrepancies

Within-category diagnostic updating (“within-category”) is characterized by a plausible sequence of events, such as overlapping symptoms, where workup results support refinement of the diagnosis. These cases typically show clear documentation of differential reasoning and progressive refinement in daily notes. The final diagnosis remains logically connected to the presenting symptoms, rather than being random.

In contrast, cross-category discrepancies reflect larger shifts between the admitting and principal diagnoses. These cases may warrant further review when the admitting-to-principal shift is not readily explained by the available diagnostic pathway or documentation. For example, admitting a patient for suspected pulmonary embolism without appropriate diagnostics and ultimately assigning a low-severity diagnosis like constipation would be flagged as a cross-category discrepancy for potential review. Table B.1 summarizes these categories. While no methodology is without limitations, our study uses cross-category discrepancies as claims-based markers of admitting-to-principal diagnostic misalignment that may warrant audit attention.

Table B.1 Types of Diagnostic Discrepancies, Acceptability, and Examples

Type	Description	Illustrative Audit Triage	Examples
Rule-Out/Workup Clarification	Admitted for possible condition, but ruled out and different diagnosis made.	Valid (if documented well)	- Admitted for “rule out myocardial infarction (MI),” final diagnosis “gastroesophageal reflux disease (GERD)” - Admitted for “diabetes with ketoacidosis (DKA),” final diagnosis “hyperglycemic state without ketoacidosis”
Diagnostic Evolution	New findings during stay change diagnosis.	Valid	- Admitted for “transient ischemic attack,” final diagnosis “acute ischemic stroke” - Admitted for “lower urinary tract infection,” final diagnosis “acute pyelonephritis”
Refinement	Admitting diagnosis is vague or symptom-based, primary diagnosis is specific.	Valid	Admitted for “abdominal pain,” final diagnosis “acute pancreatitis”
Cross-category shift flagged for review	Admitting diagnosis is unrelated to final diagnosis without clear documented pathway.	Flag for Review	- Admitted for “migraine,” final diagnosis “acute ischemic stroke” - Admitted for “heart failure,” final diagnosis “pneumonia”
Coding/Documentation Error	Wrong diagnosis code assigned due to coder or documentation mistake.	Flag for Review	Admitting diagnosis “appendicitis,” final diagnosis “cholecystitis,” but surgery notes suggest gallbladder all along

Notes. Provides an illustrative taxonomy and audit interpretation. Clarification, diagnostic evolution, and refinement may be appropriate when documented. Unexplained cross-category shifts or coding inconsistencies may warrant review. A review flag does not establish error, preventability, or inappropriate care.

Appendix C: Example ICD-9 to CCS Mapping Table

The Clinical Classifications Software (CCS) systems, developed by AHRQ, were designed to address the limitations of ICD-9-CM diagnosis codes, which number over 14,000 and are often too granular and inconsistently grouped for effective statistical or policy analysis. By aggregating these detailed ICD-9 codes into broader, clinically meaningful and mutually exclusive categories, CCS enables efficient summarization, reporting, and large-scale health services research. In a similar vein, converting granular ICD-9 codes to CCS offers a practical advantage for identifying cross-category admitting-to-principal diagnostic misalignment between admitting and principal diagnoses. Whereas small differences between very specific ICD-9 codes may be clinically insignificant, conversions to CCS categories can highlight larger shifts, such as a change from a circulatory system admission to a digestive system principal diagnosis, that could indicate diagnostic updating, reprioritization across conditions, or documentation/coding dynamics requiring further review. Thus, CCS grouping not only streamlines data analysis but also enhances the detection of larger diagnostic shifts that may be overlooked in highly granular code-to-code comparisons.

The single-level CCS system provides a straightforward, flat classification, mapping each ICD-9 diagnosis code to one of 285 mutually exclusive, clinically meaningful categories. Table C.1 illustrates examples. This structure is ideal for overview analyses, descriptive summaries, and regression modeling, as it enables consistent grouping and simplifies complex coding data for efficient reporting. In contrast, the multi-level CCS system introduces a hierarchical framework with up to four levels, allowing users to analyze diagnoses at varying degrees of specificity, from broad body systems to highly granular sub-conditions. This flexibility supports a range of analytic needs, including body-system-level comparisons, subspecialty planning, and hierarchical modeling.

Utilizing both single- and multi-level CCS groupings serves as a robust analytic approach. By aggregating diagnostic codes at different hierarchical levels, researchers can assess whether findings are sensitive to how categories are defined. Consistent results across both systems in our study provide additional evidence that conclusions are not driven by arbitrary or overly granular grouping, thus reinforcing the robustness of our framework and findings that we further illustrate in a battery of robustness checks (Appendix G).

Table C.1 Mapping of Common Diagnoses to ICD-9 and CCS Categories

Common Diagnosis	ICD-9 Examples	Mapped CCS Category	CCS Description
Chest Pain	786.50, 786.59	102	Nonspecific chest pain
Acute Myocardial Infarction (AMI)	410.xx	100	Acute myocardial infarction
Pneumonia	481, 482.xx, 486	122	Pneumonia
Urinary Tract Infection (UTI)	599.0	159	Urinary tract infections
Heart Failure	428.xx	108	Congestive heart failure
Noninfectious Gastroenteritis	558.9	154	Noninfectious gastroenteritis
Stroke (CVA)	431, 434.xx	109	Acute cerebrovascular disease
Syncope	780.2	245	Syncope
Dehydration	276.51	55	Fluid and electrolyte disorders

Notes. Illustrates ICD-9-CM-to-CCS mappings; “xx” denotes all valid subcodes. The examples show how granular codes are consolidated for admitting-to-principal comparisons and are not exhaustive.

Appendix D: Additional Model-Free Evidence

Figures D.1, D.2, and D.3 illustrate the model-free evidence for the detailed cost categories (i.e., $\ln(\text{LabCost})$, $\ln(\text{RadioCost})$, $\ln(\text{RoomCost})$, $\ln(\text{MedSurgCost})$, and $\ln(\text{PharmCost})$), clinical outcomes (i.e., $\ln(\text{LOS})$ and DispNotHome), and process-of-care outcomes (i.e., $\ln(\text{PrDay})$, PrDiag , and PrDiagOR), respectively.

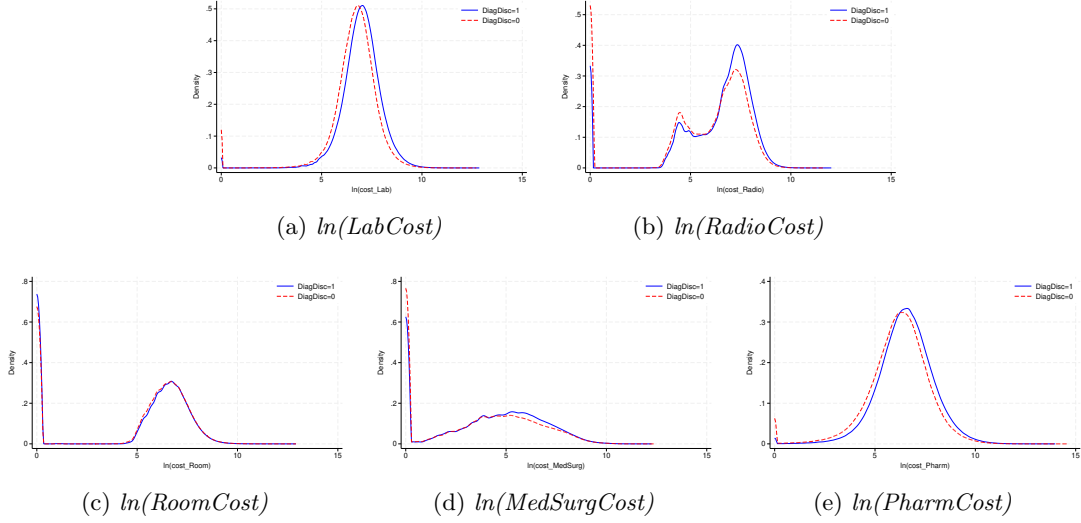


Figure D.1 Kernel Density Estimation of Other Dependent Variables: Detailed Costs

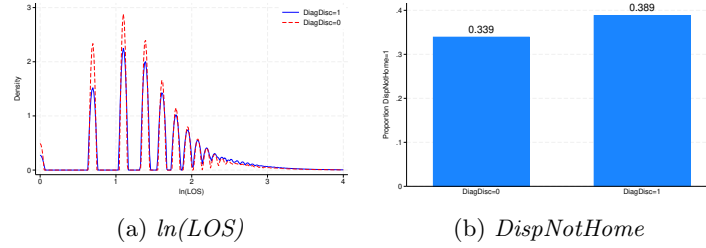


Figure D.2 Kernel Density Estimation and Bar Charts of Clinical Outcomes: $\ln(\text{LOS})$ and DispNotHome

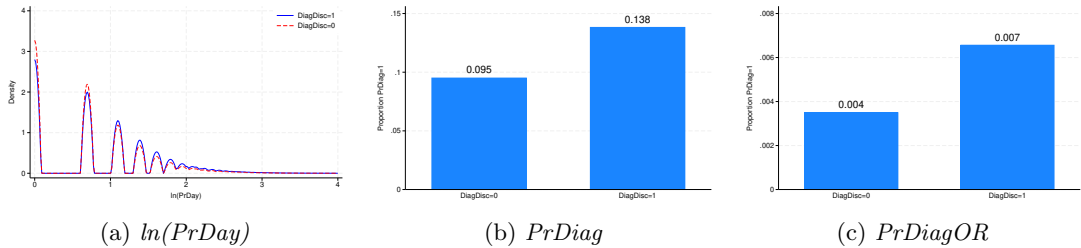


Figure D.3 Kernel Density Estimation and Bar Charts of Process-of-Care Outcomes: $\ln(\text{PrDay})$, PrDiag , and PrDiagOR

Appendix E: Robustness Analysis: Risk-Adjusted Excessive Discrepancies (RAED) and Propensity Score Matching (PSM)

A central interpretive concern is whether the documented cost association reflects the operational burden of admitting-to-principal diagnostic misalignment or is instead an artifact of underlying case-mix differences across encounters and providers. To address this, we construct a risk-adjusted excessive discrepancy (RAED) measure that benchmarks each encounter’s observed discrepancy against the discrepancy that would be *expected* given the patient’s diagnostic and clinical context, and we examine whether the association persists once this benchmark replaces the raw discrepancy indicator.

Constructing the expected-discrepancy benchmark. A credible benchmark requires that the expected discrepancy incorporate information from both the admission and the final-diagnosis stages, rather than from the principal diagnosis alone. We therefore estimate a mixed-effects logistic model of the discrepancy indicator (*DiagDisc*) on a rich covariate set that conditions on admitting CCS, principal CCS, and DRG cohort effects, together with the full patient and contextual information set (age, sex, race, primary payer, admission source, weekend admission, the Elixhauser comorbidity set, and ED workload) and hospital and physician random intercepts. Conditioning jointly on admitting-diagnosis, principal-diagnosis, and DRG cohorts ensures that the expected-discrepancy benchmark reflects the diagnostic information available at both entry and discharge that shapes whether an admitting-to-principal shift is anticipated, rather than the principal diagnosis alone.

Formally, for encounter e treated by physician i at hospital h in diagnosis cohort c , let

$$\hat{p}_{ihc} = \Pr(\text{DiagDisc}_{ihc} = 1 \mid X_{ihc}, \alpha_i, \beta_h)$$

denote the *predicted* probability of a diagnostic discrepancy, where α_i and β_h denote physician- and hospital-specific random effects, respectively, and let

$$\hat{p}_{ihc}^0 = \Pr(\text{DiagDisc}_{ihc} = 1 \mid X_{ihc}, \alpha_i = 0, \beta_h = 0)$$

denote the corresponding *expected* probability that sets the random effects to zero, capturing the discrepancy risk implied solely by observed patient and contextual covariates X_{ihc} . The encounter-level RAED is the ratio of the two, scaled by the realized discrepancy rate within the diagnosis cohort, and is the ratio-scaled “excess” analogue of the CMS excess-readmission ratio (CMS 2023, Zuckerman et al. 2017):

$$\text{RAED}_{ihc} = \frac{\hat{p}_{ihc}}{\hat{p}_{ihc}^0} \times \overline{\text{DiagDisc}}_c,$$

where $\overline{\text{DiagDisc}}_c$ is the realized mean discrepancy rate in cohort c . Values above the cohort benchmark indicate higher-than-expected misalignment given the patient’s diagnostic and clinical context. We then aggregate the encounter-level measure to physician- and hospital-level indices using a ratio-of-sums formulation. Figure E.1 plots the resulting hospital- and physician-level distributions, which reveal substantial heterogeneity in risk-adjusted discrepancy intensity across organizations and providers.

Physician- and hospital-level benchmarking logic. The physician-level analysis tests whether physicians with higher risk-adjusted excess discrepancy than peers in comparable hospital and time contexts also exhibit higher adjusted costs. The hospital-level analysis uses hospital fixed effects to test whether within-hospital increases in excess discrepancy are associated with higher adjusted costs over time. Together, these analyses assess whether the patient-level association persists at both provider and organizational levels.

Results. The association remains consistent across levels of aggregation, although each analysis captures a distinct source of variation. At the patient level (Table E.2, columns (1)–(3)), replacing the raw indicator with encounter-level RAED (Mean = 0.525, SD = 0.323, Min = 0.028, Max = 0.968) preserves the positive cost association ($\beta = 0.0489$, $p < 0.001$) and both experience interactions (focal = -0.0152 ; related = 0.0064), suggesting that the findings are not driven solely by observed patient complexity or diagnosis mix. At the physician level (Table E.3), physicians with greater-than-expected exposure to diagnostic misalignment have higher adjusted costs, while the focal- and related-experience interactions retain their expected signs and statistical significance (-0.2593 and 0.1625 , respectively; both $p < 0.001$). At the hospital level (Table E.4), periods in which a hospital

exhibits higher risk-adjusted excess discrepancy are associated with higher adjusted costs ($\beta = 0.0926$, $p < 0.01$). Because the aggregate analyses benchmark discrepancies against physician- and hospital-level exposure, we focus the hospital-level specification on the aggregate cost association. The moderation constructs are physician- and patient-level and are less naturally aggregated to the institution.

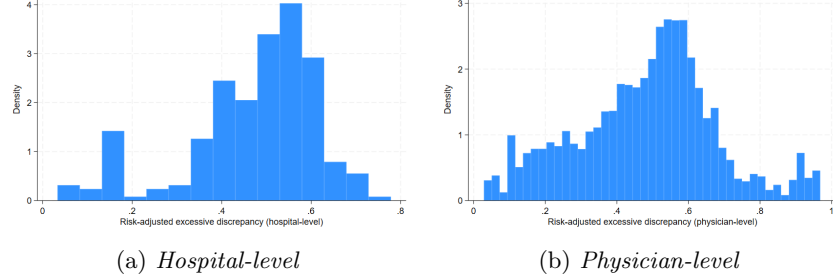


Figure E.1 Histogram of Risk-Adjusted Excessive Discrepancies (RAED)

Propensity score matching. As a complementary strategy to mitigate confounding from selection on physician characteristics (although our main models already include physician fixed effects), we implement a propensity score matching (PSM) design comparing physicians with relatively high versus low RAED but otherwise similar observable human capital profiles. We define a treatment indicator equal to one if a physician's RAED exceeds the sample mean and estimate a logistic regression on physician attributes: average years of experience (*PhyExpYear*), number of training institutions attended (*EducationCnt*), faculty appointment status (*Faculty*), possession of an M.D. degree (*DegreeMD*), presence of any non-health degree (*NonHealthDegree*), and average number of hospital affiliations (*HospAffiliationCnt*). Based on the resulting propensity scores, we perform 1:1 nearest-neighbor matching without replacement, imposing a tight caliper (0.005) and restricting the sample to the region of common support. This yields a matched sample of treated and control physicians with standardized percentage bias for all covariates reduced to near-zero and substantially overlapping propensity score distributions (Figure E.2, Table E.1). Re-estimating the main models on the matched sample (Table E.2, columns (4)-(6)) produces estimates highly consistent with the baseline ($\beta = 0.0643$; focal -0.0131 ; related 0.0065), indicating that our conclusions are robust to reweighting toward physicians with comparable training and practice characteristics.

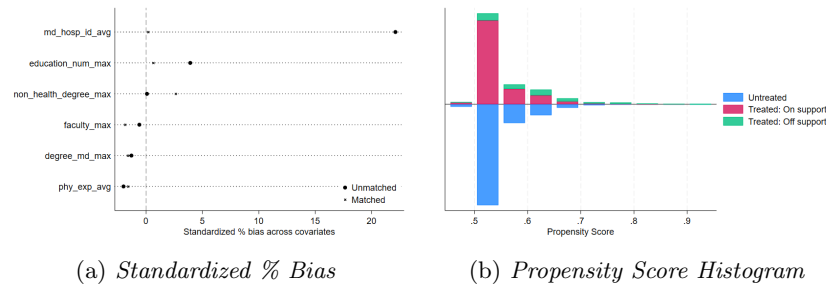


Figure E.2 (a) % Bias Before and After PSM (b) Propensity Score Histogram by Treatment Status

Interpretation. We present RAED as benchmarking robustness evidence rather than as a replacement for the high-dimensional fixed-effects specification. The consistency of both the cost association and experience-moderation pattern across patient- and physician-level RAED, using a benchmark that accounts for discrepancies predicted by observed admission- and final-diagnosis information, reduces concerns that the main findings are driven by differences in observed case mix or physician characteristics.

Table E.1 Covariate Balance Before and After Matching

	Before Matching				After Matching			
	Treated	Control	% Bias	t-stat	Treated	Control	% Bias	t-stat
<i>PhyExpYear</i>	13.889	14.100	-2.0	-1.21	13.822	13.990	-1.6	-0.92
<i>EducationCnt</i>	1.252	1.226	3.9	2.35*	1.226	1.221	0.6	0.39
<i>Faculty</i>	0.211	0.214	-0.6	-0.36	0.206	0.213	-1.9	-1.07
<i>DegreeMD</i>	0.707	0.713	-1.3	-0.78	0.707	0.714	-1.6	-0.92
<i>NonHealthDegree</i>	0.110	0.110	0.1	0.05	0.119	0.110	2.6	1.48
<i>HospAffiliationCnt</i>	1.379	1.243	22.1	13.13***	1.229	1.228	0.2	0.13

Notes. Covariate balance is evaluated using standardized percentage bias and two-sample t-tests before and after matching. Prior to matching, the treated and control groups include 8,792 and 10,753 physicians, respectively. After matching, the sample consists of 6,570 physicians in each group. The reduction in sample size following matching reflects missing values in physician characteristics used in the PSM procedure. The number of observations in the estimation models may vary due to additional missing values in fixed effects or control variables. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table E.2 Patient-Level: Risk-Adjusted Excessive Discrepancies and Propensity Score Matching

DV: $\ln(\text{TotalCost})$	Risk-Adjusted <i>DiagDisc</i> (RAED)			Matched Sample (PSM)		
	(1)	(2)	(3)	(4)	(5)	(6)
<i>DiagDisc</i>	0.0502*** (0.0037)	0.0498*** (0.0037)	0.0489*** (0.0037)	0.0655*** (0.0023)	0.0649*** (0.0023)	0.0643*** (0.0023)
$\ln(\text{PhyFocalExp})$		-0.0129*** (0.0014)	-0.0055* (0.0023)		-0.0135*** (0.0017)	-0.0075*** (0.0019)
<i>DiagDisc</i> $\times$ $\ln(\text{PhyFocalExp})$			-0.0152*** (0.0030)			-0.0131*** (0.0017)
$\ln(\text{PhyRelatedExp})$		0.0062*** (0.0007)	0.0029* (0.0014)		0.0061*** (0.0008)	0.0029* (0.0011)
<i>DiagDisc</i> $\times$ $\ln(\text{PhyRelatedExp})$			0.0064** (0.0019)			0.0065*** (0.0011)
Controls	✓	✓	✓	✓	✓	✓
Fixed Effects	✓	✓	✓	✓	✓	✓
Observations	5,724,263	5,724,263	5,724,263	2,786,379	2,786,379	2,786,379
Physicians	19,535	19,535	19,535	10,721	10,721	10,721
Hospitals	223	223	223	222	222	222
R^2	0.508	0.508	0.508	0.519	0.520	0.520

Notes. Columns (1)-(3) use continuous encounter-level *RAED*, labeled *DiagDisc*. Columns (4)-(6) use the main binary *DiagDisc* with the matched sample. Standard errors in parentheses are clustered by hospitals. + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table E.3 Physician-Level Analysis: Risk-Adjusted Excessive Discrepancies (RAED) and $\ln(\text{TotalCost})$

DV: $\ln(\text{TotalCost})$	(1)	(2)	(3)
<i>DiagDisc</i>	0.0982*** (0.0182)	0.0803*** (0.0145)	0.1253*** (0.0369)
$\ln(\text{PhyFocalExp})$		-0.0308** (0.0089)	-0.1523*** (0.0160)
<i>DiagDisc</i> $\times$ $\ln(\text{PhyFocalExp})$			-0.2593*** (0.0342)
$\ln(\text{PhyRelatedExp})$		0.0095* (0.0038)	0.0986*** (0.0117)
<i>DiagDisc</i> $\times$ $\ln(\text{PhyRelatedExp})$			0.1625*** (0.0229)
Controls	✓	✓	✓
Hospital FE	✓	✓	✓
Year and Qtr FE	✓	✓	✓
Observations	163,710	163,710	163,710
Hospitals	221	221	221
R^2	0.569	0.570	0.756

Notes. Reports physician-panel models using aggregated *RAED*, labeled *DiagDisc*. Standard errors in parentheses are clustered by hospitals. + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table E.4 Hospital-Level Analysis: Risk-Adjusted Excessive Discrepancies (RAED) and $\ln(\text{TotalCost})$

DV: $\ln(\text{TotalCost})$	(1)
<i>DiagDisc</i>	0.0926** (0.0354)
Controls	✓
Hospital FE	✓
Year and Qtr FE	✓
Observations	3,169
Hospitals	222
R^2	0.9389

Notes. Reports hospital-period estimates of aggregated *RAED*, labeled *DiagDisc*, and log total cost. Standard errors in parentheses are clustered by hospitals. ⁺ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix F: Exploratory ILI-Based 2SLS Analysis

F.1. Purpose and Interpretive Limits

The main paper uses local influenza-like-illness (ILI) burden as contextual evidence on operating environments in which the diagnostic-discrepancy flag is more likely to appear. This appendix reports an exploratory two-stage least-squares (2SLS) specification that uses HSA-quarter ILI burden as a source of variation in *DiagDisc*.

A causal interpretation of these estimates would require HSA-quarter ILI burden to affect episode-level cost only through *DiagDisc*. This exclusion restriction cannot be verified and may be violated because local ILI burden can affect bed availability, staffing strain, discharge timing, patient mix, testing intensity, and other operating conditions directly. In addition, instrumenting a retrospective claims-based flag changes the estimand from the paper’s primary conditional association to variation induced by the contextual measure. Accordingly, the second-stage estimates reported in this appendix are exploratory. They are not used to identify the paper’s primary estimand, validate a causal mechanism, or support managerial recommendations.

F.2. Contextual Measure and First-Stage Specification

For patient i treated by physician j in hospital service area h and quarter t , *ILIRateHSA* is defined as the leave-one-out share of other admissions in the same HSA-quarter with an influenza-like-illness (ILI) admitting diagnosis. The leave-one-out construction prevents the focal episode from mechanically contributing to its own contextual measure. The first-stage specification is:

$$\text{DiagDisc}_{ijt} = \alpha + \tau \cdot \text{ILIRateHSA}_{iht} + \gamma Z_{it} + \delta_j + \lambda_t + \phi_e + \mu_{ijt}, \quad (\text{F.1})$$

The model includes the same patient controls and fixed effects as the main specification, together with HSA-quarter total admissions and ILI-coded discharges. The coefficient on *ILIRateHSA* is positive and statistically significant in both the full and restricted samples. The restricted sample excludes ILI index episodes and physicians practicing at multiple hospitals. Consistent with the interpretation in the main text, this first-stage estimate is treated as contextual evidence that the diagnostic-discrepancy flag appears more frequently during periods of higher local ILI-related workload or diagnostic pressure.

F.3. Exploratory Second-Stage Specification

For completeness, the exploratory second-stage equation is:

$$y_{ijt} = \alpha + \beta^{IV} \cdot \widehat{\text{DiagDisc}}_{ijt} + \gamma Z_{it} + \delta_j + \lambda_t + \phi_e + \eta_{ijt}, \quad (\text{F.2})$$

where $\widehat{\text{DiagDisc}}_{ijt}$ is the predicted value from Equation (F.1) and β^{IV} is the exploratory 2SLS estimate for the association between cross-category diagnostic misalignment and total costs under the instrumental variable (IV) assumptions. The same comprehensive set of fixed effects in Equation (1) (Physician, Year and Quarter, CCS, CCS-Year-Quarter, Hospital, Payer, Race, Urban Rural, and ED Admission Hour) is included in both stages. The coefficient β^{IV} has a causal interpretation only under the standard IV assumptions, including independence and exclusion. First-stage strength establishes relevance only and does not establish the exclusion restriction.

Table F.1 Exploratory 2SLS: Association Between Diagnostic Discrepancy and Total Cost per Hospitalization

Sample:	2SLS	
	Full Sample (1)	Restricted Sample (2)
Second Stage (DV: $\ln(\text{TotalCost})$)		
<i>DiagDisc</i>	0.0746** (0.0276)	0.0857* (0.0393)
First Stage (DV: <i>DiagDisc</i>)		
<i>ILIRateHSA</i>	8.0257*** (0.5798)	8.4140*** (0.7683)
Controls	✓	✓
Fixed Effects	✓	✓
Observations	5,724,257	3,066,110
Physicians	19,535	12,145
Hospitals	223	218
Underid. (KP <i>rk</i> LM <i>p</i> -value)	<0.001	<0.001
Stock-Yogo Weak id. (KP <i>rk</i> Wald <i>F</i> -stat.)	191.58	119.94
Endogeneity test (χ^2 <i>p</i> -value)	0.7399	0.5475

Notes: Model (2) restricted sample excludes ILI patients and multi-site physicians. In both models, we additionally control for HSA-quarter total admission volume (*TotalAdmCt*) and the number of discharges with ILI-coded diagnosis (*ILIAdmCt*). The underidentification test and Wald *F*-statistic for weak identification are based on Kleibergen and Paap (2006). The LM test checks if the instruments are relevant (i.e., correlated with endogenous regressors), rejecting the null of underidentification, supporting full column rank. Weak identification arises from weak instrument correlation. Stock and Yogo (2005) provide a critical value of 16.38, with all models' *F*-statistics exceeding this, mitigating weak identification concerns. For the endogeneity test, the null is that endogenous regressors may be treated as exogenous; Standard errors in parentheses are clustered by HSAs because the instrument varies at the HSA-quarter level. + $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

F.4. Diagnostics and Supplementary Specifications

We report several diagnostics to characterize the exploratory 2SLS specification. First, the first-stage relationship is statistically strong. Kleibergen–Paap LM statistics reject underidentification ($p < 0.001$), and the Kleibergen–Paap *rk* Wald *F*-statistics of 191.58 and 119.94 exceed the Stock–Yogo 10% critical value of 16.38 (Stock and Yogo 2005). These statistics establish relevance of *ILIRateHSA* for *DiagDisc*, but they do not address the exclusion restriction.

Second, the specification includes controls intended to account for local ILI-related volume and burden. Specifically, the models control for HSA-quarter total admissions (*TotalAdmCt*; mean 12,837, SD 13,033) and ILI-coded discharges (*ILIAdmCt*; mean 671, SD 656), in addition to the patient covariates and fixed effects described above. We also report a restricted sample that excludes ILI index episodes and physicians practicing at multiple hospitals. This restriction reduces direct outbreak-related channels for the focal episode and limits potential cross-hospital diffusion of diagnostic practices. These design features improve comparability across observations but do not by themselves validate the exclusion restriction.

Third, we examine whether observable patient characteristics explain variation in *ILIRateHSA*. Table F.2 reports regressions of *ILIRateHSA* on observable patient characteristics, including Elixhauser comorbidity index, age, gender, payer type, zip-code and state income quartiles, race, and urban-rural indicators, along with hospital, year-quarter, and ED admission-hour fixed effects. The coefficients are mostly small and statistically insignificant, with only a few exceptions, suggesting that variation in *ILIRateHSA* is not strongly explained by observable patient mix after conditioning on the included fixed effects. This pattern is consistent with interpreting *ILIRateHSA* as a contextual diagnostic-workload measure rather than as a proxy for systematic differences in observed patient severity or socioeconomic composition.

Finally, we examine two supplementary specifications. We first re-estimate the 2SLS models using Hospital Referral Region (HRR)-based ILI measures in place of the HSA-based measure and obtain similar estimates, suggesting that the results are not sensitive to the geographic aggregation used to construct the contextual ILI measure. We then estimate overidentified 2SLS models that include both ILI rate and ILI volume measures. Hansen *J* tests fail to reject the overidentifying restrictions at conventional significance levels, and the second-stage estimates remain similar.

Taken together, these diagnostics indicate that local ILI burden provides strong contextual variation

associated with the appearance of the diagnostic-discrepancy flag. However, because local ILI burden may also affect costs through operating conditions other than *DiagDisc*, and the instrumented variation differs from the primary conditional association, we report the 2SLS estimates as exploratory evidence for completeness rather than as causal estimates or validation of the primary association.

Table F.2 Post-hoc Analysis: Factors Associated with ILIRateHSA

DV: <i>ILIRateHSA</i>	HDFE OLS	
<i>ElixCI</i>	0.0002 ⁺	(0.0001)
<i>TransferIn</i>	0.0019	(0.0021)
<i>AdmWeekend</i>	-0.0001	(0.0001)
<i>Age</i>	0.0000	(0.0000)
<i>Female</i>	0.0001	(0.0001)
<i>MedIncState</i>		
Q2	0.0003	(0.0019)
Q3	-0.0038	(0.0033)
Q4	-0.0070	(0.0054)
<i>ZipIncQtrl</i>		
Q2	0.0032	(0.0031)
Q3	0.0087	(0.0053)
Q4	0.0099	(0.0064)
<i>Payer</i>		
Medicaid	0.0006	(0.0005)
Private	0.0000	(0.0004)
Self-pay	0.0004	(0.0004)
No Charge	-0.0008	(0.0009)
Other	0.0005	(0.0006)
<i>Race</i>		
Black	0.0004	(0.0003)
Hispanic	0.0008 ⁺	(0.0004)
Asian	0.0015 [*]	(0.0008)
Native American	0.0020 ⁺	(0.0012)
Other	-0.0021	(0.0022)
<i>UrbanRural</i>		
Small Metropolitan	0.0018	(0.0015)
Micropolitan	-0.0070 ⁺	(0.0041)
Not Metropolitan	0.0002	(0.0011)
Controls	✓	
Other FEs		
Observations	5,722,933	
Clusters (HSA)	108	
<i>R</i> ²	0.855	

Notes. Regresses assigned HSA-quarter *ILIRateHSA* on observed characteristics with hospital, year-quarter, and ED-hour fixed effects. The reference categories are Q1 (*MedIncState*), Q1 (*ZipIncQtrl*), Medicare (*Payer*), White (*Race*), Large Metropolitan (*UrbanRural*). Small, mostly insignificant coefficients indicate limited explanatory power of observable patient composition. Standard errors in parentheses are clustered by HSAs. ⁺ $p < 0.10$, ^{*} $p < 0.05$, ^{**} $p < 0.01$, ^{***} $p < 0.001$

Appendix G: Robustness Checks

G.1. Alternative Diagnostic Discrepancies: Secondary Diagnoses and DRG Family

Table G.1 Robustness Checks: Alternative Diagnostic Discrepancies on Secondary Diagnoses and DRG Family

Alternative <i>DiagDisc</i> :	<i>DD_Secondary</i>			<i>DD_DRGFamily</i>		
DV: $\ln(\text{TotalCost})$	(1)	(2)	(3)	(4)	(5)	(6)
<i>DiagDisc</i>	0.0188*** (0.0019)	0.0185*** (0.0019)	0.0177*** (0.0020)	0.0723*** (0.0021)	0.0718*** (0.0021)	0.0703*** (0.0021)
$\ln(\text{PhyFocalExp})$		-0.0131*** (0.0015)	-0.0111*** (0.0016)		-0.0122*** (0.0015)	-0.0082*** (0.0018)
<i>DiagDisc</i> $\times$ $\ln(\text{PhyFocalExp})$			-0.0063*** (0.0012)			-0.0102*** (0.0019)
$\ln(\text{PhyRelatedExp})$		0.0062*** (0.0007)	0.0059*** (0.0008)		0.0059*** (0.0007)	0.0037*** (0.0009)
<i>DiagDisc</i> $\times$ $\ln(\text{PhyRelatedExp})$			0.0011 (0.0009)			0.0051*** (0.0011)
Controls	✓	✓	✓	✓	✓	✓
Fixed Effects	✓	✓	✓	✓	✓	✓
Observations	5,724,263	5,724,263	5,724,263	5,724,263	5,724,263	5,724,263
Physicians	19,535	19,535	19,535	19,535	19,535	19,535
Hospitals	223	223	223	223	223	223
R^2	0.508	0.508	0.508	0.509	0.509	0.509

Notes. Columns (1)-(3) require no match among principal or secondary diagnoses. Columns (4)-(6) use DRG-family alignment. Standard errors in parentheses are clustered by hospitals. ⁺ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

G.2. Alternative Diagnostic Discrepancies based on Multi-Level CCS

Table G.2 Robustness: Alternative Diagnostic Discrepancy Metrics based on Multi-Level CCS

Multi-Level CCS: Alternative <i>DiagDisc</i> : DV: $\ln(\text{TotalCost})$	<i>Discrepancy based on Principal Diagnosis Code Only</i>					
	Level 1 <i>DDMultiCCSlv1_Pri</i> (1)		Level 2 <i>DDMultiCCSlv2_Pri</i> (2)		Level 3 <i>DDMultiCCSlv3_Pri</i> (3)	
<i>DiagDisc</i>	0.0662*** (0.0021)		0.0644*** (0.0021)		0.0654*** (0.0020)	
$\ln(\text{PhyFocalExp})$	-0.0101*** (0.0016)		-0.0054** (0.0017)		-0.0063*** (0.0016)	
<i>DiagDisc</i> $\times$ $\ln(\text{PhyFocalExp})$	-0.0069*** (0.0013)		-0.0158*** (0.0017)		-0.0138*** (0.0013)	
$\ln(\text{PhyRelatedExp})$	0.0053*** (0.0008)		0.0022* (0.0009)		0.0027** (0.0009)	
<i>DiagDisc</i> $\times$ $\ln(\text{PhyRelatedExp})$	0.0020* (0.0008)		0.0080*** (0.0010)		0.0067*** (0.0008)	
R^2	0.509		0.512		0.509	
Multi-Level CCS: Alternative <i>DiagDisc</i> : DV: $\ln(\text{TotalCost})$	<i>Encompassing Principal and Secondary Diagnosis Codes</i>					
	Level 1 <i>DDMultiCCSlv1_Sec</i> (4)		Level 2 <i>DDMultiCCSlv2_Sec</i> (5)		Level 3 <i>DDMultiCCSlv3_Sec</i> (6)	
<i>DiagDisc</i>	0.0050* (0.0021)		0.0063** (0.0020)		0.0264*** (0.0019)	
$\ln(\text{PhyFocalExp})$	-0.0124*** (0.0015)		-0.0113*** (0.0015)		-0.0113*** (0.0015)	
<i>DiagDisc</i> $\times$ $\ln(\text{PhyFocalExp})$	-0.0072*** (0.0016)		-0.0071*** (0.0013)		-0.0079*** (0.0012)	
$\ln(\text{PhyRelatedExp})$	0.0062*** (0.0007)		0.0058*** (0.0007)		0.0058*** (0.0008)	
<i>DiagDisc</i> $\times$ $\ln(\text{PhyRelatedExp})$	0.0006 (0.0012)		0.0017+ (0.0009)		0.0022* (0.0009)	
R^2	0.508		0.511		0.508	
Controls	✓		✓		✓	
Fixed Effects	✓		✓		✓	
Observations	5,724,263		5,485,019		4,990,652	
Physicians	19,535		19,431		18,998	
Hospitals	223		223		221	

Notes. Columns (1)-(3) compare principal diagnoses at CCS Levels 1-3. Columns (4)-(6) use secondary-inclusive definitions. The cost association and opposing experience interactions generally persist, with smaller inclusive-measure estimates. Sample sizes vary by mapping coverage. Level 4 is omitted for high missingness. Standard errors in parentheses are clustered by hospitals. ⁺ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

G.3. Alternative Diagnostic Discrepancies: Based on First 3-Digit of ICD-9 Codes

Table G.3 Robustness Checks: Alternative Diagnostic Discrepancy based on First 3-Digit of ICD-9 Codes

Alternative <i>DiagDisc</i> :	<i>DD_Principal</i>			<i>DD_Secondary</i>		
DV: $\ln(\text{TotalCost})$	(1)	(2)	(3)	(4)	(5)	(6)
<i>DiagDisc</i>	0.0671*** (0.0021)	0.0667*** (0.0021)	0.0664*** (0.0020)	0.0228*** (0.0018)	0.0226*** (0.0018)	0.0224*** (0.0018)
$\ln(\text{PhyFocalExp})$		-0.0122*** (0.0014)	-0.0052** (0.0018)		-0.0131*** (0.0015)	-0.0114*** (0.0016)
<i>DiagDisc</i> $\times$ $\ln(\text{PhyFocalExp})$			-0.0133*** (0.0013)			-0.0044*** (0.0011)
$\ln(\text{PhyRelatedExp})$		0.0060*** (0.0007)	0.0023* (0.0009)		0.0062*** (0.0007)	0.0059*** (0.0008)
<i>DiagDisc</i> $\times$ $\ln(\text{PhyRelatedExp})$			0.0066*** (0.0008)			0.0008 (0.0008)
Controls	✓	✓	✓	✓	✓	✓
Fixed Effects	✓	✓	✓	✓	✓	✓
Observations	5,724,263	5,724,263	5,724,263	5,724,263	5,724,263	5,724,263
Physicians	19,535	19,535	19,535	19,535	19,535	19,535
Hospitals	223	223	223	223	223	223
R^2	0.509	0.509	0.509	0.508	0.508	0.508

Notes. Columns (1)-(3) compare the first three ICD-9-CM digits of admitting and principal diagnoses. Columns (4)-(6) credit secondary matches. Standard errors in parentheses are clustered by hospitals. ⁺ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

G.4. Alternative Diagnostic Discrepancies: Multi-scheme

Table G.4 Multi-scheme Diagnostic Discrepancy Metric

DV: $\ln(\text{TotalCost})$	(1)	(2)	(3)
<i>DiagDisc</i> _{Robust}	0.0319*** (0.0020)	0.0315*** (0.0020)	0.0265*** (0.0021)
$\ln(\text{PhyFocalExp})$		-0.0126*** (0.0014)	-0.0096** (0.0016)
<i>DiagDisc</i> _{Robust} $\times$ $\ln(\text{PhyFocalExp})$			-0.0062*** (0.0014)
$\ln(\text{PhyRelatedExp})$		0.0061*** (0.0007)	0.0049*** (0.0009)
<i>DiagDisc</i> _{Robust} $\times$ $\ln(\text{PhyRelatedExp})$			0.0024* (0.0009)
Controls	✓	✓	✓
Fixed Effects	✓	✓	✓
Observations	5,485,019	5,485,019	5,485,019
Physicians	19,431	19,431	19,431
Hospitals	223	223	223
R^2	0.5121	0.5122	0.5122

Notes. This table reports robustness analyses using a multi-scheme measure of admission-to-principal diagnostic discordance. *DiagDisc*_{Robust} equals one when the admitting and principal diagnoses are classified into different diagnostic groups across major alternative coding schemes considered, including single-level CCS, first-three-digit ICD-9 grouping, DRG family, and multi-level CCS groupings up to level 2, along with consideration of secondary diagnoses. A total of nine coding variants were considered. *DiagDisc*_{Robust} flags 19.26% of non-missing encounters. It therefore captures the most conservative set of diagnosis pairs that remain discordant across plausible taxonomies. Standard errors in parentheses are clustered by hospitals. ⁺ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

G.5. Subsample on Sepsis Patients

Table G.5 Robustness Checks: Alternative Sample on Sepsis Patients

Subsample:	Excluding Sepsis			Sepsis Only		
DV: $\ln(\text{TotalCost})$	(1)	(2)	(3)	(4)	(5)	(6)
<i>DiagDisc</i>	0.0676*** (0.0021)	0.0673*** (0.0021)	0.0660*** (0.0021)	0.0045 (0.0051)	0.0044 (0.0051)	0.0082 (0.0053)
$\ln(\text{PhyFocalExp})$		-0.0094*** (0.0015)	-0.0036* (0.0018)		-0.0072** (0.0025)	-0.0034 (0.0040)
$\text{DiagDisc} \times \ln(\text{PhyFocalExp})$			-0.0123*** (0.0018)			-0.0051 (0.0036)
$\ln(\text{PhyRelatedExp})$		0.0044*** (0.0007)	0.0010 (0.0010)		0.0093*** (0.0019)	0.0114*** (0.0032)
$\text{DiagDisc} \times \ln(\text{PhyRelatedExp})$			0.0068*** (0.0010)			-0.0031 (0.0035)
Controls	✓	✓	✓	✓	✓	✓
Fixed Effects	✓	✓	✓	✓	✓	✓
Observations	5,387,983	5,387,983	5,387,983	333,793	333,793	333,793
Physicians	19,386	19,386	19,386	8,356	8,356	8,356
Hospitals	223	223	223	206	206	206
R^2	0.499	0.499	0.499	0.470	0.470	0.470

Notes. Columns (1)-(3) exclude sepsis encounters. Columns (4)-(6) retain only sepsis encounters. The cost association and opposing experience interactions remain strong outside sepsis but are attenuated and generally insignificant within sepsis. Thus, the aggregate pattern is concentrated outside the sepsis-only sample and is not explained solely by sepsis encounters. Standard errors in parentheses are clustered by hospitals. ⁺ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

G.6. Alternative Scopes of Physician Focal and Related Experience

Table G.6 Robustness Checks: Principal Diagnosis-Based Alternative Measures of Physician Experience

Code Scope:	Principal Diagnosis					
	CCS			DX (3-Digit)		
	1-Year	Half Year	1-Qtr	1-Year	Half Year	1-Qtr
DV: $\ln(\text{TotalCost})$	(1)	(2)	(3)	(4)	(5)	(6)
<i>DiagDisc</i>	0.0640*** (0.0021)	0.0641*** (0.0021)	0.0641*** (0.0021)	0.0636*** (0.0021)	0.0636*** (0.0021)	0.0638*** (0.0021)
$\ln(\text{PhyFocalExp})$	-0.0073*** (0.0019)	-0.0050** (0.0018)	-0.0049* (0.0020)	-0.0083*** (0.0018)	-0.0073*** (0.0019)	-0.0052* (0.0020)
$\text{DiagDisc} \times \ln(\text{PhyFocalExp})$	-0.0131*** (0.0016)	-0.0145*** (0.0017)	-0.0153*** (0.0020)	-0.0162*** (0.0015)	-0.0178*** (0.0017)	-0.0193*** (0.0019)
$\ln(\text{PhyRelatedExp})$	0.0027* (0.0013)	0.0020* (0.0010)	0.0024* (0.0011)	0.0032** (0.0011)	0.0032** (0.0011)	0.0025* (0.0011)
$\text{DiagDisc} \times \ln(\text{PhyRelatedExp})$	0.0084*** (0.0011)	0.0075*** (0.0010)	0.0076*** (0.0010)	0.0084*** (0.0010)	0.0081*** (0.0009)	0.0078*** (0.0009)
Controls	✓	✓	✓	✓	✓	✓
Fixed Effects	✓	✓	✓	✓	✓	✓
Observations	5,724,263	5,724,263	5,724,263	5,724,263	5,724,263	5,724,263
Physicians	19,535	19,535	19,535	19,535	19,535	19,535
Hospitals	223	223	223	223	223	223
R^2	0.509	0.509	0.509	0.510	0.510	0.509

Notes. Re-estimates joint moderation using principal-diagnosis experience over one year, six months, or one quarter at CCS or three-digit ICD levels. Measures are logged and mean-centered. Column (2) is the main specification. Standard errors in parentheses are clustered by hospitals. ⁺ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table G.7 Robustness Checks: Admitting Diagnosis-Based Alternative Measures of Physician Experience

Code Scope: Time Scope DV: $\ln(\text{TotalCost})$	Admitting Diagnosis					
	CCS			DX (3-Digit)		
	1-Year (1)	Half Year (2)	1-Qtr (3)	1-Year (4)	Half Year (5)	1-Qtr (6)
<i>DiagDisc</i>	0.0652*** (0.0024)	0.0652*** (0.0024)	0.0649*** (0.0024)	0.0645*** (0.0025)	0.0644*** (0.0025)	0.0640*** (0.0025)
$\ln(\text{PhyFocalExp})$	0.0010 (0.0015)	0.0012 (0.0016)	0.0018 (0.0018)	0.0018 (0.0014)	0.0021 (0.0015)	0.0028+ (0.0017)
$\text{DiagDisc} \times \ln(\text{PhyFocalExp})$	-0.0054** (0.0018)	-0.0061** (0.0019)	-0.0072*** (0.0021)	-0.0032+ (0.0017)	-0.0037* (0.0018)	-0.0046* (0.0020)
$\ln(\text{PhyRelatedExp})$	-0.0006 (0.0010)	-0.0004 (0.0010)	-0.0003 (0.0009)	-0.0007 (0.0009)	-0.0006 (0.0009)	-0.0005 (0.0009)
$\text{DiagDisc} \times \ln(\text{PhyRelatedExp})$	0.0050*** (0.0013)	0.0053*** (0.0012)	0.0056*** (0.0011)	0.0032** (0.0012)	0.0036** (0.0012)	0.0042*** (0.0011)
Controls	✓	✓	✓	✓	✓	✓
Fixed Effects	✓	✓	✓	✓	✓	✓
Observations	5,724,263	5,724,263	5,724,263	5,724,263	5,724,263	5,724,263
Physicians	19,535	19,535	19,535	19,535	19,535	19,535
Hospitals	223	223	223	223	223	223
R^2	0.509	0.509	0.509	0.509	0.509	0.509

Notes. Constructs experience from admitting diagnoses over one year, six months, or one quarter at CCS or three-digit ICD levels. Measures are logged and mean-centered. Standard errors in parentheses are clustered by hospitals. + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

G.7. Relative Focal Experience and Robustness to Prevalence and Volume

Table G.8 Robustness to Diagnosis Prevalence, Hospital Volume, and Relative Focal Experience

DV: $\ln(\text{TotalCost})$	Main Results (1)	Additional Controls (2) (3)		Relative FocalExp (4) (5)	
<i>DiagDisc</i>	0.0641*** (0.0021)	0.0597*** (0.0022)	0.0598*** (0.0022)	0.0642*** (0.0021)	0.0586*** (0.0023)
$\ln(\text{PhyFocalExp})$	-0.0050** (0.0018)	-0.0049** (0.0016)	-0.0051** (0.0016)		
$\text{DiagDisc} \times \ln(\text{PhyFocalExp})$	-0.0145*** (0.0017)	-0.0140*** (0.0017)	-0.0135*** (0.0020)		
<i>RelFocal</i>				-0.0065*** (0.0019)	-0.0054** (0.0016)
$\text{DiagDisc} \times \text{RelFocal}$				-0.0119*** (0.0021)	-0.0128*** (0.0021)
$\ln(\text{PhyRelatedExp})$	0.0020* (0.0010)	0.0007 (0.0014)	0.0006 (0.0014)	0.0015+ (0.0008)	0.0006 (0.0014)
$\text{DiagDisc} \times \ln(\text{PhyRelatedExp})$	0.0075*** (0.0010)	0.0136*** (0.0014)	0.0137*** (0.0014)	0.0049*** (0.0009)	0.0137*** (0.0015)
$\ln(\text{CCSPrev})$		0.0333*** (0.0074)	0.0358*** (0.0078)		0.0298*** (0.0077)
$\ln(\text{HospCCSVol})$		-0.0100** (0.0034)	-0.0114** (0.0041)		-0.0110** (0.0041)
$\text{DiagDisc} \times \ln(\text{CCSPrev})$			-0.0048 (0.0032)		-0.0092** (0.0032)
$\text{DiagDisc} \times \ln(\text{HospCCSVol})$			0.0027 (0.0028)		0.0021 (0.0028)
Controls	✓	✓	✓	✓	✓
Fixed Effects	✓	✓	✓	✓	✓
Observations	5,724,263	4,993,715	4,993,715	5,724,253	4,993,705
Physicians	19,535	18,401	18,401	19,535	18,401
Hospitals	223	221	221	223	221
R^2	0.5090	0.5092	0.5092	0.5089	0.5092

Notes. $\ln(\text{CCSPrev})$ measures the number of patients in the focal CCS category at the focal hospital during the physician-experience lookback window, and $\ln(\text{HospCCSVol})$ measures the hospital-level volume in the focal CCS category. *RelFocal* benchmarks a physician's focal experience against peer physicians treating the same CCS category in the same hospital-period, thereby capturing physician-specific diagnostic depth relative to the local service-line environment. Standard errors in parentheses are clustered by hospitals. + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

G.8. Alternative Diagnosis Fixed-Effect Specifications

Table G.9 Robustness Analysis: Association Between Diagnostic Discrepancy and Total Cost Under Alternative Diagnosis Fixed-Effect Specifications, with Moderating Effects of Physicians' Focal and Related Experience

DV:	$\ln(\text{TotalCost})$ (1)	$\ln(\text{TotalCost})$ (2)	$\ln(\text{TotalCost})$ (3)	$\ln(\text{TotalCost})$ (4)	$\ln(\text{TotalCost})$ (5)	$\ln(\text{TotalCost})$ (6)
<i>DiagDisc</i>	0.0641*** (0.0021)	0.0466*** (0.0020)	0.0545*** (0.0020)	0.0844*** (0.0022)	0.0644*** (0.0021)	0.0528*** (0.0016)
$\ln(\text{PhyFocalExp})$	-0.0050** (0.0018)	-0.0045** (0.0016)	-0.0065*** (0.0017)	-0.0088*** (0.0016)	-0.0019 (0.0014)	-0.0046*** (0.0010)
<i>DiagDisc</i> $\times$ $\ln(\text{PhyFocalExp})$	-0.0145*** (0.0017)	-0.0128*** (0.0015)	-0.0120*** (0.0016)	-0.0063*** (0.0014)	-0.0145*** (0.0017)	-0.0073*** (0.0010)
$\ln(\text{PhyRelatedExp})$	0.0020* (0.0010)	0.0017+ (0.0009)	0.0030** (0.0009)	0.0037*** (0.0009)	0.0011 (0.0009)	0.0022** (0.0007)
<i>DiagDisc</i> $\times$ $\ln(\text{PhyRelatedExp})$	0.0075*** (0.0010)	0.0053*** (0.0009)	0.0055*** (0.0009)	0.0039*** (0.0009)	0.0072*** (0.0011)	0.0026*** (0.0007)
Controls	✓	✓	✓	✓	✓	✓
Physician FE	✓	✓	✓	✓	✓	✓
Year and Qtr FE	✓	✓	✓	✓	✓	✓
Hospital FE	✓	✓	✓	✓	✓	✓
Payer FE	✓	✓	✓	✓	✓	✓
Race FE	✓	✓	✓	✓	✓	✓
Urban Rural FE	✓	✓	✓	✓	✓	✓
ED Adm Hour FE	✓	✓	✓	✓	✓	✓
Principal CCS FE	✓	✓	✓	✓	✓	✓
Principal CCS $\times$ Year-Qtr FE	✓	✓	✓	✓	✓	✓
Additional diagnosis fixed effects:						
DRG FE		✓				✓
ICD-9 FE			✓			✓
Admitting CCS FE				✓		✓
Principal CCS $\times$ Hospital FE					✓	✓
Observations	5,724,263	5,724,260	5,723,189	5,724,263	5,720,540	5,719,433
Physicians	19,535	19,535	19,533	19,535	19,517	19,514
Hospitals	223	223	223	223	221	221
R^2	0.509	0.597	0.537	0.514	0.528	0.626

Note: All columns share the same baseline controls and fixed effects; columns (2)-(5) each add one alternative diagnosis-code fixed effect to the column (1) baseline, and (6) includes all four. Standard errors in parentheses are clustered by hospitals. + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

G.9. Alternative Winsorization and Data Trimming

Table G.10 Robustness Checks: Alternative Winsorization and Data Trimming

DV: $\ln(\text{TotalCost})$	Winsor 3% (1)	Winsor 5% (2)	Trim 1% (3)	Trim 3% (4)	Trim 5% (5)	Full Data (6)
<i>DiagDisc</i>	0.0626*** (0.0020)	0.0608*** (0.0020)	0.0618*** (0.0020)	0.0569*** (0.0019)	0.0528*** (0.0017)	0.0646*** (0.0022)
$\ln(\text{PhyFocalExp})$	-0.0051** (0.0017)	-0.0050** (0.0017)	-0.0051** (0.0017)	-0.0056*** (0.0017)	-0.0057*** (0.0015)	-0.0051** (0.0018)
<i>DiagDisc</i> $\times$ $\ln(\text{PhyFocalExp})$	-0.0136*** (0.0016)	-0.0127*** (0.0016)	-0.0132*** (0.0016)	-0.0110*** (0.0016)	-0.0093*** (0.0014)	-0.0150*** (0.0017)
$\ln(\text{PhyRelatedExp})$	0.0018* (0.0009)	0.0017+ (0.0009)	0.0018* (0.0009)	0.0018* (0.0008)	0.0019* (0.0008)	0.0022* (0.0010)
<i>DiagDisc</i> $\times$ $\ln(\text{PhyRelatedExp})$	0.0073*** (0.0010)	0.0070*** (0.0009)	0.0072*** (0.0010)	0.0063*** (0.0009)	0.0053*** (0.0008)	0.0075*** (0.0010)
Controls	✓	✓	✓	✓	✓	✓
Fixed Effects	✓	✓	✓	✓	✓	✓
Observations	5,724,263	5,724,263	5,631,225	5,410,784	5,185,418	5,724,263
Physicians	19,535	19,535	19,403	19,176	18,980	19,535
Hospitals	223	223	223	222	221	223
R^2	0.507	0.502	0.488	0.451	0.418	0.508

Notes. Tests alternative winsorization, trimming, and the full cost distribution. The main analysis uses 1% winsorization. Observation counts decline with trimming because extreme cases are removed. Standard errors in parentheses are clustered by hospitals. + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix H: Balance Diagnostics and Residual Disease Complexity

Claims data do not reveal patients' underlying true diagnoses or full clinical presentation at admission. The principal diagnosis is an administrative final code that reflects clinical evolution, documentation, and coding practices. Accordingly, conditioning on principal CCS reduces broad diagnostic heterogeneity but cannot eliminate unobserved differences in severity or complexity within a diagnosis category.

Table H.1 shows that observable balance improves substantially within principal-CCS categories. The standardized differences for *ElixCI*, age, and *NPR* decline from 0.238, 0.168, and 0.154 overall to 0.096, 0.043, and 0.008 within CCS, respectively, and most other covariates are close to zero. The stability of the estimates under admitting-CCS, DRG, and hospital-by-principal-CCS fixed effects, together with the post-matching balance diagnostics, reduces concern that the main results are explained solely by observed patient-complexity differences. However, these checks cannot eliminate residual episode complexity. We therefore interpret diagnostic discrepancy as an adjusted marker of operational burden that may partly reflect residual clinical complexity, rather than as an independently identified source of higher resource use.

Table H.1 Covariate Balance Diagnostics by Diagnostic Discrepancy Status

Variable	Mean		Standardized Difference	
	Discrepancy	No Discrepancy	Overall	Within-CCS
<i>ElixCI</i>	3.549	3.013	0.238	0.096
<i>TransferIn</i>	0.058	0.056	0.007	-0.004
<i>Weekend</i>	0.259	0.255	0.009	0.013
<i>ORProc</i>	0.163	0.156	0.017	-0.003
<i>Age</i>	60.835	57.041	0.168	0.043
<i>Female</i>	0.533	0.526	0.016	0.009
<i>MedIncStQ</i>	2.282	2.268	0.013	0.001
<i>ZipIncQtl</i>	1.897	1.898	-0.001	-0.012
<i>HospBusyness</i>	0.688	0.689	-0.004	0.001
<i>NDX</i>	11.911	9.908	0.308	0.112
<i>NPR</i>	1.423	1.090	0.154	0.008

Note: An absolute standardized difference below 0.10 indicates negligible imbalance. The first two columns report covariate means by discrepancy status. The overall standardized difference is $(\bar{x}_1 - \bar{x}_0) / \sqrt{(s_1^2 + s_0^2)/2}$ on the full sample, where subscripts 1 and 0 denote the discrepancy and no-discrepancy groups, respectively. The within-CCS standardized difference is the coefficient on the discrepancy indicator from a regression absorbing principal-CCS fixed effects, scaled by the pooled within-CCS residual standard deviation.

Appendix I: Subsample Analysis: Patient and Hospital Characteristics

Table I.1 Subsample Analysis: High vs. Low Patient Complexity, Private vs. Public Hospitals, Urban vs. Rural

Subsample:	<i>ElixCI</i> High	<i>ElixCI</i> Low	PrivateHosp	PublicHosp	Urban	Rural
DV: $\ln(\text{TotalCost})$	(1)	(2)	(3)	(4)	(5)	(6)
<i>DiagDisc</i>	0.0500*** (0.0026)	0.0686*** (0.0022)	0.0660*** (0.0033)	0.0615*** (0.0026)	0.0644*** (0.0020)	0.0571*** (0.0071)
$\ln(\text{PhyFocalExp})$	-0.0066** (0.0024)	-0.0047** (0.0017)	-0.0017 (0.0023)	-0.0069** (0.0024)	-0.0054** (0.0018)	0.0044 (0.0044)
$\text{DiagDisc} \times \ln(\text{PhyFocalExp})$	-0.0097*** (0.0019)	-0.0148*** (0.0018)	-0.0151*** (0.0020)	-0.0146*** (0.0024)	-0.0145*** (0.0017)	-0.0148*** (0.0039)
$\ln(\text{PhyRelatedExp})$	0.0040** (0.0013)	0.0013 (0.0010)	0.0007 (0.0014)	0.0030* (0.0012)	0.0022* (0.0010)	0.0000 (0.0033)
$\text{DiagDisc} \times \ln(\text{PhyRelatedExp})$	0.0053*** (0.0013)	0.0078*** (0.0011)	0.0075*** (0.0015)	0.0072*** (0.0013)	0.0075*** (0.0010)	0.0052 (0.0032)
Controls	✓	✓	✓	✓	✓	✓
Fixed Effects	✓	✓	✓	✓	✓	✓
Observations	1,619,970	4,101,465	2,211,085	3,511,744	5,423,908	297,660
R^2	0.450	0.486	0.502	0.516	0.511	0.516

Notes. Estimates joint moderation for patient-complexity, ownership, and urban-rural subsamples. Standard errors in parentheses are clustered by hospitals. + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix J: Post-hoc Analysis: Factors Associated with Diagnostic Discrepancy

We examine factors potentially associated with diagnostic discrepancies. Although our data cannot fully capture all determinants, such as detailed symptom descriptions at admission or ED triage processes, we assess whether discrepancies are related to patient or physician characteristics. To do so, we estimate a linear probability model (LPM) with a binary dependent variable (*DiagDisc*) to obtain average marginal effects. The LPM is appropriate given our large sample and a diagnostic discrepancy rate of 52.44%, for which it provides unbiased and consistent estimates (Wooldridge 2010).

Table J.1 reports the results. None of the physician characteristics, including $\ln(\text{PhyFocalExp})$, $\ln(\text{PhyRelatedExp})$, years of experience (*ExpYr*), training breadth (*MedInstCt*), faculty status (*Faculty*), M.D. degree (*DegreeMD*), non-health degree (*NonHealthDeg*), or multiple hospital affiliations (*MultiHosp*), are statistically significant at the 5% level. In contrast, several patient characteristics, such as higher clinical complexity, advanced age, and absence of inter-provider transfer, are significantly associated with higher discrepancy rates. Because these factors are largely difficult to modify directly, the findings suggest that the operational relevance of the flag lies in understanding variation in resource-use patterns once cross-category misalignment is observed. We examine this variation by focusing on physicians' focal and related experience.

Table J.1 Post-hoc Analysis: Factors Associated with Diagnostic Discrepancy

DV: <i>DiagDisc</i>	LPM	
$\ln(\text{PhyFocalExp})$	-0.0023	(0.0016)
$\ln(\text{PhyRelatedExp})$	0.0005	(0.0007)
<i>ExpYr</i>	0.0010	(0.0016)
<i>MedInstCt</i>	-0.0112	(0.0259)
<i>Faculty</i>	-0.0408	(0.0439)
<i>DegreeMD</i>	-0.0329	(0.0734)
<i>NonHealthDeg</i>	-0.0331	(0.0290)
<i>MultiHosp</i>	-0.0017	(0.0013)
<i>ElixCI</i>	0.0009***	(0.0002)
<i>TransferIn</i>	-0.0022*	(0.0009)
<i>AdmWeekend</i>	0.0001	(0.0004)
<i>Age</i>	0.0001***	(0.0000)
<i>Female</i>	-0.0012**	(0.0004)
<i>EDWorkloadRatio</i>	-0.0064+	(0.0034)
Controls	✓	
Fixed Effects	✓	

Notes. Reports a linear probability model for *DiagDisc*. Coefficients are probability changes in decimal units. Physician human-capital and affiliation variables are insignificant at 5%, whereas comorbidity, age, transfer status, and sex show small associations, indicating a primarily case- and context-related condition. Standard errors in parentheses are clustered by hospitals. + $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

References

- Buck, C. J. 2015. *2015 ICD-9-CM for Hospitals, Volumes 1, 2 and 3 Professional Edition-E-Book*. Elsevier Health Sciences.
- CMS. 2023. Hospital Readmissions Reduction Program: Readmission Measures Methodology Report, CMS QualityNet (Accessed Dec 18, 2025). URL <https://www.cms.gov/files/document/hybrid-hospital-wide-readmission-methodology-report-03202023.pdf>.
- Hong, L. S., X. Garcia-Albeniz, D. Friesen, N. Foskett, R. Beau-Lejdstrom. 2024. Use of clinical classifications software to address icd coding transition in large healthcare databases analyzed via high-dimensional propensity scores. *Pharmacoepidemiology and Drug Safety* **33**(1) e5702.
- Kleibergen, F., R. Paap. 2006. Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics* **133**(1) 97–126.
- Stock, J. H., M. Yogo. 2005. *Testing for Weak Instruments in Linear IV Regression*. Cambridge University Press, New York, 80–108.
- Wooldridge, J. M. 2010. *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, MA.
- Zuckerman, R. B., K. E. Joynt Maddox, S. H. Sheingold, L. M. Chen, A. M. Epstein. 2017. Effect of a hospital-wide measure on the readmissions reduction program. *New England Journal of Medicine* **377**(16) 1551–1558.