CAHPR CENTER FOR ADVANCING HEALTH POLICY THROUGH RESEARCH

School of Public Health
BROWN UNIVERSITY

AI Adoption, Hospital Throughput, and Employment[1]

Working Paper

Daniel R. Arnold,* Jonathan Cantor,** and Christopher M. Whaley***
August 2026

Abstract

We combine 2021–2024 data on artificial intelligence (AI) adoption across U.S. short-term general hospitals with national measures of hospital finances, volume, employment, and measured quality. Using synthetic difference-in-differences, we find that AI adoption is followed by approximately 3% higher net patient revenue, 3% higher total paid hours, and 7% higher patient volume. Total and clinical expenses also rise. By contrast, estimates for administrative expenses, administrative hours, and employee full-time equivalents are imprecise under inference clustered at the hospital-system level. Measured risk-adjusted mortality declines for several conditions, but unadjusted mortality and claims-based clinical-process measures do not show corresponding improvements, while documented severity increases. The results therefore point most clearly to operational expansion, throughput, and richer documentation; they do not establish administrative cost savings, per-unit productivity gains, or lower underlying mortality.

Keywords: Artificial intelligence, labor substitution, hospital performance

---

Supported by NIA R01AG073286. We thank Ben Ukert, Andy Ryan, Ashvin Gandhi, Jonathan Kolstad, Pinar Karaca Mandic, David Chan, Dan Zeltzer, David Cutler, and seminar participants at AHEC, Brown, ASHE, and the NBER Applications of Artificial Intelligence in Healthcare conference for helpful comments.
*Brown University, daniel_arnold@brown.edu
**RAND, jcantor@rand.org
***Brown University, christopher_whaley@brown.edu

1 Introduction

Spurred by OpenAI's release of ChatGPT in November 2022, the use of generative artificial intelligence (GenAI) has rapidly accelerated across the global economy (Jackson, 2023; Brynjolfsson, Li and Raymond, 2025; Singla et al., 2025; Humlum and Vestergaard, 2025*a*). Since 2021, an estimated $1.1 trillion has been invested in AI technologies (Stanford University Human-Centered Artificial Intelligence, 2025). Across industries, the potential impacts of AI are both multifaceted and uncertain (Brynjolfsson, Chandar and Chen, 2025; Buenaventura et al., 2025; Humlum and Vestergaard, 2025*b*). The adoption of AI technologies can improve firm efficiencies and operations (Babina et al., 2024). At the same time, these efficiencies often are expected to arise from the reduction or replacement of human-performed tasks, which may reduce human employment (Acemoglu, Autor and Johnson, 2023; Autor, 2024; Baird et al., 2026; Eloundou et al., 2024; Frank et al., 2019). In addition, the impacts of the adoption of AI on product quality and the corresponding impacts on consumers are uncertain. Thus, a fundamental question is whether the adoption of AI technologies will lead to the substitution of human labor and expertise or contribute to their augmentation (Hui, Reshef and Zhou, 2024). Similarly, the impacts on overall firm productivity and consumers are not known. Although AI adoption has been underway for more than a decade, the impacts of this adoption on both firm performance and labor markets remain poorly understood (Hampole et al., 2025).

A particularly important industry in which to study the impacts of AI adoption is the health care sector (Wachter and Brynjolfsson, 2024; Chen and Zink, 2025; Zink et al., 2023; Nguyen et al., 2025; Obermeyer, 2025; Sahni et al., 2026). Health care firms have many tasks for which AI technologies are expected to improve performance (e.g., using AI to interpret radiology images). Automation within the sector is expected to differ based on the skills and roles of workers (Dranove and Garthwaite, 2024). The health care industry also has a high proportion of administrative tasks (e.g., medical billing, patient scheduling, and clinical note-taking) that are currently manually performed and can be automated or improved using new technologies (Sahni et al., 2024, 2026). One recent report suggested that hospitals spend approximately 56% of their total operating revenue on labor (Deloitte, 2024). As a result, there is considerable potential that AI can both improve performance and reduce administrative burden in the U.S. health care system (Howell, Corrado and DeSalvo, 2024; Shuaib, 2024).

In this paper, we examine the multidimensional impacts of AI adoption in one of the most important segments of the U.S. health care sector: hospitals. We leverage national data from several sources that include detailed measures of AI adoption, as well as granular measures of firm performance, including patient throughput, staffing, and capacity. We also measure hospital quality using national data on hospital quality metrics. In addition, our analyses include detailed claims data from the universe of Medicare-covered hospitalizations that allow us to examine the underlying clinical mechanisms and assess *how* AI adoption impacts hospital performance and patient outcomes.

A key challenge with examining this question is the inherent endogeneity between AI adoption and firm performance. Following previous studies that examine the impacts of firm technology adoption (e.g., Agha, 2014; Brynjolfsson, Li and Raymond, 2025; Freedman, Lin and Prince, 2018; Horn, Sacarny and Zhou, 2022), we use a generalized event study empirical approach. To adjust for underlying differences between hospitals that adopt AI technologies, we further leverage a synthetic difference-in-differences approach that combines standard difference-in-differences and synthetic control approaches (Arkhangelsky et al., 2021). As an additional test, we examine the robustness of our results to using more standard event study approaches (both OLS two-way fixed effects and Sun-Abraham estimators) and find similar results. We also use a number of placebo tests to assess the robustness of our findings to alternative explanations besides AI adoption.

We find that AI adoption is followed by operational expansion rather than broad labor displacement. Under the baseline persistent-adopter definition, employee full-time equivalents and total paid hours rise by 1.3 and 3.2 percent. Under the stronger system-clustered inference, the increase in total hours remains precise, while the FTE estimate is positive but imprecise. Administrative paid hours have a negative point estimate, but the confidence interval is wide and includes both meaningful decreases and increases. We therefore do not interpret the administrative labor result as evidence of substitution.

Hospital finances and volume expand at the same time. Baseline estimates indicate 3.3 percent higher net patient revenue, 2.9 percent higher total expenses, 2.8 percent higher clinical expenses, and 6.6 percent higher discharge equivalents. A contemporaneous treatment definition with 200 bootstrap replications clustered by hospital system produces similar estimates and confirms the revenue, total-expense, clinical-expense, volume, and total-hours results after adjustment for multiple testing. Administrative expenses remain too imprecise to support a cost-savings claim. Because output and inputs both rise, these totals establish expansion but do not by themselves establish per-unit productivity, profitability, or welfare gains.

To understand the sources of observed revenue gains, we also examine the mechanisms through which AI adoption appears to affect hospital performance. The strongest evidence points to an operational channel. In follow-up analyses, AI adoption is associated with higher occupancy and more emergency department visits, but not with a corresponding increase in the hospital-wide case-mix index. We also find evidence of denser documentation and coding: adopting hospitals code more diagnoses, especially secondary diagnoses, and hospital-wide Medicare claims show more diagnoses and higher submitted charges per admission, even though realized Medicare reimbursement per admission does not increase. By contrast, the direct claims-based process measures we are able to test provide much less support for a narrow bedside clinical pathway such as faster diagnostics, less intensive care unit (ICU) escalation, or improved respiratory rescue. Taken together, the mechanism evidence is most consistent with AI helping hospitals expand throughput, use capacity more intensively, and improve documentation and risk capture. A descriptive decomposition shows that throughput measures absorb substantially more of the revenue association than coding measures, but this exercise does not causally allocate the revenue effect across mechanisms.

The implications for patients are less clear. CMS condition-specific measures show declines in measured *risk-adjusted* mortality for heart failure, pneumonia, chronic obstructive pulmonary disease (COPD), and stroke. These are hospital-level risk-adjusted performance measures, not counts of deaths prevented, and we do not extrapolate them into avoided deaths. Several patient-experience measures are unchanged, while nurse communication and receipt of help decline modestly.

Yet, as discussed above, one potential impact of the adoption of AI technologies is to increase coding intensity to increase per-patient reimbursement. Because our mortality measures are risk-adjusted, these changes to coding intensity may increase documented patient severity. In such a case, it is possible for reductions in risk-adjusted mortality to reflect changes in observed patient complexity, rather than any actual reduction in deaths.

To examine this interpretation, we use Medicare hospitalization data to calculate unadjusted mortality for pooled condition-specific cohorts and test clinical processes that could contribute to lower mortality, such as the use and timing of echocardiography among heart-failure admissions. We find no sustained change in unadjusted mortality and little change in the measured processes, while severe comorbidity and risk coding increase. These findings are consistent with richer documentation contributing to the measured risk-adjusted improvement. They do not prove that all of the change is coding-related, because our claims-derived risk index is not the exact CMS risk-adjustment model and the samples and measurement windows differ.

Within this set of overall results, we also examine sources of heterogeneity. We find that AI adoption is associated with larger revenue increases for for-profit and system-affiliated hospitals. We also find larger and more consistent increases in patient discharges among patients with either Medicare or private sources of insurance. We find smaller and noisier increases in lower-paying Medicaid patient discharges.

We subject the findings to treatment-timing, donor-weight, pandemic, hospital-sample, EHR-vendor, and alternative-estimator checks. Revenue, total and clinical expenses, discharge equivalents, and total hours are the most stable findings. Estimates are similar under a contemporaneous treatment definition, after adjustment for county COVID-19 hospitalization burden, and when hospital EHR vendor is interacted with year, although several magnitudes attenuate. Results that depend on the 2021 cohort or a strict observed-transition definition are less stable. Administrative expenses and administrative hours are consistently noisy and should not be treated as robust savings. CARES Act funding placebos generally do not reproduce the positive revenue estimate, but no single placebo eliminates all possible pandemic-era confounding.

Our study makes several important contributions. First, the emergence of AI has led to high-profile policy discussions about both firm performance and whether AI technologies will serve as substitutes or complements for human-performed tasks. Yet, despite the importance of these questions, empirical evidence remains limited. Perhaps most closely related to our study is Brynjolfsson, Li and Raymond (2025), which examines the impacts of AI adoption on the productivity of approximately 5,000 customer-support workers, but does not examine impacts on broader labor markets or firm performance. Similarly, Baird et al. (2026) find that adoption of GitHub Copilot, a generative AI coding platform, led to an increase in hiring software engineers. Both studies are also limited to a narrow worker population. Related studies find evidence that AI spurs firm innovation (Babina et al., 2024, 2025). Other work finds firm exposure to ChatGPT led to a 0.4 percent increase in firm public market valuation and a decrease in worker wages. The literature on the immediate impacts of generative AI builds upon studies of earlier technologies. Autor (2024) examines how technological innovations since 1940 have segmented labor markets by worker skill. Yet within this literature, studies examining how AI or other technology adoption impacts the combination of firms, workers, and consumers, are limited.

Second, we more narrowly contribute to the nascent literature specific to how AI tools are used within the health care sector and their effects. Much of what is currently known about the impacts of recent use of AI by hospitals is related to the characteristics that are associated with adoption: system member hospitals and hospitals with higher operating margins are more likely to be AI adopters (Nong et al., 2025; Bin Abdul Baten, 2024; Goldfarb, He and Teodoridis, 2025), while hospitals located in areas with a high Social Deprivation Index, hospitals located in provider shortage areas, critical access hospitals, and rural hospitals are less likely to use AI (Nong et al., 2025; Chen and Yan, 2025).

While differences in hospital adoption of AI technologies are well documented, the impacts of hospital AI adoption on hospital labor and performance are not known. If hospital adoption of AI materially changes hospital finances or operations, then underlying resource disparities across providers may be exacerbated. Further, if AI tools, in addition to impacting hospital operations, also affect patient outcomes or quality of care, then disparities in health outcomes are also likely to be exacerbated (Nong et al., 2025; Nong, Maurer and Dwivedi, 2025). However, one recent study quantified trends in the quality of care delivered by hospitals that adopted AI and found mixed results (Hwang et al., 2026).

As AI technologies become prevalent, understanding their effects on hospitals and patients is critical. AI can automate some tasks, but it is not clear whether, at scale, it complements or substitutes for human labor, expands capacity, or changes care. To inform policymakers, administrators, and clinicians, this paper uses synthetic difference-in-differences to examine how hospital AI adoption is associated with operations, employment, documentation, and measured quality.

2 Applications of AI in Health Care

The introduction of AI has prompted speculation about its potential to transform the health care sector (Cutler, 2023; Obermeyer, 2025; Kocher, Zhao and Duffy, 2026). Hospitals are important healthcare providers that are likely to be uniquely impacted by AI adoption. AI may affect hospital healthcare delivery and cost through several mechanisms (Sahni et al., 2026). AI may substitute for human labor in administrative tasks such as billing, management, scheduling, and documentation; complement clinicians by integrating structured and unstructured data in ways that support decision-making; and generate efficiency gains in monitoring, diagnosis, and workforce allocation (Dranove and Garthwaite, 2024; Sahni et al., 2024).

AI can impact hospital spending and performance through several mechanisms. Savings and improvements may come from improvements in clinical operations, quality and safety, continuity of care, reimbursement, corporate functions, or clinical analytics (Sahni et al., 2024). One estimate suggests that hospitals could achieve net savings of $60 billion to $120 billion annually within five years without reducing quality or access to care, with roughly 40 percent of savings coming from lower administrative costs and 60 percent from lower medical costs (Sahni et al., 2024). These potential savings could arise from replacing some administrative labor, improving the allocation of existing staff, and reducing inefficiencies in hospital operations. For example, Duke Health reported that it reduced temporary labor staffing by 50 percent after using AI tools to better align staffing levels with patient demand (Kass-Hout, 2025).

At the same time, AI is potentially more likely to complement than replace skilled clinical labor. Many clinical tasks require professional judgment, interpretation of unstructured information, and accountability for decisions, which limit the scope for full automation. Instead, AI may augment clinicians by improving information processing, documentation, and decision support. Potential benefits include faster and more accurate diagnosis (Jiang, Edwards and Newstead, 2021; Khalifa and Albadawy, 2024), improved patient engagement and adherence (Babel et al., 2021; Sekandi et al., 2023), and lower documentation burden. For example, one recent study found that physician use of an AI scribe was associated with 13.4 fewer minutes of electronic health record time, 16.0 fewer minutes of documentation time, and 0.49 additional weekly visits delivered (Rotenstein et al., 2026). Examples from AI companies point to similar mechanisms: Abridge reports that its tools assist clinicians before, during, and after visits by supporting documentation and decision-making, while Aidoc reports shorter turnaround times for positive pulmonary embolism findings after deployment of its radiology tools.[1] AI tools that allow clinicians to see more patients or spend less time per encounter while maintaining quality could increase labor productivity and lower the cost per visit without reducing the demand for clinical workers.

AI tools may also alter where and by whom care is delivered. Lower-cost monitoring, triage, and diagnostic capabilities could shift some care from more expensive inpatient settings to less expensive outpatient or telehealth settings. AI may also support earlier intervention before conditions escalate in severity and cost. In addition, some services currently delivered by specialists may become feasible by lower-cost generalists when helped by algorithmic support. For example, AdventHealth reported that AI-enabled patient flow management reduced patient movement times within their hospital by 15 minutes and improved patient placement by 20 minutes (AdventHealth, 2022).

Yet, empirical evidence on the effects of AI adoption on hospital performance is limited. Prior research has focused primarily on rates of self-reported AI adoption by hospitals (Nguyen et al., 2025; Nong et al., 2025; Bin Abdul Baten, 2024; Goldfarb, He and Teodoridis, 2025; Hwang et al., 2026). The existing empirical literature on AI and hospital quality is mixed (Hwang et al., 2026). Thus, while AI may reduce administrative costs, increase clinician productivity, and shift care toward less costly settings and provider types, the magnitude of these effects in hospitals remains unknown.

3 Data

3.1 Hospital AI Adoption

We use data from the American Hospital Association (AHA) Annual Survey to classify hospitals as AI adopters or non-adopters. The AHA Annual Survey began asking questions

[1] Sources: www.abridge.com/platform/clinicians and info.aidoc.com/radiology-ai-demo.

related to AI in 2021. We used 2021 to 2024 AHA Survey data to track AI adoption by hospitals. The 2024 Annual Survey included hospital characteristics for all 6,100 hospitals in the United States, of which 5,121 were community hospitals (i.e., nonfederal, short-term general, and other specialty hospitals) (American Hospital Association, 2026). For this study, we only considered survey responses from the 4,214 short-term general hospitals within the larger group of community hospitals as they accounted for an overwhelming majority (95%) of community hospital admissions and are more easily comparable (in terms of services offered) than specialty and long-term hospitals.

The survey questions related to AI changed for the 2024 survey and contained more detail than the AI question in the 2021-2023 surveys, which simply asked "Does your hospital use artificial intelligence (AI) or machine learning in the following? (Check all that apply): 1) predicting staffing needs, 2) predicting patient demand, 3) staff scheduling, 4) automating routine tasks, 5) optimizing administrative and clinical workflows, 6) none of the above." In 2024, the survey asked about clinical vs. operational use of AI and whether hospitals were not implementing AI, exploring it, piloting/testing it, expanding it, or had fully integrated it (see Figure A1 for the survey questions).

Among the short-term general hospitals in the United States in 2024, 2,406 (57%) responded to this question. The response rate was similar in 2023 (53%) and 2022 (56%), but lower in 2021 (20%). Among the 4,093 short-term general hospitals in our analytic sample between 2021 and 2024, 2,856 (70%) answered the AI question in at least one year of the survey.

Our baseline specification uses a *persistent-adopter* definition. A hospital enters the treated group in the first year during FY2021–FY2023 that it checked any AI-use box, provided that its 2024 response was "expanding" or "fully integrated." Hospitals whose later response did not meet that threshold are excluded from the baseline treated group. In 2024, hospitals enter treatment if they report "expanding" or "fully integrated." Hospitals that never report qualifying use, including nonrespondents, form the baseline comparison group. This rule prioritizes sustained, substantive implementation, but it uses information observed after some hospitals' initial reports. We therefore also estimate a contemporaneous definition based only on information available in each survey year, a definition that excludes the sparsely observed 2021 cohort, and a strict observed-transition definition. Appendix Table A4 reports these sensitivity analyses, and a responders-only analysis addresses possible hidden adopters among survey nonrespondents.

3.2 Hospital Performance and Quality

We used Healthcare Cost Report Information System (HCRIS) data to construct measures of hospital performance and financial status. HCRIS is a comprehensive collection of financial and statistical information that Medicare-certified institutional health care providers are required to submit annually to CMS. For hospitals, the data provide a detailed breakdown of a facility's operations, including its costs, charges, utilization, and finances. We extracted 2017-2024 data for eight primary outcomes: net patient revenue, total expenses, expenses for clinical activities, expenses for administrative activities, discharge equivalents (i.e., volume), employees on payroll (full-time equivalents), total paid hours, and administrative and general paid hours. These data have previously been used to measure hospital performance (Prager and Schmitt, 2021; Schmitt, 2017).

To measure care quality, we used 2017-2024 quality of care data from CMS's Hospital Compare dataset. We assessed the association between AI adoption and seven quality outcome measures: six mortality measures – the 30-day mortality rate after being admitted to the hospital for heart attack, heart failure, pneumonia, chronic obstructive pulmonary disease (COPD), stroke, or coronary artery bypass graft (CABG) surgery, respectively – and a hospital's all-cause 30-day readmission rate. We additionally extracted five patient experience measures – communication with nurses, communication with doctors, receiving help, rating the hospital a 9 or 10, and recommending the hospital. The mortality and readmission measures are commonly used in the literature (Oseran et al., 2023; Werner and

Bradlow, 2006), as are the patient communication and hospital recommendation measures (Jha et al., 2008).

We supplemented these data with 2017-2024 Medicare administrative claims and enrollment records. The Medicare Provider Analysis and Review (MedPAR) file contains stay-level records for inpatient admissions among Medicare fee-for-service (FFS) beneficiaries, including the treating hospital, admission and discharge dates, diagnosis and procedure codes, present-on-admission indicators, utilization, charges, and Medicare reimbursement. We linked these admissions to annual beneficiary enrollment files to measure age, sex, race and ethnicity, dual-eligibility status, and Medicare Advantage enrollment during the year. We used these data to construct hospital-year measures of ED-origin admissions, patient composition, diagnosis-coding intensity, utilization, and mortality among patients admitted for COPD, pneumonia, heart failure, and stroke. The risk-adjusted mortality analysis begins in 2018 because it requires prior-year claims history. For the surgical-volume analysis, we additionally linked a 20% sample of FFS Carrier claims to MedPAR stays to identify inpatient surgical procedures and the performing surgeon.

4 Empirical Approach

A key challenge when assessing the impacts of AI on hospital performance is that hospitals that adopt AI likely differ from those that do not. For example, adopters may have more financial resources or operate in more favorable payer environments. As a solution, we used the Synthetic Difference-in-Differences (SDID) method to estimate the association between AI adoption and changes in our outcome measures (Arkhangelsky et al., 2021).

The goal of SDID is to estimate the Average Treatment Effect on the Treated (ATT) by constructing a credible counterfactual for what would have happened to the treated units in the absence of treatment. From synthetic control (SC), SDID borrows the idea of reweighting untreated (control) units to create a “synthetic” control group. This synthetic control is constructed such that its outcome trends in the pre-treatment period closely match the trends of the treated group. SDID incorporates the core logic of difference-in-differences (DID) by allowing for time-invariant differences between the treated and control groups (i.e., unit fixed effects), meaning the method focuses on matching pre-treatment trends rather than requiring a perfect match in pre-treatment outcome levels, thereby relaxing a key assumption of the standard SC method.

A key innovation of SDID is the introduction of time weights. In addition to weighting control units, SDID also weights the pre-treatment time periods. It gives more importance to periods that are more predictive of outcomes in the post-treatment period, which can improve the estimator's precision and robustness.

The core estimating equation is:

$$(\hat{\tau}^{sdid}, \hat{\mu}, \hat{\alpha}, \hat{\beta}) = \underset{\tau,\mu,\alpha,\beta}{\arg\min} \left\{ \sum_{i=1}^{N} \sum_{t=1}^{T} (Y_{it} - \mu - \alpha_i - \beta_t - W_{it}\tau)^2 \hat{\omega}_i^{sdid} \hat{\lambda}_t^{sdid} \right\},$$

where Y_{it} is the outcome for unit i at time t, W_{it} indicates treatment, and α_i and β_t are unit and time fixed effects, respectively. This hybrid approach makes SDID robust in settings where traditional DID may fail due to non-parallel trends and where traditional SC may fail because the treated unit's outcomes are outside the range of the control units.

In this study, we were faced with both non-parallel trends and differences in pre-treatment outcome levels. The trends were frequently different across treated and control groups and there were sizable level differences given that treated hospitals are generally larger.

Our analysis uses a balanced panel of 4,093 U.S. short-term general hospitals from 2017 through 2024. Under the baseline persistent-adopter definition, the treated group consists of 1,511 hospitals first treated between 2021 and 2024, and the comparison group consists of 2,582 hospitals that do not meet the adoption definition during the study period. The main dynamic figures use this definition. Because it conditions on sustained implementation observed in 2024, we also report estimates using contemporaneous survey-year status and other treatment-timing rules.

Estimation proceeds separately by adoption cohort. For cohort a, the pre-treatment period runs from 2017 through $a-1$, and the donor pool contains hospitals that never meet the relevant treatment definition. Unit weights ($\hat{\omega}_{ja}$) construct a synthetic control that reproduces the treated cohort's pre-adoption path, while time weights ($\hat{\lambda}_{ta}$) determine the pre-adoption benchmark. Let $\bar{Y}_{a,t}$ be the mean outcome for treated hospitals in cohort a. With event time zero defined as the adoption year, the cohort-by-event-time effect is

$$\widehat{\tau}_{a,l} = \left(\bar{Y}_{a,a+l} - \sum_{j} \widehat{\omega}_{ja} Y_{j,a+l}\right) - \sum_{t<a} \widehat{\lambda}_{ta} \left(\bar{Y}_{a,t} - \sum_{j} \widehat{\omega}_{ja} Y_{j,t}\right).$$

Building on this foundation, the SDID framework can be extended to an event-study analysis to understand the dynamic, period-by-period effects of a treatment. This extension disaggregates the overall ATT into a series of estimates, showing how the treatment effect evolves over time after implementation.

The event-study extension reports these effects separately by event time. The cohort-specific SDID estimate averages the observed post-treatment event times for that cohort.

In a staggered adoption setting where different cohorts receive treatment at different times, it is useful to have a single estimate for the effect at a specific lag l (e.g., the average effect two years post-treatment). This is achieved by calculating the aggregated dynamic estimator ($\hat{\tau}_l^{sdid}$), which is a weighted average of the cohort-specific dynamic effects ($\hat{\tau}_{a,l}^{sdid}$) across all cohorts observed l periods after their treatment began. The weights are based on the relative number of treated units in each cohort, giving more influence to larger cohorts. The formula for this aggregated dynamic estimator is:

$$\hat{\tau}_l^{sdid} = \sum_{a \in A_l} \frac{N_a}{N_l} \hat{\tau}_{a,l}^{sdid},$$

where $N_l = \sum_{a \in A_l} N_a$ is the number of treated hospitals observed at event time l. Our overall treated-hospital-year (THY) estimand weights each observable cohort-by-event-time effect by the number of treated hospitals contributing to that cell:

$$\widehat{\tau}^{THY} = \frac{\sum_a \sum_{l \in L_a} N_a \widehat{\tau}_{a,l}}{\sum_a \sum_{l \in L_a} N_a} = \sum_l \frac{N_l}{\sum_s N_s} \widehat{\tau}_l.$$

For the eight primary outcomes in Table 4, inference uses 200 nonparametric bootstrap replications clustered by hospital system, with independent hospitals treated as separate clusters. We report 95% intervals and control family-wise error across the eight outcomes using Holm-adjusted p-values; Benjamini–Hochberg false-discovery-rate adjustments yield the same substantive conclusions. The event-study figures use 50 hospital-clustered bootstrap replications and are interpreted as descriptive evidence about dynamics rather than as the primary multiplicity-adjusted inference. For completeness, we also estimate traditional two-way fixed-effects (TWFE), Sun and Abraham (2021), and propensity-score-matched models. The matched stacked event studies and summary difference-in-differences regressions are estimated by OLS with matched cohort-by-hospital and matched cohort-by-year fixed effects and hospital-clustered standard errors; they are not Sun–Abraham estimates.

We assess the credibility of the synthetic controls using pre-period normalized root mean squared prediction error, effective donor counts, the maximum donor weight, treated-outcome support, exclusion of same-system donors, uniform rather than estimated time weights, and leave-one-out removal of the highest-weight donor. Appendix Table A3 summarizes these diagnostics. All eight primary outcomes lie within donor support and no single donor is influential. Pre-period fit is strongest for revenue, expenses, and volume and materially weaker for labor hours, especially administrative hours; this difference informs our interpretation of the results.

5 Results

5.1 Descriptive Characteristics of Hospital AI Adoption

We identified 1,511 hospitals that adopted AI over the 2021-2024 period. 624 hospitals first reported using AI in 2021, while 165, 100, and 622 hospitals adopted in 2022, 2023, and 2024, respectively. Figure 1 maps where the 1,511 AI-adopting hospitals are located throughout the U.S. Adoption does not appear to be concentrated in any particular region of the country. Table 1 compares characteristics of the AI adopter and non-adopter hospitals at baseline in 2017. AI adopter hospitals were larger (221 beds vs. 114 beds), more likely to be part of a system (0.87 vs. 0.61), less likely to be critical access hospitals (0.19 vs. 0.39), more often located in urban areas (urban continuum code 2.7 vs. 4.1), less likely to be for-profit (0.07 vs. 0.22), more often a major teaching hospital (0.13 vs. 0.05), and had higher case-mix indices (1.68 vs. 1.56) than non-adopter hospitals.

5.2 Labor Market Impacts

Figure 2 shows a 1.3% increase in employee FTEs and a 3.2% increase in total paid hours under the baseline persistent-adopter definition. The stronger contemporaneous-treatment analysis in Table 4 yields a similar 3.2% increase in total hours with a system-clustered 95% confidence interval of 1.8% to 4.6%. The FTE estimate is positive but imprecise under that inference. Administrative paid hours have a negative point estimate in both specifications, but their confidence intervals are wide; we therefore treat the administrative-hours result as inconclusive. The event studies are useful for describing timing, while the system-clustered table provides the primary inference.

These results do not support broad employment displacement during the study period. They are consistent with AI accompanying an expansion in hospital activity and total labor input, although the data do not identify task-level substitution or changes in the occupational mix.

5.3 Hospital Finances and Volume

Turning to finances and volume, the baseline estimates show a 3.3% increase in net patient revenue, a 2.9% increase in total expenses, a 2.8% increase in clinical expenses, and a 6.6% increase in discharge equivalents (Figure 3). The corresponding contemporaneous-treatment estimates in Table 4 are 2.8%, 2.1%, 2.7%, and 7.1%, respectively, and all remain precise after system-level clustering and Holm adjustment. Outcomes estimated in logs are transformed to percent changes. Because revenue, expenses, volume, and labor inputs all rise, these findings show organizational expansion but do not by themselves demonstrate per-unit productivity or efficiency gains.

The event-study plots for the five finance and volume outcomes are shown in Figure 4. Net patient revenue rises after adoption and the point estimates grow at later event times (Panel A), although later estimates increasingly reflect the early-adopting cohort. Administrative-expense estimates are negative at some event times but become noisy, and the stronger inference in Table 4 includes zero. We therefore do not interpret the administrative-expense series as evidence of robust cost savings.

5.3.1 Hospital finances and volume mechanisms

Several mechanisms may underlie the observed expansion. AI may alter administrative tasks such as billing, scheduling, and management or complement clinicians through diagnostic support, risk stratification, and workflow optimization. We examine increased patient throughput, coding intensity, and patient selection. The evidence is strongest for operational throughput and documentation-related mechanisms, while the claims-based measures provide much less evidence of a direct clinical pathway.

Mechanism 1: Throughput and capacity expansion We first examine changes in patient throughput and capacity. To do so, we extend our main finding of a 6.6% increase in discharge equivalents. We first examine whether the volume increases are contained to a particular service type (e.g., inpatient) or payer (e.g., commercial) or applied more broadly. Figures A2 and A3 in the Appendix show volume increased across the board – by type (inpatient, outpatient) and payer (commercial, Medicare, and Medicaid). Most notably, we find similar increases in commercial insurance and Medicare volume, but smaller changes in lower-paying Medicaid patient volume.

To further explore changes in throughput, we next examine whether increases in patient volume are due to increases in physical capacity or changes in the use of *existing* capacity. Notably, finding changes in physical capacity could be a potential confounder for our core results. In addition, hospital beds did not increase at adopting hospitals until three years post-adoption and when bed expansion did occur it was at a much smaller level (2.0%) than the increase in patient volume. However, we find that AI adoption increases hospital occupancy by about 6.4% in logs (or about 3.0 percentage points in levels). Thus, it does not appear to be the case that the observed increase in hospital patient volume following AI adoption is due to increased physical capacity, but rather an increase in the use of existing capacity.

We next examine changes in throughput that occur through the emergency department (ED) channel. Emergency departments are a major pathway to inpatient care: in 2022, 11.5% of ED visits resulted in admission to the same hospital, and 2018 benchmarking data suggest that approximately 70% of hospital admissions were processed through the ED.[2] Inpatient volume could increase if emergency departments see more patients or if hospitals admit a greater share of ED patients. We use the MedPAR ER-charge field to identify FFS inpatient stays that originated in the ED.[3] In the national FFS MedPAR data, the share of inpatient admissions with an ER charge rose steadily over the sample period, from 69.3% in 2017 to 76.9% in 2024, consistent with an increasing role for the ED in inpatient admission pathways.

[2]Source: National Center for Health Statistics. National Hospital Ambulatory Medical Care Survey: 2022 Emergency Department Summary Tables and (Melnick et al., 2022).

[3]The ER-charge field is available in our FFS MedPAR files but not in the corresponding MA encounter files.

Figure 5 reports synthetic difference-in-differences event-study estimates for ED-origin admissions and the share of inpatient admissions originating in the ED. The pre-adoption placebo estimates are economically close to zero. Following AI adoption, ED-origin admissions increase progressively, rising by 3.7% in the adoption year, 8.4% after one year, 13.3% after two years, and 21.0% after three years. In contrast, the ED-origin share remains flat to modestly lower, with post-adoption estimates ranging from approximately −1.7% to −3.2%; only the year-one estimate is statistically distinguishable from zero. Thus, AI-adopting hospitals admit more patients through the ED in levels, but this growth is part of a broader increase in inpatient volume rather than a shift toward routing a larger fraction of admissions through the ED. This pattern is consistent with other evidence that operational changes can increase ED throughput without materially changing hospitals' reliance on the ED admission pathway (Richards and Whaley, 2024).

Mechanism 2: Surgical volume Volume could increase because hospitals hire more doctors, existing doctors perform more procedures, or some combination of additional hiring and increased per doctor caseload. We examine inpatient surgical volume to assess how much each channel matters. One possibility is that AI adoption improves hospital throughput or capacity management, allowing hospitals to accommodate more procedures without a proportional increase in fixed inputs. Surgical care is a useful mechanism test because inpatient procedures are high-revenue, capacity-constrained episodes where changes in scheduling, perioperative coordination, documentation, triage, or bed management could plausibly translate into measurable volume growth. We measure these patterns using Carrier (e.g., physician) claims linked to MedPAR stays, which identify inpatient surgical stays and the surgeons associated with those stays. To reduce sensitivity to one-off provider appearances and small hospital-year samples, our preferred specification counts a surgeon NPI only when it is associated with at least five linked surgical stays in that hospital-year and restricts the sample to hospitals with at least 50 linked surgical stays in every year.

Using these more conservative definitions, AI adoption is associated with a 6.2% average increase in qualifying surgeon NPIs and a 2.5% increase in stays per qualifying surgeon. The event-study estimates increase gradually, reaching 10.5% and 4.6%, respectively, three years after adoption, while the pre-adoption placebo estimates remain close to zero. Thus, the increase in surgical volume appears to operate through both broader surgeon participation and somewhat greater activity among repeatedly observed surgeons, with the participation margin accounting for the larger share. Because NPIs identify surgeons observed performing procedures rather than hospital employees, the first estimate should not be interpreted as a direct measure of hiring. Instead, it indicates that AI-adopting hospitals retain or engage more repeatedly active surgeons relative to their synthetic comparison hospitals (Figure 6).

Mechanism 3: Changes in patient composition We similarly examine if the increases in capacity are due to changes in patient composition. To do so, we examine changes in observable patient characteristics (e.g., age, sex, race, and income status) using the MedPAR data. To measure changes in patient income, we measure dual enrollment in both Medicare and Medicaid. As with the ED analysis, we limit the MedPAR sample to the FFS population. Although beneficiary demographic fields are well populated for both FFS and MA enrollees, we restrict these analyses to FFS MedPAR admissions because our MA encounter file was not complete through 2024 and therefore could not support a consistent 2017–2024 panel.

The SDID event study for mean patient age shows little evidence of a compositional shift. Post-adoption estimates range from approximately -0.02% to 0.08% and are statistically indistinguishable from zero, while the pre-adoption placebo estimates are similarly small. Thus, the increase in volume does not appear to be driven by AI hospitals admitting meaningfully younger or older patients (Figure 7). Other patient characteristics point to a similar conclusion. Estimated changes in the female share are small and statistically insignificant. We also find no overall meaningful changes in hospitals' case-mix indices, which capture expected patient complexity (Table 2). However, we do find small increases in case-mix several years after AI adoption (Figure 7). Taken together, these results suggest that the post-adoption volume increase is not primarily explained by a large observable shift in the composition of admitted patients.

Mechanism 4: Denser documentation and coding The fourth mechanism we examine is broader documentation intensity and risk capture. Across insurers, US hospitals are largely paid based on patient and condition severity. A key and long-appreciated challenge is that "true" severity is unobserved, leading to incentives to "upcode" patients and conditions (Cutler and Zeckhauser, 2000; Dafny, 2005). Indeed and as summarized above, a key mechanism of many health care AI, as well as previous technologies, is to increase coding and documentation intensity. To examine this mechanism, we use the MedPAR data to construct several measures of coding intensity, including total diagnoses per admission, secondary diagnoses per admission, and newly coded conditions.

As highlighted in Table 3, we find meaningful increases in coding intensity. The primary coding-intensity estimates use the pooled COPD, pneumonia, heart failure, and stroke cohorts. We then test whether the findings generalize in a separate hospital-wide panel containing all Medicare FFS acute inpatient admissions. AI adoption increases the total number of diagnoses and secondary diagnoses coded on the claim. The follow-up present-on-admission tests point in the same direction. After adoption, hospitals code more diagnoses that are flagged as already present when the patient arrived, but they also code more diagnoses that are *not* flagged that way. Because both categories rise, the percentage marked present on admission falls slightly. This is not the pattern we would expect if AI were mainly helping hospitals uncover a fixed set of previously under-documented pre-existing comorbidities at admission. Instead, it suggests broader growth in documentation intensity over the course of the stay. Consistent with that interpretation, we also do not find evidence of an increase in newly recorded chronic conditions that were absent from the patient's prior-year claims history.

The results presented in Table 3 that use the hospital-wide Medicare FFS panel points in the same direction, but also clarifies what this mechanism is *not*. After dropping any hospital that ever has a nonpositive mean Medicare reimbursement or ever falls below 10 Medicare FFS admissions in a year, AI adopters have more diagnoses per Medicare admission and higher submitted charges per admission. Under a stricter sensitivity screen that drops any hospital ever below 25 Medicare FFS admissions, those patterns remain similar. However, average Medicare reimbursement per admission falls rather than rises, even as total Medicare FFS discharges increase substantially. Because the main paper's revenue result is hospital-wide net patient revenue across all payers, it need not move in parallel with Medicare FFS reimbursement per admission.

These results are not consistent with a mechanism in which denser coding mainly boosts Medicare payment per stay. Instead, the financial value of denser documentation appears more likely to come from broader organizational channels: more complete records, stronger revenue-cycle support, and better measured severity in settings where risk adjustment matters. In this sense, the coding results align with the mechanism that some of the observed increases in risk-adjusted quality reflect hospital risk documentation increases conditional on documented patient severity, even if average Medicare FFS payment per admission does not rise. Taken together, these results do not suggest aggressive fabrication of new chronic conditions or straightforward "upcoding" into higher Medicare payment per stay, but more complete documentation and risk capture overall, combined with higher throughput.

Relative contribution of candidate mechanisms. As a descriptive accounting exercise, we add candidate mechanism indices to a two-way fixed-effects model of log net patient revenue and measure how much the reduced-form AI coefficient attenuates. A throughput index combining discharge equivalents, inpatient discharges, and occupancy attenuates the coefficient by 58.0%, whereas a coding and billed-intensity index attenuates it by 4.3% (2.7% when coding is measured only by diagnoses per claim). These quantities are not causal mediation estimates: the candidate mechanisms are post-treatment outcomes, may be jointly determined, and are measured with error. They show only that observed throughput measures account for more of the revenue association than the available coding measures.

5.3.2 Primary-Outcome Inference

Table 4 reports the eight prespecified operational outcomes using the contemporaneous treatment definition, 200 bootstrap replications clustered by hospital system, and multiplicity-adjusted p-values. Revenue, total expenses, clinical expenses, discharge equivalents, and total paid hours remain statistically distinguishable from zero after Holm adjustment. Employee FTEs, administrative expenses, and administrative paid hours do not. These results motivate our distinction between robust evidence of expansion and uncertain evidence about administrative savings or employment composition.

5.4 Impacts on Hospital Quality

Our next core result examines changes in hospital quality following AI adoption. Changes to hospital quality are both a measurable and important way to quantify the impact of AI technologies on consumer welfare. To examine changes in quality, we first measure changes in *risk-adjusted* mortality using "Hospital Compare" measures that CMS uses to monitor hospital quality. The Hospital Compare metrics include both hospital clinical quality, measured through condition-specific readmission and mortality rates, and patient experience and satisfaction metrics.[4]

As highlighted in Figure 8, we find meaningful reductions in *risk-adjusted* mortality for several clinical categories. Specifically, AI adoption is associated with a 0.22 percentage point (pp) decrease in the risk-adjusted heart failure mortality rate. The event studies in Figure 9 show steady reductions in risk-adjusted mortality following AI adoption. The mean

[4]The Hospital Compare clinical metrics are constructed using Medicare claims data, while the patient experience and satisfaction metrics are constructed using random surveys of hospitalized Medicare beneficiaries.

heart failure mortality rate at baseline for AI adopters was 11.8%, so this corresponds to a 1.9% relative decrease. AI adoption was also associated with a 0.41 pp decrease in the pneumonia mortality rate, a 0.21 pp decrease in the COPD mortality rate, and a 0.22 pp decrease in the stroke mortality rate, which correspond to 2.5%, 2.4%, and 1.6% relative decreases, respectively. The other three quality outcome measures we analyzed were not associated with AI adoption. These are measured risk-adjusted mortality outcomes, so they should be interpreted with some caution. As discussed in the mechanisms section below, the follow-up coding results suggest that some portion of the apparent improvement may reflect richer documentation and risk capture rather than a single direct bedside clinical pathway.

For patient experience (Figure 10), three of the five quality patient experience measures we analyzed likewise were not associated with AI adoption, with nurse communication and "receiving help" being the exceptions. AI adoption was associated with a 0.44 pp decrease in the percentage of patients who said their nurses always communicated well. At baseline, 80% of patients at AI-adopting hospitals reported their nurses always communicated well, so this effect corresponds to a less than 1% relative decrease. AI adoption was associated with a 0.21 pp decrease in the percentage of patients who said they received help when needed. At baseline, 67% of patients at AI-adopting hospitals reported receiving help when needed, so this effect corresponds to a less than 1% relative decrease. Notably, the patient experience changes show *decreased* patient experience with areas involving communication. One potential mechanism is that AI technologies may replace in-person communication.

5.4.1 Patient Quality Mechanisms

The reductions in risk-adjusted mortality are sizable as measured, but they should not be equated with deaths prevented. Because the CMS measures are risk-adjusted, they can change through underlying outcomes, documented risk, or both. As highlighted above, diagnosis density and billed charges also increase, although realized Medicare reimbursement per admission does not. To assess the mortality results, we explore clinical pathways that could influence mortality and changes in underlying risk coding.

Clinical pathways We first leverage the Medicare claims data to examine clinical outcomes and processes that should influence condition-specific mortality. Finding granular evidence of clinical improvements following AI adoption would suggest that our observed mortality declines are driven by changes in actual performance. To do so, we use a set of measures, summarized in Table 5, that should impact condition-specific mortality (e.g., use and speed of echocardiogram diagnostic tests for heart failure patients).

However, the claims-based process measures do not provide evidence of a strong direct clinical pathway that would explain the lower mortality (Table 5 summarizes these results). We do not find faster echocardiograms in heart failure, faster head CT in stroke, more frequent echocardiograms in heart failure, or more frequent head CT in stroke. We also do not find meaningful reductions in ICU use or ICU days for COPD, pneumonia, heart failure, or stroke. The respiratory process measures are similarly non-suggestive: the non-invasive stabilization measure is essentially zero and the intubation or respiratory-failure measure is slightly positive rather than negative. These null and uncertain results for these clinical tests cast doubt on the observed reductions in risk-adjusted mortality being driven by actual changes in mortality.

Changes to Risk Adjustment Our main quality results should be interpreted as improvements in measured, risk-adjusted mortality rather than evidence of lower underlying mortality. To decompose risk-adjustment and actual mortality improvements, we constructed a claims-based unadjusted mortality panel for the four conditions with apparent CMS quality improvements: heart failure, pneumonia, COPD, and stroke. To measure the influence of risk scores, we focus on severe risk coding, defined as the number of serious comorbidities and acuity markers captured on the inpatient claim, because these documented conditions can increase measured patient severity and, in some cases, affect both risk-adjusted quality metrics and inpatient reimbursement.

Figure 11 combines these two metrics across all four conditions. In this event study, AI adoption is followed by a clear increase in severe claims-based risk and comorbidity coding. After AI adoption, severe risk coding increases gradually and persistently, rising by roughly 1.2 to 1.6 percent in years 1 through 3 after adoption. In contrast, unadjusted mortality is flat. The mortality estimate falls in the first post-adoption year but then returns close to zero in subsequent years. These patterns suggest that AI adoption is associated with more intensive documentation of severe comorbidities, while there is no evidence for a sustained reduction in underlying mortality.

Overall interpretation. Taken together, our findings points most strongly to AI impacting hospital operations. After AI adoption, hospitals move more patients through the organization, use capacity more intensively, and handle more emergency department volume without a corresponding increase in hospital-wide case mix. These patterns suggest that AI is helping hospitals expand throughput and tighten capacity use rather than simply shifting toward observably sicker patients. A second, but still important, channel is documentation and risk capture. Adopting hospitals code more diagnoses, especially secondary diagnoses, and do so on both the present-on-admission and non-present-on-admission margins. In the hospital-wide Medicare panel, billed charges per admission also rise, but realized Medicare reimbursement per admission does not. That pattern is more consistent with more complete documentation, NLP extraction, revenue-cycle support, and stronger measured severity in risk-adjusted settings than with a simple payment-per-claim upcoding story.

At the same time, the current claims-based process measures provide much less support for improvements in underlying clinical outcomes. We do not observe faster diagnostic testing, less ICU escalation, or clearer signs of improved respiratory rescue after adoption. The overall picture is therefore one of expanded throughput and more intensive capacity use, alongside denser documentation, coding completeness, and measured severity. Because both output and inputs rise, we do not label this pattern a per-unit efficiency gain. The combination of denser coding, flat unadjusted mortality in the claims-based cohorts, and largely null process measures suggests that richer documentation may contribute to the improvement in measured risk-adjusted mortality. That conclusion remains qualified: our severe-risk index is not the exact CMS model, and the CMS and claims analyses use different outcome construction and windows. The evidence does not establish that AI raises Medicare payment per admission or changes a specific acute clinical decision observable in claims.

5.5 Additional Results

We additionally find that AI adoption is associated with larger increases in revenues among for-profit, non-critical access, and system hospitals, but subgroup estimates were not statistically different from one another, largely because of reduced sample size. We find little heterogeneity across the same three dimensions in terms of effects on administrative expenses (Figures A4 and A5).

6 Robustness Tests and Heterogeneity

We assess alternative treatment definitions, adoption cohorts, comparison groups, pandemic-era confounding, hospital samples, EHR vendors, and estimators. We also report direct diagnostics of the synthetic controls. These exercises distinguish findings that recur across demanding specifications from findings that are sensitive to timing or inference.

The clearest recurring pattern is higher volume together with higher revenue, clinical spending, and total paid hours. The exact magnitudes attenuate under some treatment-timing and EHR specifications, and the strict observed-transition design is less supportive. Administrative expenses and administrative hours are not robustly different from zero. The results therefore support a narrower conclusion about operational expansion, not a general claim of administrative savings or efficiency.

6.1 Types of AI Technology Adoption

Table A1 summarizes the types of clinical and operational AI hospitals are adopting. Among the more than 2,000 hospital respondents to the 2024 AI survey questions, summing the "expanding" and "fully integrated" categories makes clear that hospital AI adoption is concentrated in specific clinical applications. Notably, Clinical Decision Support Tools (43%), Predictive Analytics for Patient Care (39%), and AI-Assisted Diagnostics (37%) demonstrate substantially higher adoption rates compared to the rest of the surveyed applications. While operational applications generally lag behind, Revenue Cycle Management does show some meaningful traction, with 29% of hospitals reporting it as expanding or fully integrated. Overall, however, the highest rates of deep integration are concentrated within patient-facing and diagnostic tools, suggesting that advanced clinical AI implementation is currently more common and mature than operational AI use cases in hospitals.

To examine whether the effects of AI differ by the type of technology adopted, we classify the 1,511 AI-adopting hospitals into three mutually exclusive groups based on their 2024 AHA survey responses: clinical only (566 hospitals, 37%), operational only (156 hospitals, 10%), and both clinical and operational (789 hospitals, 52%). A hospital is classified as a clinical adopter if any clinical AI variable is "expanding" or "fully integrated" and no operational variable meets that threshold, and analogously for operational-only adopters. Hospitals meeting the threshold in both domains are classified as "both." For hospitals that first adopted AI in 2021–2023 based on the earlier binary survey questions, we assign the type classification from their 2024 responses.

Figure 12 presents SDID estimates by adoption type for six key outcomes. All three groups show positive and significant revenue effects, but hospitals adopting both clinical and operational AI exhibit the largest gains (3.8%), compared with 2.7% for operational-only and 2.4% for clinical-only adopters. This pattern is consistent with complementarities between clinical and operational AI. The volume effects are similarly largest for the "both" group (7.6% increase in discharge equivalents) and operational-only adopters (7.2%), suggesting that operational AI—particularly applications like patient flow optimization and demand forecasting—may be especially effective at increasing throughput.

Clinical-only adopters have negative point estimates for administrative expenses and hours, while operational-only adopters have estimates near zero and hospitals adopting both types lie between them. These subgroup estimates are exploratory: samples are substantially smaller, confidence intervals overlap, and the main system-clustered administrative estimates are imprecise. We therefore do not infer technology-specific administrative savings from these differences.

6.2 Cohort-Specific Effects

Our staggered adoption design allows us to examine heterogeneity by adoption timing. Figure 13 presents separate estimates for the 2021–2024 cohorts. The 2021 cohort generally has the largest point estimates and is also the only cohort observed for four post-treatment years. Because the cohort is identified from a low-response survey year and later event times contain progressively fewer cohorts, this pattern cannot by itself be interpreted as a dose response.

For example, discharge equivalents increase 8.3% for the 2021 cohort versus 4.5% for 2022 and 1.1% for 2024. Administrative outcomes remain imprecise across cohorts. Excluding the 2021 cohort attenuates the revenue, volume, clinical-expense, and labor estimates, and a strict observed-transition definition attenuates them further (Appendix Table A4). The cohort evidence is therefore timing-sensitive and supports caution about exact magnitudes and dynamic accumulation.

6.3 Alternative Treatment Timing

The contemporaneous definition adds 346 adopters that reported AI use in an earlier year but did not meet the later persistent-use screen. Its estimates are close to the baseline for most outcomes: net patient revenue is 2.8% rather than 3.3%, discharge equivalents are 7.1% rather than 6.6%, and total hours are 3.2% in both. Excluding the 2021 cohort yields smaller estimates, while requiring a directly observed transition from nonuse to use yields still smaller volume and revenue estimates. Administrative outcomes are unstable under every definition. These results preserve the qualitative expansion pattern under the most natural contemporaneous rule, but show that exact magnitudes and the contribution of the earliest cohort are uncertain.

6.4 Business Stealing

To address the possibility that our estimates reflect business stealing from nearby non-adopting hospitals, we re-estimated the main revenue and volume models after excluding any non-adopter hospital located within 15 miles of an AI adopter. This restriction removes 787 nearby control hospitals, leaving 1,795 non-adopting hospitals in the control group. The exercise directly targets the concern that nearby controls may violate the Stable Unit Treatment Value Assumption because their revenue or volume could be affected by treatment at a neighboring AI hospital.

The results do not support the interpretation that the main volume findings are driven by local business stealing. After excluding nearby controls, estimated effects remain positive for both volume outcomes: inpatient discharges increase by 10.1% in the two-way fixed effects model and 8.6% in the SDID model, while discharge equivalents increase by 7.8% and 6.7%, respectively. These estimates are very similar to the estimates using the original control group. The revenue effect is smaller after the exclusion, falling from 4.2% to 2.8% in the two-way fixed effects model and from 3.3% to 2.6% in the SDID model, but it remains positive. Overall, the robustness check suggests that the observed increase in hospital volume is not explained by treated hospitals simply drawing patients away from nearby non-adopting hospitals.

6.5 Placebo Test: COVID-19 Public Health Emergency Funding

A potential concern with our estimates is that the revenue increases attributed to AI adoption could instead reflect differential receipt of CARES Act funding, which was distributed to hospitals beginning in 2020 (Cantor et al., 2021). Related work finds hospitals respond to financial improvements by investing part of those gains in technology (Richards, Shi and Whaley, 2025). If AI-adopting hospitals—which tend to be larger and better resourced—received disproportionately more federal relief funding, this could confound our revenue estimates.

To investigate, we conduct a placebo test that replaces AI adoption with high CARES Act funding as the treatment variable. We construct a cumulative CARES Act funding measure from the cost reports by setting missing values to zero and accumulating the running total over time for each hospital. We then classify hospitals that received above-median cumulative funding as "treated" beginning in 2020 (when CARES Act funding was first distributed) and estimate the effect on log net patient revenue using our SDID framework. If differential CARES Act funding were driving the revenue gains we attribute to AI adoption, we would expect a significant positive effect in this specification.

Table 6 examines absolute CARES funding and funding normalized by 2019 revenue or beds. The absolute-dollar and median normalized specifications produce negative revenue estimates, the normalized never-AI estimate is near zero, and only the top-quartile normalized threshold is slightly positive (0.36%). Thus, the CARES placebos do not reproduce the magnitude of the AI estimate. Because relief funding is only one pandemic-related shock and the exercise is not a complete model of pandemic exposure, it does not rule out all pandemic-era confounding.

6.6 Local COVID-19 Burden

We additionally merge official HHS facility-week reports of confirmed COVID-19 admissions, aggregate them to county-year admissions per 100,000 residents, and include the log of this rate in the outcome models. Coverage exceeds 99% in every study year. As Appendix Table A6 shows, county adjustment changes the estimates little: for example, net patient revenue moves from 2.83% to 2.73%, discharge equivalents from 7.13% to 7.01%, and total hours from 3.20% to 3.29%. A leave-one-out county measure that subtracts the focal hospital's admissions is similarly stable. This check addresses observed local hospital COVID burden, but not every local pandemic shock.

6.7 Excluding Critical Access Hospitals

Among the 4,093 short-term general hospitals in our sample, 1,298 (32%) are critical access hospitals (CAHs). CAHs differ from non-CAH short-term general hospitals in size, rurality, payment rules, and staffing constraints, all of which could affect both AI adoption and its measured effects. As a sensitivity test, we re-estimated our models after removing all CAHs from our sample. The test indicates that our headline results are not being driven by our decision to include CAHs. If anything, the core finance and labor effects are somewhat larger in the non-CAH sample: net patient revenue rises about 4.5% instead of 3.3%, total expenses about 4.1% instead of 2.9%, clinical expenses about 3.3% instead of 2.8%, and employee FTEs about 2.2% instead of 1.3%, while total paid hours are almost unchanged at roughly 3.1% versus 3.2% in the main sample. Volume effects also remain clearly positive, though slightly smaller, with discharge equivalents increasing about 5.5% rather than 6.6% in the full sample.

6.8 AI Adoption Intensity

As a more demanding treatment-definition robustness check, we re-estimate our main models after redefining the treated group as the subset of 168 hospitals that, in the 2024 AHA survey, reported the highest implementation level, "fully integrated," in all three of the core clinical AI applications: AI-assisted diagnostics, predictive analytics for patient care, and clinical decision support tools. We refer to these hospitals as heavy users because they are not simply reporting any AI adoption; they report full implementation across multiple central clinical AI domains. The control group remains the same 2,582 non-adopting hospitals, while the remaining adopters are excluded from this specification. This test therefore addresses the concern that our baseline treated group may include hospitals with only limited or ambiguous AI use.[5] It should be interpreted as a stricter eventual-heavy-adopter definition rather than the real-time effect of becoming fully implemented, since this implementation status is observed in 2024.

Even under this much narrower treatment definition, the core patterns in the paper largely survive. Total expenses remain positive and similar in magnitude to the main estimates, clinical expenses are essentially unchanged, discharge equivalents increase even more than in the baseline results, and labor inputs remain positive, with employee FTEs rising more and total paid hours remaining clearly positive. At the same time, this robustness check clarifies which findings are strongest and which are more sensitive. The net patient revenue estimate becomes smaller, falling from about 3.3 percent in the full adopter sample to about 1.7 percent, and is no longer statistically significant, potentially because of a smaller sample size. The modest administrative-expense reduction in the main sample also disappears, becoming slightly positive, and the negative patient-experience effects for nurse communication and receiving help are attenuated and no longer statistically distinguishable from zero. By contrast, several of the measured risk-adjusted mortality improvements remain and in some cases become larger in magnitude, especially for heart failure, pneumonia, and COPD, although stroke no longer shows a clear improvement. Taken together, these results suggest that the paper's main volume, labor, and several quality findings are not being driven by marginal or weak adopters, while the revenue and administrative results are more sensitive once treatment is restricted to a much smaller set of clearly intensive AI users.

[5]Though as shown in Section 7, among adopters who respond to the AHA IT supplement nearly 90% say they have AI integrated into their EHRs. So to the extent that AI in EHRs is one of the most impactful ways hospitals adopt AI, the broader sample has heavy use of AI as well.

6.9 AHA Survey Respondents Only

As an additional robustness check, we restricted the sample to hospitals that responded at least once to the AI adoption questions between 2021 and 2024, including hospitals that explicitly reported no AI use. This addresses the concern that some hospitals in our baseline control group may in fact have adopted AI but simply failed to answer the survey items. Under this restriction, the treated group remains unchanged at 1,511 adopters, while the control group falls from 2,582 to 1,364 explicit non-adopters. Re-estimating the main models in this responders-only sample yields results that are very similar to our baseline findings.

The core finance, volume, and labor patterns remain in place. Net patient revenue is estimated to rise by about 3.1 percent, clinical expenses remain positive, and both employment and total paid hours increase. Discharge equivalents rise by about 5.2 percent and inpatient discharges by about 6.8 percent. Administrative expenses retain a negative point estimate but remain part of the paper's less precise administrative evidence. The measured risk-adjusted mortality results are similar in sign to the main analysis. Taken together, this test suggests that hidden adopters among survey nonrespondents do not explain the main expansion pattern.

6.10 Controls for EHR Vendor

We estimate the main outcomes in the EHR-linked sample with EHR-vendor-by-year interactions, allowing each vendor's hospitals to experience different annual shocks, such as new vendor programs. Under the persistent-adopter definition, the revenue estimate moves from 2.86% to 2.44%, discharge equivalents from 5.46% to 3.48%, and total hours from 2.44% to 1.45% (Appendix Table A5). Clinical expenses remain positive, while the administrative-expense estimate moves from -1.01% to 0.03%. The contemporaneous-definition results show similar attenuation. These estimates indicate that vendor-specific changes explain part, but not all, of the finance and volume associations.

The available vendor variable is first observed in 2022 for this linked sample, so the vendor-by-year specification is not a clean control based exclusively on pre-treatment information. It may absorb genuine AI implementation that occurs through the EHR as well as vendor-level confounding. We therefore present it as a demanding sensitivity analysis rather than the preferred causal specification.

6.11 Alternative Estimation Approaches

To test robustness to the SDID approach, we estimate matched DiD models using cohort-specific propensity-score matching. For each cohort, we match adopters to never-adopters on baseline characteristics and pre-adoption outcome levels and trends, stack the matched cohorts, and estimate the event-study and summary regressions by OLS with matched cohort-by-hospital and matched cohort-by-year fixed effects. The summary estimates are directionally consistent with the primary SDID results but generally smaller: finance and volume outcomes remain positive, labor estimates are less precise, and administrative expenses and hours are not distinguishable from zero. Figures A6–A8 report the results.

7 Conclusion

The launch of ChatGPT and other AI applications has intensified interest in how AI may change health care (Fui-Hoon Nah et al., 2023). In U.S. hospitals, we find that AI adoption is followed by higher revenue, clinical expenses, patient volume, and total paid hours. Volume rises across service and payer types and is accompanied by higher occupancy rather than a commensurate expansion in beds. These findings are most naturally described as operational expansion and more intensive capacity use. Because output and measured inputs rise together, they do not establish per-unit productivity or net efficiency gains. Administrative-expense and administrative-hours estimates are too imprecise to support a savings claim.

There are clear differences in the characteristics of hospitals that have adopted AI and those that have not. AI-adopting hospitals are larger, more urban, more likely to be major teaching hospitals, and more likely to treat complex patient populations, which is consistent with previous studies (Nong et al., 2025; Bin Abdul Baten, 2024; Goldfarb, He and Teodoridis, 2025; Poon et al., 2025). They are also much less likely to be critical access hospitals than non-adopters. Organizations with greater resources are more capable of financing AI adoption and hiring staff with AI development expertise (Nong, Maurer and Dwivedi, 2025; Burns et al., 2025). These differences in AI adoption raise the question of whether smaller, more rural (and presumably financially less stable) hospitals have access to the benefits of AI and what the future looks like for these hospitals, their providers, and their patients if they are unable to access the same level of benefits.

These findings should be interpreted in light of several limitations. First, treatment is self-reported, survey wording changes in 2024, and the baseline persistent-adopter definition uses later implementation status to screen earlier reports. Contemporaneous and alternative timing rules preserve several core patterns but show that exact magnitudes, especially those involving the 2021 cohort, are sensitive. Second, some hospitals labeled as non-adopters may be unobserved adopters, although the responders-only analysis is similar. Third, SDID requires the synthetic controls to remain valid counterfactuals after treatment. The diagnostics show broad donor support and diffuse weights, with strong pre-period fit for finance and volume but weaker fit for labor hours. Fourth, pandemic, EHR-vendor, and local-market shocks may still confound the estimates despite the new sensitivity analyses. Finally, our measured outcomes do not permit a complete welfare or per-unit productivity analysis.

The mechanism evidence points most clearly to throughput expansion and tighter capacity use, alongside denser documentation and coding, rather than to a single dominant bedside clinical pathway. The declines in measured risk-adjusted mortality should be interpreted cautiously: unadjusted mortality and the available process measures do not show corresponding sustained improvements, while documented severity rises. Future research should identify which tools generate which changes, measure per-unit productivity and patient welfare directly, and disentangle clinical improvement from documentation and risk capture. It should also examine whether implementation can be made accessible to smaller and rural hospitals rather than remaining concentrated in large, well-resourced systems.

References

Acemoglu, Daron, David Autor, and Simon Johnson. 2023. "Can we have pro-worker AI?" Centre for Economic Policy Research. https://cepr.org/system/files/publication-files/191183-policy_insight_123_can_we_have_pro_worker_ai_choosing_a_path_of_machines_in_service_of_minds.pdf.

AdventHealth. 2022. "AdventHealth Mission Control Drives Improved Emergency Department Outcomes." *https://www.adventhealth.com/news/adventhealth-mission-control-drives-improved-emergency-department-outcomes*, Accessed July 5, 2026.

Agha, Leila. 2014. "The effects of health information technology on the costs and quality of medical care." *Journal of Health Economics*, 34: 19–30.

American Hospital Association. 2026. "Fast Facts on U.S. Hospitals, 2026." American Hospital Association. https://www.aha.org/statistics/fast-facts-us-hospitals.

Arkhangelsky, Dmitry, Susan Athey, David A Hirshberg, Guido W Imbens, and Stefan Wager. 2021. "Synthetic difference-in-differences." *American Economic Review*, 111(12): 4088–4118.

Autor, David. 2024. "Applying AI to rebuild middle class jobs." National Bureau of Economic Research. Working paper 32140.

Babel, Aditi, Richi Taneja, Franco Mondello Malvestiti, Alessandro Monaco, and Shaantanu Donde. 2021. "Artificial intelligence solutions to increase medication adherence in patients with non-communicable diseases." *Frontiers in Digital Health*, 3: 669869.

Babina, Tania, Anastassia Fedyk, Alex He, and James Hodson. 2024. "Artificial Intelligence, Firm Growth, and Product Innovation." *Journal of Financial Economics*, 151: 103745.

Babina, Tania, Anastassia Fedyk, Alex He, and James Hodson. 2025. "Artificial Intelligence Makes Firm Operating Performance Less Volatile." *AEA Papers and Proceedings*, 115: 35–39.

Baird, Matthew, Mar Carpanelli, Brian Xu, and Kevin Xu. 2026. "Firms' GitHub Copilot Adoption and Labor Market Outcomes for Software Engineers." *Contemporary Economic Policy.*

Bin Abdul Baten, Redwan. 2024. "How are US hospitals adopting artificial intelligence? Early evidence from 2022." *Health Affairs Scholar*, 2(10): qxae123.

Brynjolfsson, Erik, Bharat Chandar, and Ruyu Chen. 2025. "Canaries in the Coal Mine? Six Facts about the Recent Employment Effects of Artificial Intelligence." Stanford University. https://digitaleconomy.stanford.edu/publications/canaries-in-the-coal-mine/.

Brynjolfsson, Erik, Danielle Li, and Lindsey Raymond. 2025. "Generative AI at work." *The Quarterly Journal of Economics*, 140(2): 889–942.

Buenaventura, Maya, Anton Shenk, Anujin Nergui, Sean Mann, Flannery C. Dolan, Karishma V. Patel, Jim Mignano, and Jonathan W. Welburn. 2025. *Artificial Intelligence Adoption and Sectoral Transformation: Implications for Health Care, Financial Services, Climate and Energy, and Transportation.* Santa Monica, CA:RAND Corporation.

Burns, Michael L, Ssu-Ying Chen, Chu-An Tsai, John Vandervest, Balaji Pandian, Paige Nong, David A Hanauer, Andrew Rosenberg, and Jodyn Platt. 2025. "Generative AI costs in large healthcare systems, an example in revenue cycle." *npj Digital Medicine*, 8(1): 579.

Cantor, Jonathan, Nabeel Qureshi, Brian Briscombe, Justin Chapman, and Christopher M. Whaley. 2021. "Association Between COVID-19 Relief Funds and Hospital Characteristics in the US." *JAMA Health Forum*, 2(10): e213325.

Chen, Irene Y, and Anna Zink. 2025. "AI tools in human hands: measuring real-world impact in the emergency department."

Chen, Jie, and Alice Shijia Yan. 2025. "Hospital artificial intelligence/machine learning adoption by neighborhood deprivation." *Medical Care*, 63(3): 227–233.

Cutler, David M. 2023. "What Artificial Intelligence Means for Health Care." *JAMA Health Forum*, 4(7): e232652.

Cutler, David M., and Richard J. Zeckhauser. 2000. "Chapter 11 - The Anatomy of Health Insurance." In *Handbook of Health Economics*. Vol. 1 of *Handbook of Health Economics*, , ed. Anthony J. Culyer and Joseph P. Newhouse, 563–643. Elsevier.

Dafny, Leemore S. 2005. "How Do Hospitals Respond to Price Changes?" *American Economic Review*, 95(5): 1525–1547.

Deloitte. 2024. "How AI can help hospitals strengthen their financial performance and reduce clinician burnout." Deloitte. https://www.deloitte.com/us/en/services/consulting/articles/artificial-intelligence-in-hospitals-financial-performance-clinical-burnout.html.

Dranove, David, and Craig Garthwaite. 2024. "Artificial Intelligence, the Evolution of the Health Care Value Chain, and the Future of the Physician." In *The Economics of Artificial Intelligence: Health Care Challenges. Working Paper Series*. University of Chicago Press.

Eloundou, Tyna, Sam Manning, Pamela Mishkin, and Daniel Rock. 2024. "GPTs Are GPTs: Labor Market Impact Potential of LLMs." *Science*, 384(6702): 1306–1308.

Frank, Morgan R, David Autor, James E Bessen, Erik Brynjolfsson, Manuel Cebrian, David J Deming, Maryann Feldman, Matthew Groh, José Lobo, Esteban Moro, et al. 2019. "Toward understanding the impact of artificial intelligence on labor." *Proceedings of the National Academy of Sciences*, 116(14): 6531–6539.

Freedman, Seth, Haizhen Lin, and Jeffrey Prince. 2018. "Information Technology and Patient Health: Analyzing Outcomes, Populations, and Mechanisms." *American Journal of Health Economics*, 4(1): 51–79.

Fui-Hoon Nah, Fiona, Ruilin Zheng, Jingyuan Cai, Keng Siau, and Langtao Chen. 2023. "Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration."

Goldfarb, Avi, Xianda He, and Florenta Teodoridis. 2025. "Patterns of Artificial Intelligence Adoption by Hospitals." *AEA Papers and Proceedings*.

Hampole, Menaka, Dimitris Papanikolaou, Lawrence D.W. Schmidt, and Bryan Seegmiller. 2025. "Artificial Intelligence and the Labor Market." National Bureau of Economic Research Working Paper 33509.

Horn, Danea, Adam Sacarny, and Annetta Zhou. 2022. "Technology adoption and market allocation: The case of robotic surgery." *Journal of Health Economics*, 86: 102672.

Howell, Michael D, Greg S Corrado, and Karen B DeSalvo. 2024. "Three epochs of artificial intelligence in health care." *JAMA*, 331(3): 242–244.

Hui, Xiang, Oren Reshef, and Luofeng Zhou. 2024. "The Short-Term Effects of Generative Artificial Intelligence on Employment: Evidence from an Online Labor Market." *Organization Science*, 35(6): 1977–1989.

Humlum, Anders, and Emilie Vestergaard. 2025*a*. "Large Language Models, Small Labor Market Effects." National Bureau of Economic Research Working Paper 33777.

Humlum, Anders, and Emilie Vestergaard. 2025*b*. "The unequal adoption of ChatGPT exacerbates existing inequalities among workers." *Proceedings of the National Academy of Sciences*, 122(1): e2414972121.

Hwang, Yeon-Mi, Madelena Y. Ng, Malvika Pillai, Michelle P. Sahai, and Tina Hernandez-Boussard. 2026. “The Landscape of AI Implementation in US Hospitals.” *Nature Health*, 1(1): 99–112.

Jackson, Amber. 2023. “ChatGPT one year on: adoption, transformation, regulation.” *AI Magazine*.

Jha, Ashish K, E John Orav, Jie Zheng, and Arnold M Epstein. 2008. “Patients’ perception of hospital care in the United States.” *New England Journal of Medicine*, 359(18): 1921–1931.

Jiang, Yulei, Alexandra V Edwards, and Gillian M Newstead. 2021. “Artificial intelligence applied to breast MRI for improved diagnosis.” *Radiology*, 298(1): 38–46.

Kass-Hout, Taha. 2025. “How AI Is Improving Care and Helping Hospitals Reduce Costs.” *https://www.gehealthcare.com/en-us/insights/article/how-ai-is-improving-care-and-helping-hospitals-reduce-costs*, Accessed July 5, 2026.

Khalifa, Mohamed, and Mona Albadawy. 2024. “AI in diagnostic imaging: revolutionising accuracy and efficiency.” *Computer Methods and Programs in Biomedicine Update*, 5: 100146.

Kocher, Bob, Brian Zhao, and Erin Duffy. 2026. “Thinking about the Impact of Artificial Intelligence on U.S. Health Care Costs and Spending Growth.” *NEJM Catalyst*, 7(8): CAT.25.0509.

Melnick, Glenn, June F. O’Leary, Ben A. Zaniello, and Luis Abrishamian. 2022. “COVID–19 driven decline in emergency visits: Has it continued, is it permanent, and what does it mean for emergency physicians?” *The American Journal of Emergency Medicine*, 61: 64–67.

Nguyen, Thuy D, Christopher M Whaley, Kosali Simon, Neil Mehta, Hao Yu, Ryan K McBain, Ateev Mehrotra, and Jonathan H Cantor. 2025. "Adoption of Artificial Intelligence in the Health Care Sector." *JAMA Health Forum*, 6(11): e255029.

Nong, Paige, Eric Maurer, and Roli Dwivedi. 2025. "The urgency of centering safety-net organizations in AI governance." *npj Digital Medicine*, 8(1): 117.

Nong, Paige, Julia Adler-Milstein, Nate C Apathy, A Jay Holmgren, and Jordan Everson. 2025. "Current Use And Evaluation Of Artificial Intelligence And Predictive Models In US Hospitals: Article examines uses and evaluation of artificial intelligence and predictive models in US hospitals." *Health Affairs*, 44(1): 90–98.

Obermeyer, Ziad. 2025. "Bedside to Bench — AI and the New Science of Medicine." *New England Journal of Medicine*, 393(23): 2287–2289.

Oseran, Andrew S, Rishi K Wadhera, E John Orav, and Jose F Figueroa. 2023. "Effect of Medicare Advantage on hospital readmission and mortality rankings." *Annals of Internal Medicine*, 176(4): 480–488.

Poon, Eric G, Christy Harris Lemak, Juan C Rojas, Janet Guptill, and David Classen. 2025. "Adoption of artificial intelligence in healthcare: survey of health system priorities, successes, and challenges." *Journal of the American Medical Informatics Association*, 32(7): 1093–1100.

Prager, Elena, and Matt Schmitt. 2021. "Employer consolidation and wages: Evidence from hospitals." *American Economic Review*, 111(2): 397–427.

Richards, Michael R., and Christopher M. Whaley. 2024. "Hospital behavior over the private equity life cycle." *Journal of Health Economics*, 97: 102902.

Richards, Michael R., Maggie Shi, and Christopher M. Whaley. 2025. "Economic Shocks and Healthcare Capital Investments."

Rotenstein, Lisa S., A. Jay Holmgren, Robert Thombley, et al. 2026. "Changes in Clinician Time Expenditure and Visit Quantity With Adoption of Artificial Intelligence–Powered Scribes: A Multisite Study." *JAMA*. Advance online publication.

Sahni, Nikhil R., George Stein, Rodney Zemmel, and David Cutler. 2024. "The Potential Impact of Artificial Intelligence on Health Care Spending." In *The Economics of Artificial Intelligence: Health Care Challenges. NBER Chapters*. University of Chicago Press.

Sahni, Nikhil R., Kate O'Gorman, Rahul Agarwal, Heather Bello Thornhill, and David Cutler. 2026. "The Value Opportunity from Artificial Intelligence in U.S. Health Care Spending." *NEJM Catalyst*, 7(4): CAT.26.0037.

Schmitt, Matt. 2017. "Do hospital mergers reduce costs?" *Journal of Health Economics*, 52: 74–94.

Sekandi, Juliet Nabbuye, Weili Shi, Ronghang Zhu, Patrick Kaggwa, Ernest Mwebaze, and Sheng Li. 2023. "Application of artificial intelligence to the monitoring of medication adherence for tuberculosis treatment in Africa: algorithm development and validation." *JMIR AI*, 2(1): e40167.

Shuaib, Ali. 2024. "Transforming healthcare with AI: Promises, pitfalls, and pathways forward." *International Journal of General Medicine*, 1765–1771.

Singla, Alex, Alexander Sukhovetsky, Bryce Hall, Lareina Yee, and Michael Chui. 2025. "The state of AI in 2025: agents, innovation, and transformation." McKinsey & Company. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai#/.

Stanford University Human-Centered Artificial Intelligence. 2025. "Artificial Intelligence Index Report 2025." Stanford University. https://hai.stanford.edu/ai-index/2025-ai-index-report.

Sun, Liyang, and Sarah Abraham. 2021. "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects." *Journal of Econometrics*, 225(2): 175–199.

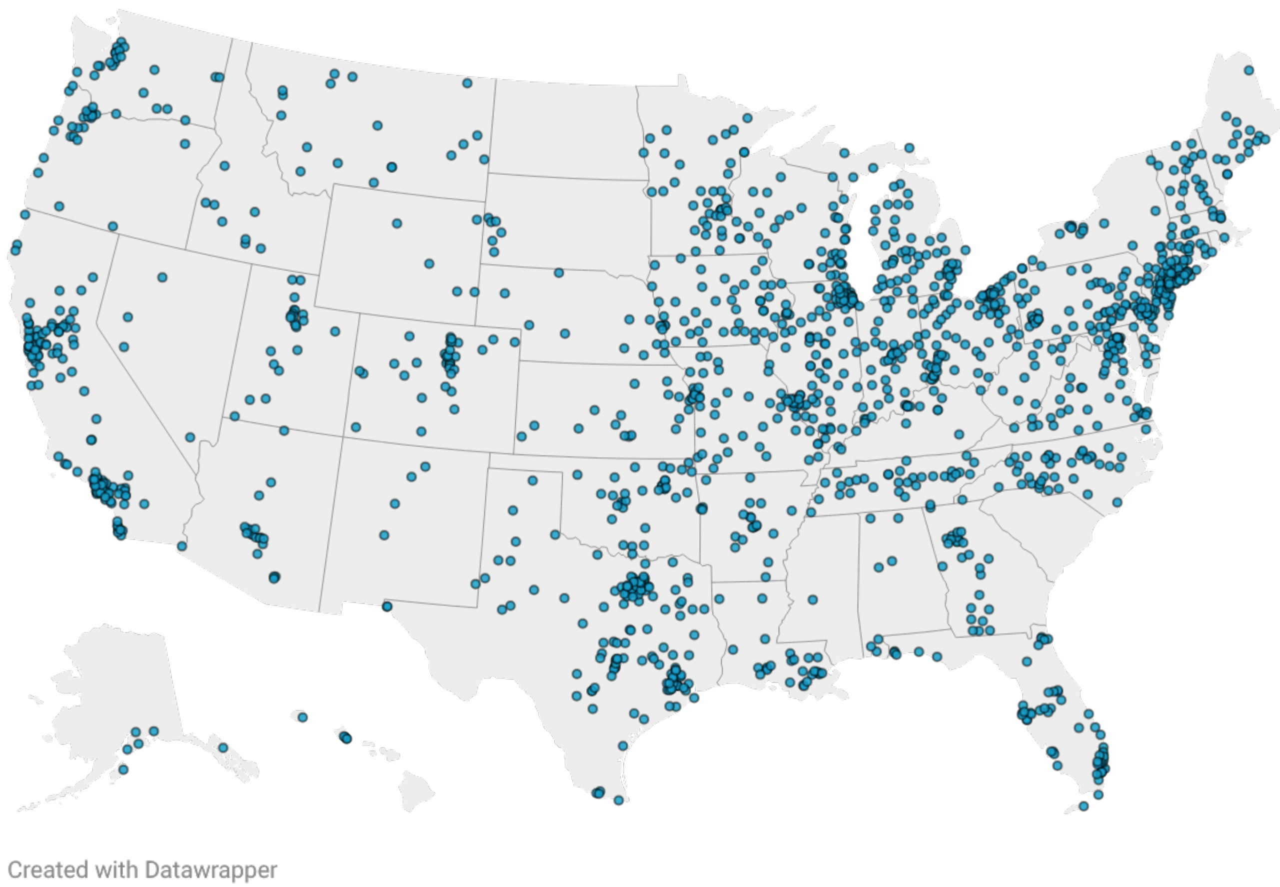
Wachter, Robert M, and Erik Brynjolfsson. 2024. "Will generative artificial intelligence deliver on its promise in health care?" *JAMA*, 331(1): 65–69.

Werner, Rachel M, and Eric T Bradlow. 2006. "Relationship between Medicare's hospital compare performance measures and mortality rates." *JAMA*, 296(22): 2694–2702.

Zink, Anna, Sarah Morriss, Anuj Gangopadhyaya, and Ziad Obermeyer. 2023. "Building Equitable Artificial Intelligence in Health Care." *Urban Institute*.

Tables and Figures

Figure 1: Map of AI-adopting hospitals



Notes: The figure shows the locations of all 1,511 AI-adopting hospitals in our analytic sample.

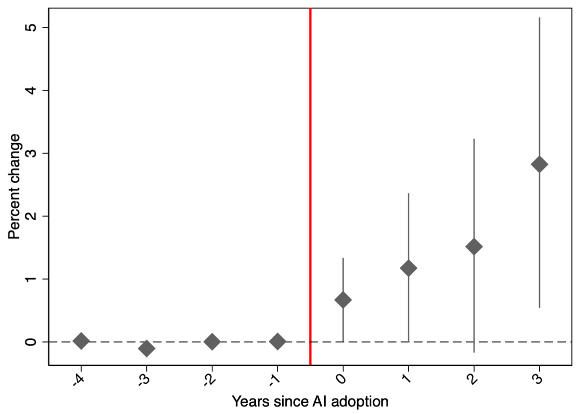
Table 1: Baseline characteristics of hospital AI adopters and non-adopters, 2017

	AI adopters	**Non-adopters**	**Std. diff.**
N	1,511	2,582	
Finances and Volume			
Net patient revenue (millions $)	339 (483)	139 (215)	0.54
Total expenses (millions $)	342 (491)	139 (227)	0.53
Expenses for clinical activities (millions $)	178 (254)	72 (116)	0.54
Expenses for administrative activities (millions $)	60 (82)	25 (39)	0.55
Discharge equivalents	18,768 (20,242)	9,064 (11,865)	0.58
Labor			
Employees on payroll (FTEs)	1,595 (2,221)	704 (1,084)	0.51
Total paid hours (thousands)	4,094 (4,956)	2,271 (2,681)	0.46
Paid hours, administrative and general (thousands)	416 (641)	244 (315)	0.34
Quality Outcomes			
30-day mortality: heart attack (%)	13.4 (1.2)	13.7 (1.2)	-0.29
30-day mortality: heart failure (%)	11.8 (1.6)	12.1 (1.6)	-0.21
30-day mortality: pneumonia (%)	15.7 (1.9)	16.2 (2.0)	-0.24
30-day mortality: COPD (%)	8.1 (1.1)	8.1 (1.1)	-0.07
30-day mortality: stroke (%)	14.5 (1.7)	14.7 (1.6)	-0.11
30-day mortality: CABG surgery (%)	3.1 (0.8)	3.4 (0.9)	-0.33
30-day all-cause readmission rate (%)	15.3 (0.8)	15.3 (0.8)	-0.11
Patient Experience			
Nurses "Always" communicated well (%)	79.8 (4.4)	80.1 (5.5)	-0.06
Doctors "Always" communicated well (%)	80.9 (4.3)	82.1 (5.7)	-0.24
"Always" received help as soon as wanted (%)	66.6 (7.6)	68.5 (9.3)	-0.22
Hospital rating of 9 or 10 (%)	73.0 (7.1)	72.1 (8.6)	0.12
Would definitely recommend the hospital (%)	73.0 (8.0)	70.9 (9.3)	0.24
Hospital Characteristics			
Beds	221 (241)	114 (139)	0.54
Part of a system	0.87 (0.33)	0.61 (0.49)	0.64
Rural-urban continuum code	2.7 (2.2)	4.1 (2.7)	-0.58
Critical Access Hospital (CAH)	0.19 (0.39)	0.39 (0.49)	-0.46
For-profit	0.07 (0.25)	0.22 (0.41)	-0.43
Major teaching hospital	0.13 (0.34)	0.05 (0.21)	0.31
Medicare case-mix index (CMI)	1.68 (0.30)	1.56 (0.32)	0.41
Medicare inpatient discharge share	0.36 (0.15)	0.44 (0.17)	-0.49
Medicaid inpatient discharge share	0.09 (0.09)	0.10 (0.10)	-0.08

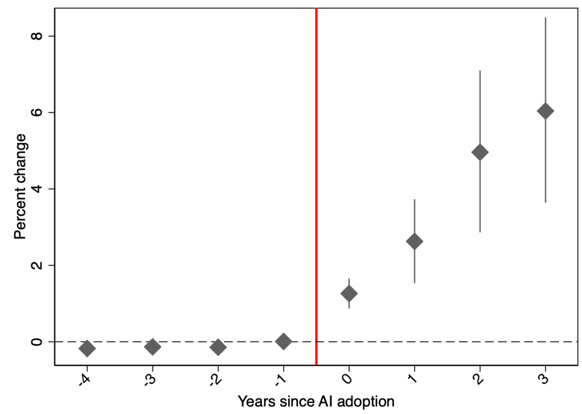
Sources: AHA's Annual Survey (AI adoption), CMS's Hospital Compare (quality measures), Hospital Cost Report data (all other measures). *Notes:* Entries are means, with standard deviations in parentheses. Binary variables are reported as proportions. Standardized differences equal the adopter mean minus the non-adopter mean, divided by the pooled standard deviation. Statistics use available observations, so sample sizes may differ across measures.

Figure 2: Labor Market Impact Event Studies

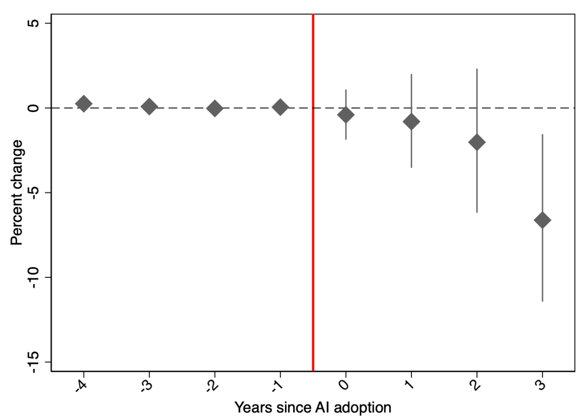
(a) Employees on payroll (FTEs)



(b) Total paid hours



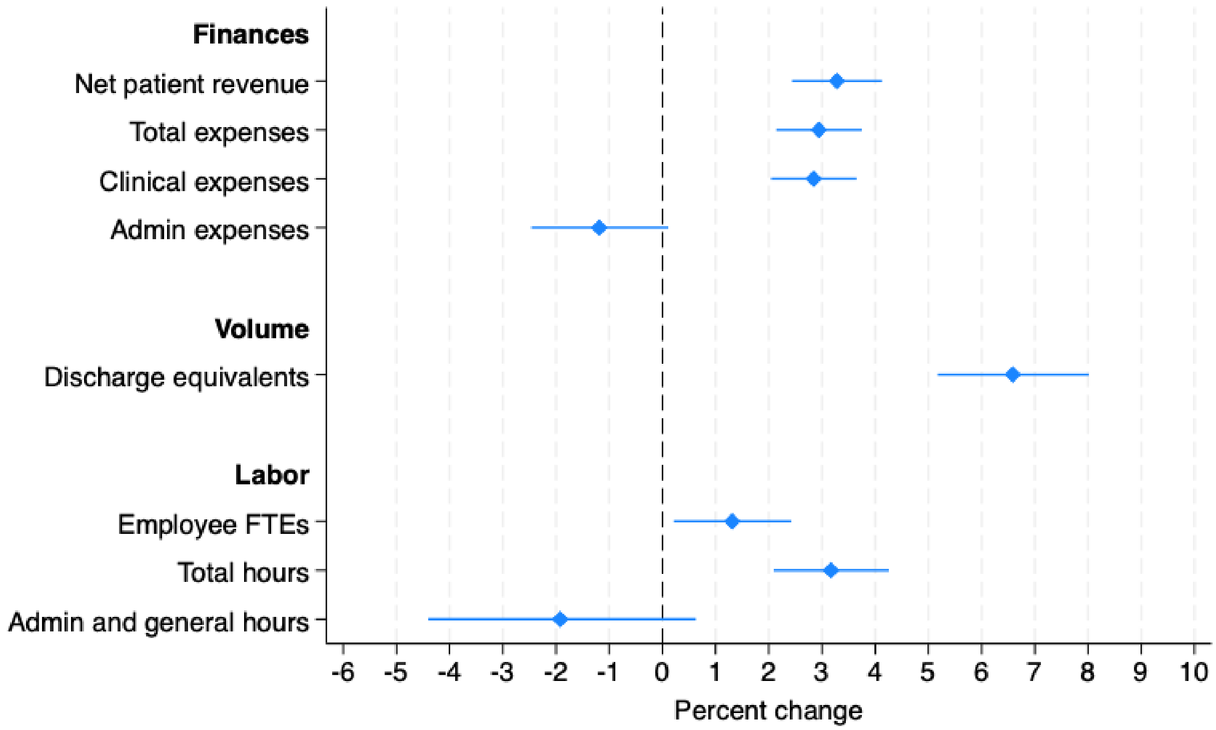
(c) Paid hours, administrative and general



Sources: AHA's Annual Survey (AI adoption) and Hospital Cost Report data as collected by RAND (outcomes).

Notes: The figures show estimates from a synthetic difference-in-differences event study where the log of each panel's title is the dependent variable. Coefficient estimates have been transformed using the formula $100 \times (\exp(\text{coef}) - 1)$ to create the percentage effects shown in the figures. Inference uses the SDID bootstrap procedure, with resampling clustered at the hospital level and 50 bootstrap replications. The error bars represent 95% bootstrap confidence intervals.

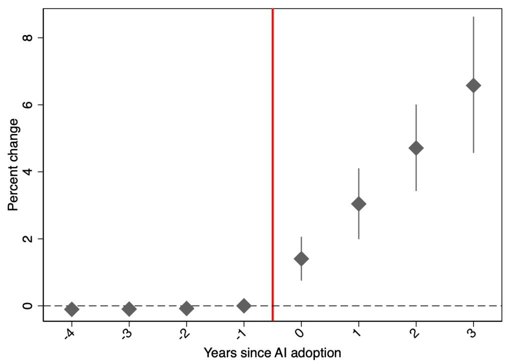
Figure 3: Summary of AI adoption effects on revenues, expenses, volume, employment, efficiency



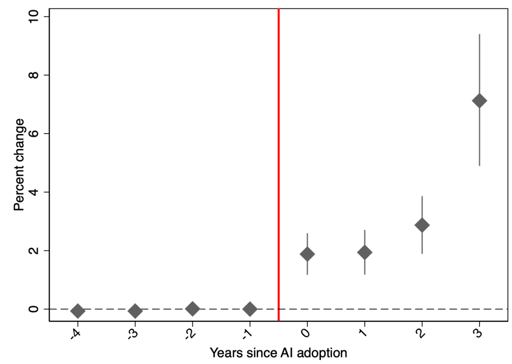
Notes: The figure shows estimates from a synthetic difference-in-differences model where the dependent variable for each model corresponds to the logged version of the variable listed on the vertical axis. Coefficient estimates have been transformed using the formula $100 \times (\exp(\text{coef}) - 1)$ to create the percentage effects shown in the figure. Inference uses the SDID bootstrap procedure, with resampling clustered at the hospital level and 50 bootstrap replications. The error bars represent 95% bootstrap confidence intervals.

Figure 4: Finances and volume event studies

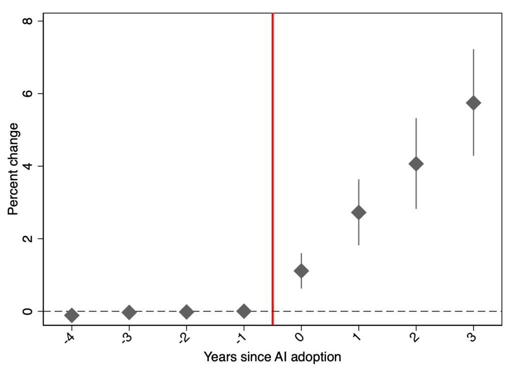
(a) Net patient revenue



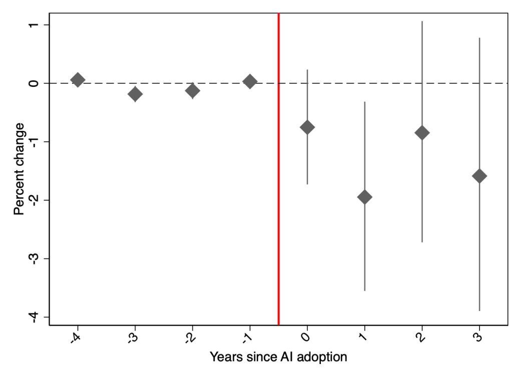
(b) Total expenses



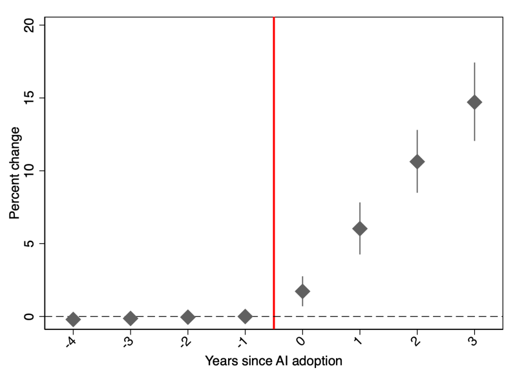
(c) Expenses for clinical activities



(d) Administrative expenses



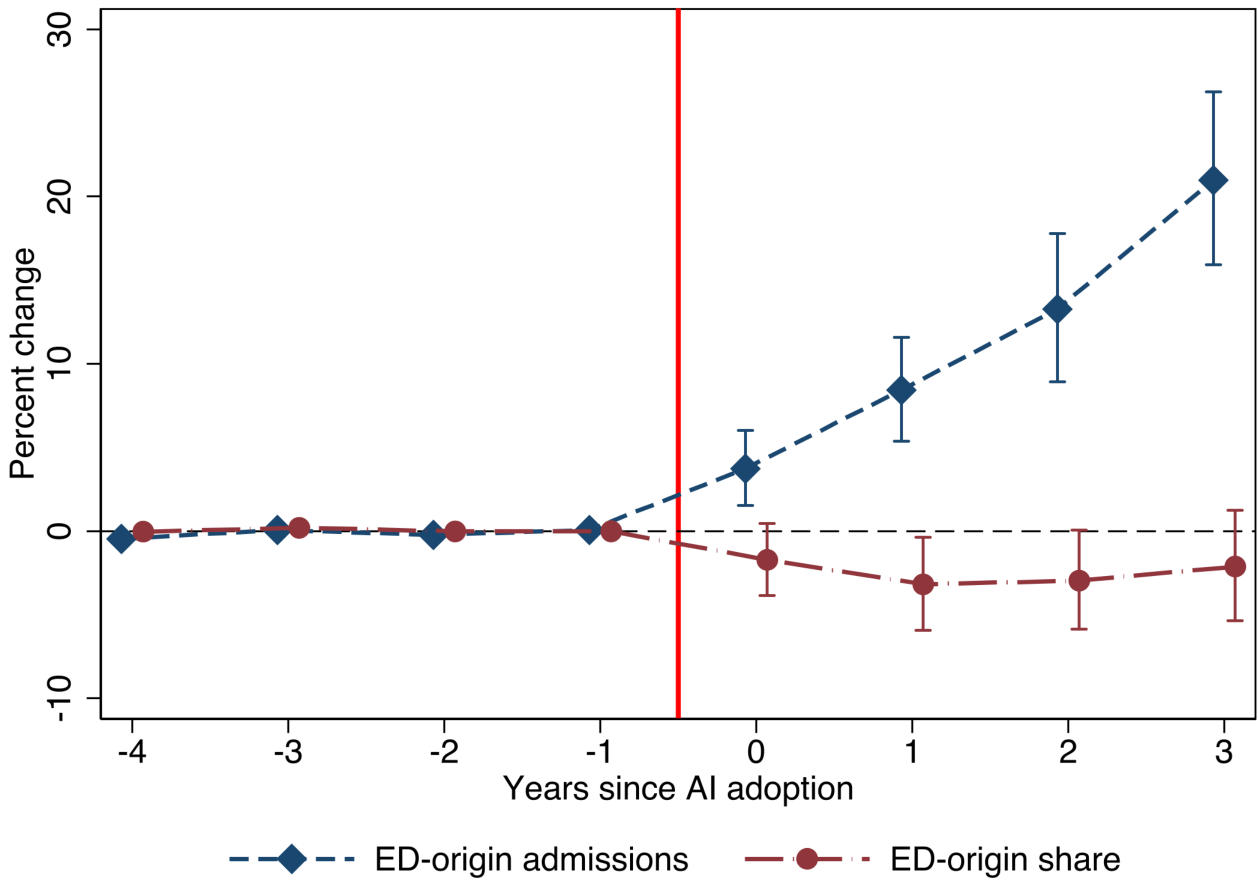
(e) Discharge equivalents



Sources: AHA's Annual Survey (AI adoption) and Hospital Cost Report data as collected by RAND (expenses and discharge equivalents).

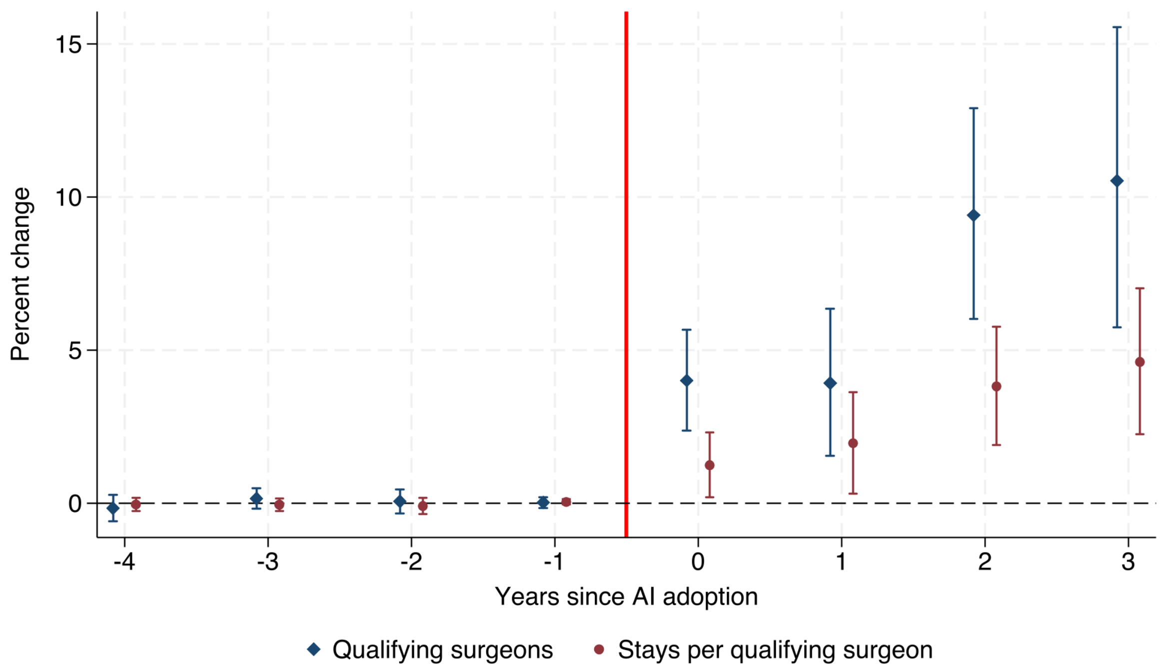
Notes: The figures show estimates from a synthetic difference-in-differences event study where the log of each panel's title is the dependent variable. Coefficient estimates have been transformed using the formula $100 \times (\exp(\text{coef}) - 1)$ to create the percentage effects shown in the figures. Inference uses the SDID bootstrap procedure, with resampling clustered at the hospital level and 50 bootstrap replications. The error bars represent 95% bootstrap confidence intervals.

Figure 5: Changes in ED volume and originating admissions following AI adoption



Notes: The figure presents synthetic difference-in-differences event-study estimates using a balanced 2017–2024 panel of 3,317 hospitals in the FFS MedPAR data. An inpatient admission is classified as originating in the emergency department if the MedPAR stay contains a positive emergency-room service charge. The outcomes are the logged number of ED-origin admissions and the logged share of inpatient admissions originating in the ED. Estimates are transformed using $100 \times [\exp(\widehat{\beta}) - 1]$ and therefore represent percentage changes. Event times -4 through -1 are pre-adoption placebo estimates, and the vertical line denotes AI adoption. Inference uses the SDID bootstrap procedure with resampling clustered at the hospital level and 50 bootstrap replications. Error bars represent 95% bootstrap confidence intervals.

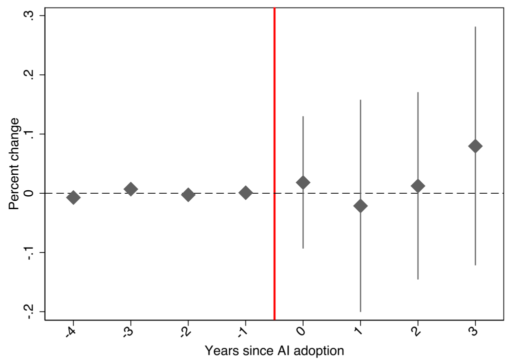
Figure 6: Changes in Surgical Volume Following AI adoption



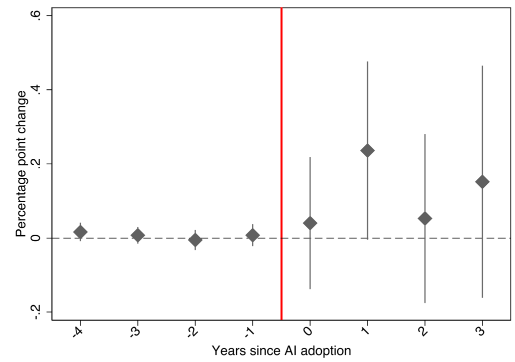
Notes: The figure reports synthetic difference-in-differences event-study estimates using Medicare FFS Carrier claims linked to MedPAR inpatient stays from 2017–2024. A surgeon NPI is counted as qualifying only if it is associated with at least five linked surgical stays at that hospital during the year. The sample is restricted to hospitals with at least 50 linked surgical stays in every sample year. The outcomes are the log number of qualifying surgeon NPIs and the log number of stays per qualifying surgeon. Estimates are transformed using $100 \times [\exp(\widehat{\beta}) - 1]$ and therefore represent percentage changes. Event time zero denotes the year of AI adoption; pre-adoption estimates are SDID placebo estimates. Error bars show 95% confidence intervals obtained using the SDID bootstrap procedure, with resampling clustered at the hospital level and 50 bootstrap replications.

Figure 7: Patient Composition Event Studies

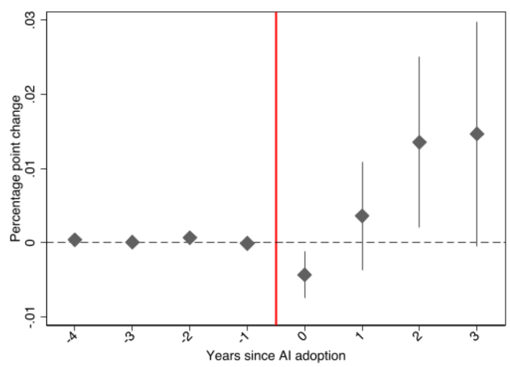
(a) Patient age



(b) Share Female



(c) Case-mix index



Notes: The figure presents synthetic difference-in-differences event-study estimates. Panels A and B use FFS MedPAR admissions from 2017–2024; Panel C uses the hospital-wide Medicare inpatient case-mix index reported in the Hospital Cost Report data. The case-mix index measures average expected resource use based on patients' assigned diagnosis-related groups. Mean age and the case-mix index are estimated in logs and transformed using $100 \times [\exp(\widehat{\beta}) - 1]$, so their estimates represent percentage changes. The female-share estimates represent percentage-point changes. Event time zero denotes the year of AI adoption, and the pre-adoption estimates are SDID placebo estimates. Inference uses a nonparametric bootstrap procedure with resampling clustered at the hospital level and 50 bootstrap replications. Error bars represent 95% bootstrap confidence intervals.

Table 2: Operational and Throughput Mechanism Evidence

Outcome	SDID estimate	p-value	Interpretation
Discharge equivalents	+6.6%	< 0.001	Broad increase in hospital volume after AI adoption.
ED visits, log specification	+0.0427 ($\approx$ 4.4%)	< 0.001	More emergency visits are handled by adopting hospitals.
ED visits, level specification	+2,190 visits / hospital-year	< 0.001	Large absolute increase in ED throughput.
Occupancy, log specification	+0.0638 ($\approx$ 6.4%)	< 0.001	Higher bed utilization and tighter capacity use.
Occupancy, level specification	+0.0301	< 0.001	About a 3.0 percentage-point increase in occupancy.
Hospital-wide inpatient case-mix index, log specification	+0.0007	0.656	No evidence that adopters are treating observably sicker inpatient cases overall.
Hospital-wide inpatient case-mix index, level specification	+0.0046	0.110	Reinforces the null case-mix result.

Notes: The occupancy and case-mix estimates come from dedicated follow-up SDID models, and the ED estimates come from the dedicated emergency-department volume analysis. Estimates use the persistent-adopter definition. p-values are based on 50 nonparametric bootstrap replications with resampling at the hospital level.

Table 3: Documentation and Coding-Intensity Mechanism Evidence

Outcome	SDID estimate	p-value	Interpretation
Total diagnoses per admission	+0.236	0.011	Adopters code more diagnoses overall.
Secondary diagnoses per admission	+0.236	0.002	Extra coding is especially visible on secondary diagnoses.
Newly coded chronic conditions absent from prior claims	−0.001	0.279	No evidence of a surge in newly "discovered" chronic conditions.
Unadjusted 30-day mortality	−0.0017	0.139	Raw mortality is essentially flat in the pooled cohort panel.
Share of diagnoses marked present on admission	−0.0063	0.017	The present-on-admission share falls slightly.
Present-on-admission diagnosis count	+0.153	0.007	Hospitals code more diagnoses as present on admission after adoption.
Non-present-on-admission diagnosis count	+0.111	0.004	Hospitals also code more diagnoses not marked present on admission.
Hospital-wide diagnoses per Medicare admission (screened sample)	+0.088	0.085	Broad diagnosis density rises modestly outside the four focal cohorts as well.
Hospital-wide submitted charges per Medicare admission (screened sample)	+$758	0.004	Billed intensity per Medicare admission increases after adoption.
Hospital-wide Medicare reimbursement per Medicare admission (screened sample)	−$353	< 0.001	Realized Medicare payment per stay falls rather than rises, inconsistent with a simple payment-per-claim upcoding story.

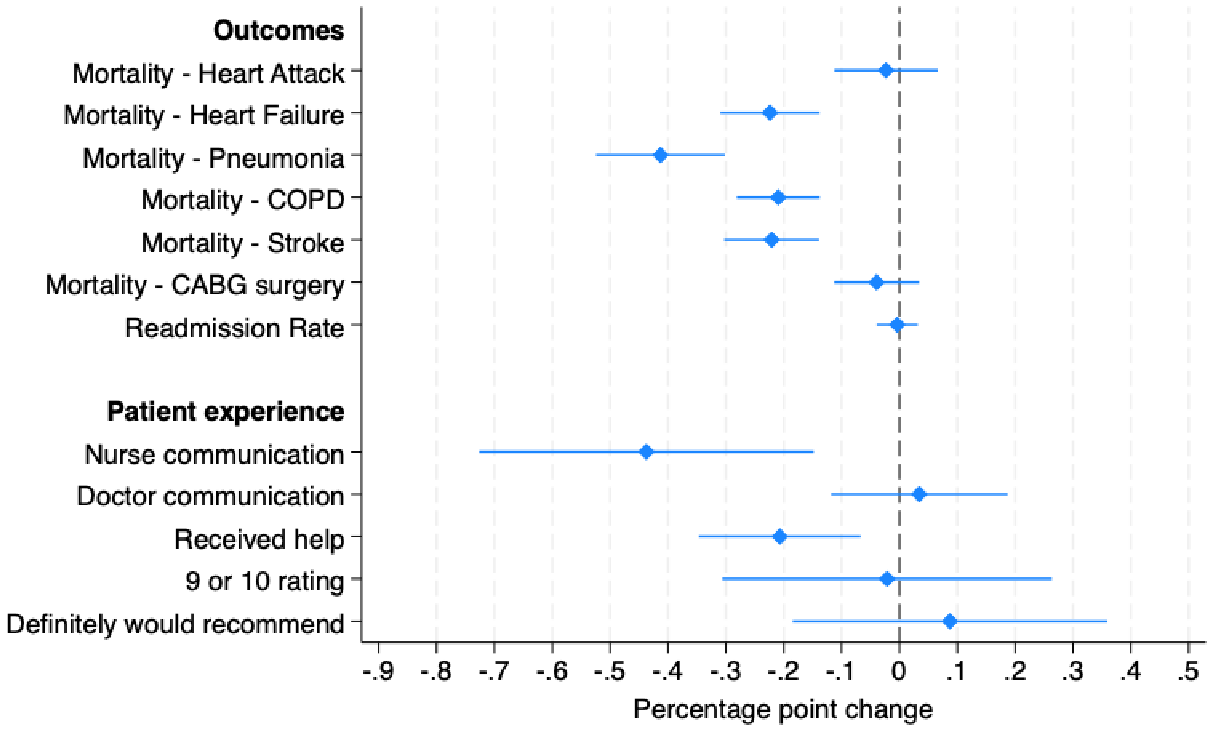
Notes: The first seven rows pool admissions with a principal diagnosis of COPD, pneumonia, heart failure, or stroke; these are the underlying clinical-condition cohorts used for the condition-specific coding and mortality analyses. The final three rows instead use all Medicare FFS acute inpatient admissions at each hospital and therefore test whether the coding patterns extend beyond those four conditions. The hospital-wide rows use the cleaned sample that drops any hospital ever showing nonpositive mean Medicare reimbursement or fewer than 10 Medicare FFS admissions in a year; a stricter screen using fewer than 25 admissions yields very similar patterns. The estimates describe changes in coding and billing measures and do not by themselves identify a causal coding contribution to revenue. p-values are based on 50 nonparametric bootstrap replications with resampling at the hospital level.

Table 4: Primary Outcomes with System-Clustered Inference

Outcome	Estimate (percent)	95% CI Lower	95% CI Upper	Holm-adjusted p
Net patient revenue	2.82	1.63	4.02	< 0.001
Total expenses	2.06	1.00	3.12	< 0.001
Clinical expenses	2.68	1.55	3.82	< 0.001
Administrative expenses	−1.39	−3.90	1.18	0.571
Discharge equivalents	7.05	5.09	9.05	< 0.001
Employee FTEs	1.06	−0.27	2.41	0.358
Total paid hours	3.19	1.82	4.59	< 0.001
Administrative paid hours	−2.58	−8.57	3.80	0.571

Notes: This table reports SDID estimates under the contemporaneous-adoption definition, which classifies a hospital as treated beginning in the first year it reports AI use without using later survey responses to revise earlier treatment status. Log-point estimates are transformed using $100 \times [\exp(\widehat{\tau}) - 1]$. Confidence intervals are based on 200 nonparametric bootstrap replications clustered by hospital system; independent hospitals form their own clusters. The final column applies Holm's family-wise-error correction across the eight outcomes. Benjamini–Hochberg adjustment yields the same substantive classification. Dynamic figures elsewhere in the paper use the persistent-adopter definition and are retained to show timing; Appendix Table A4 compares definitions directly.

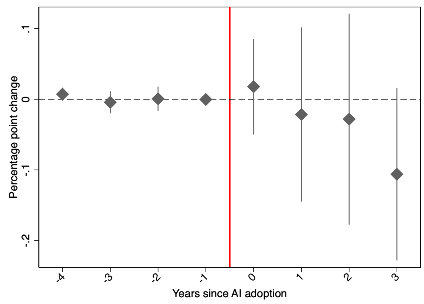
Figure 8: Summary of AI adoption effects on quality



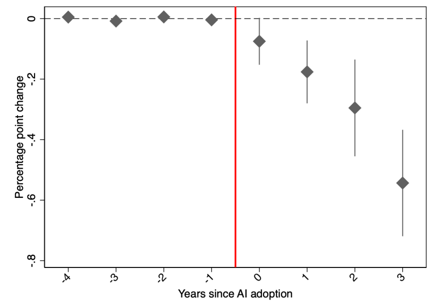
Notes: The figure shows estimates from a synthetic difference-in-differences model where the variable listed on the vertical axis is the dependent variable for each model. Inference uses the SDID bootstrap procedure, with resampling clustered at the hospital level and 50 bootstrap replications. The error bars represent 95% bootstrap confidence intervals.

Figure 9: Condition-specific *risk-adjusted* mortality event studies

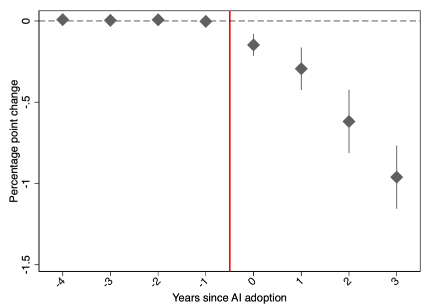
(a) Mortality rate - heart attack



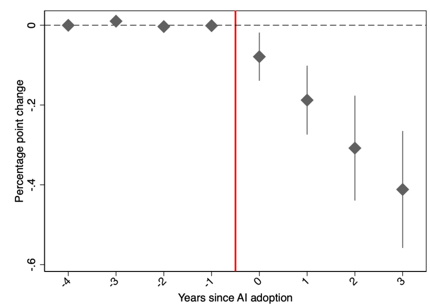
(b) Mortality rate - heart failure



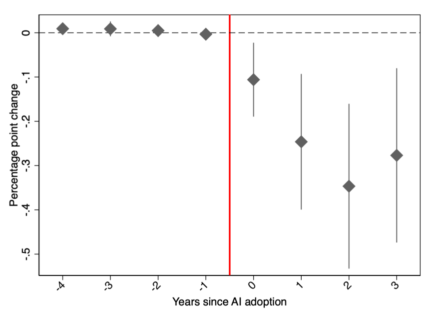
(c) Mortality rate - pneumonia



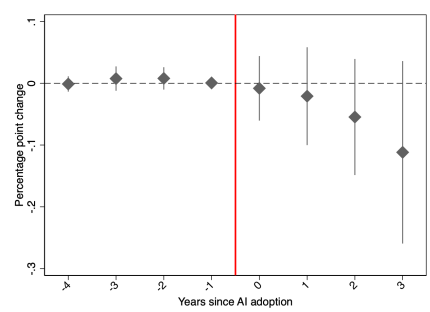
(d) Mortality rate - COPD



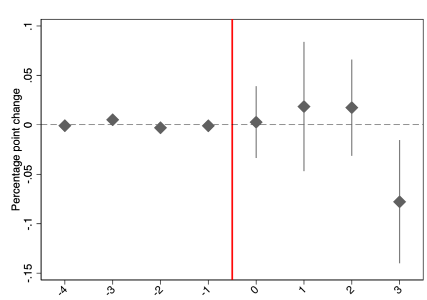
(e) Mortality rate - stroke



(f) Mortality rate - CABG surgery



(g) All-cause readmission rate

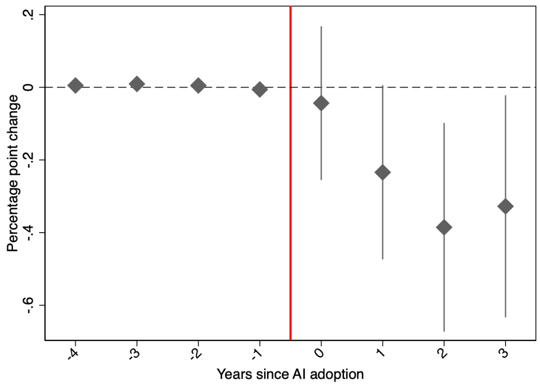


Sources: CMS Hospital Compare.

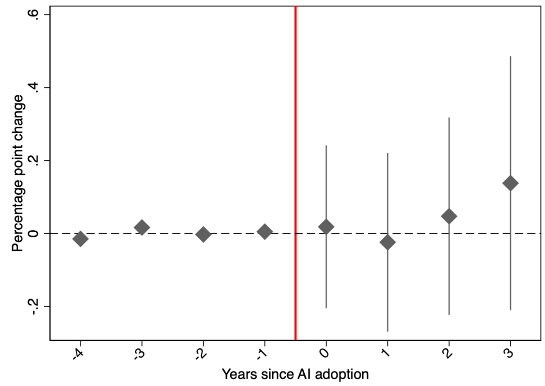
Notes: The figures show estimates from a synthetic difference-in-differences event study where each panel's title is the dependent variable. Inference uses the SDID bootstrap procedure, with resampling clustered at the hospital level and 50 bootstrap replications. The error bars represent 95% bootstrap confidence intervals.

Figure 10: Patient experience event studies

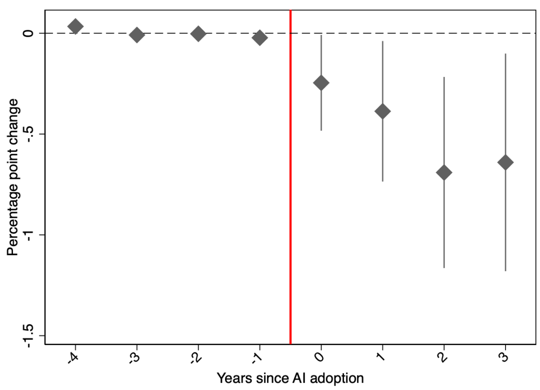
(a) Nurses communicated well



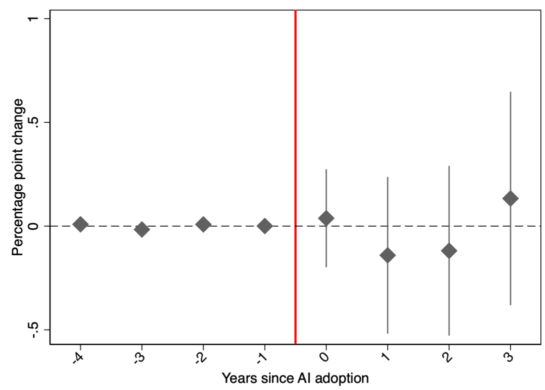
(b) Doctors communicated well



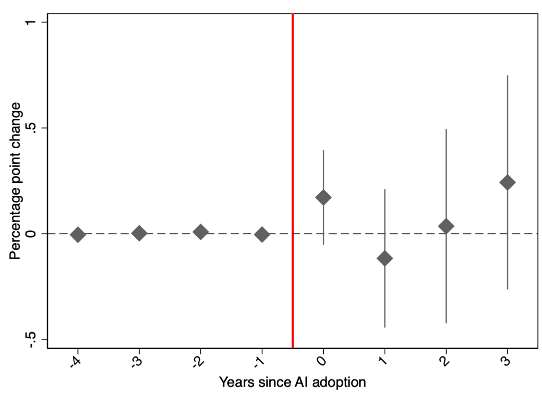
(c) Received help



(d) Hospital rating 9 or 10



(e) Recommend hospital



Sources: CMS Hospital Compare.

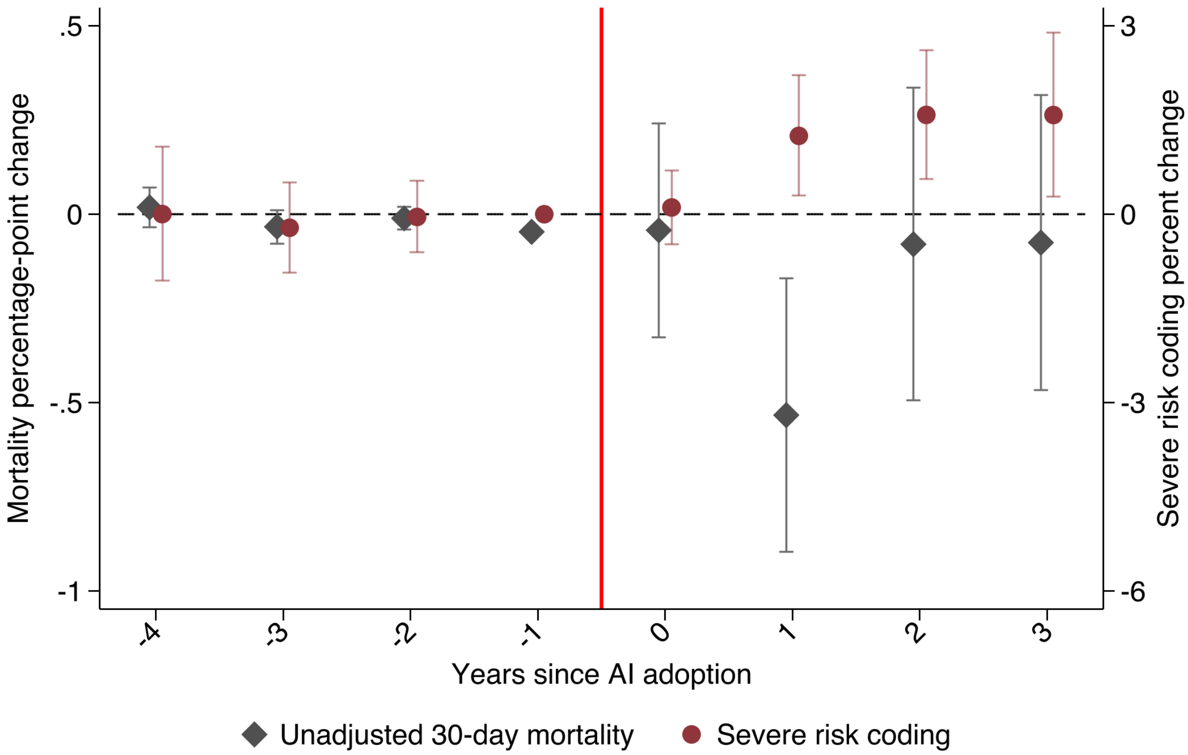
Notes: The figures show estimates from a synthetic difference-in-differences event study where each panel's title is the dependent variable. Inference uses the SDID bootstrap procedure, with resampling clustered at the hospital level and 50 bootstrap replications. The error bars represent 95% bootstrap confidence intervals.

Table 5: Direct Clinical Process Mechanism Evidence

Outcome	SDID estimate	p-value	Interpretation
Days to echo among HF admissions with echo	+0.104 days	0.575	No evidence that AI speeds echocardiography.
Days to head CT among stroke admissions with CT	+0.020 days	0.144	No evidence that AI speeds head CT.
Share of heart-failure stays with an echocardiogram	+0.0057	0.768	No evidence AI increases the share of HF stays with an echo.
Share of stroke stays with a head CT	−0.0018	0.778	No evidence AI increases the share of stroke stays with a head CT.
ICU days (COPD, PNA, HF, stroke) per admission	+0.0348 days	0.403	No consistent evidence of reduced escalation to ICU care.
Any ICU use (COPD, PNA, HF, stroke)	+0.00058	0.945	No consistent evidence of reduced escalation to ICU care.
Non-invasive respiratory stabilization	Essentially zero	0.890	No evidence of better non-invasive stabilization in the pooled respiratory panel.
Intubation / respiratory failure measure	+0.0014	0.100	Point estimate is opposite the hypothesized sign and not conventionally significant.

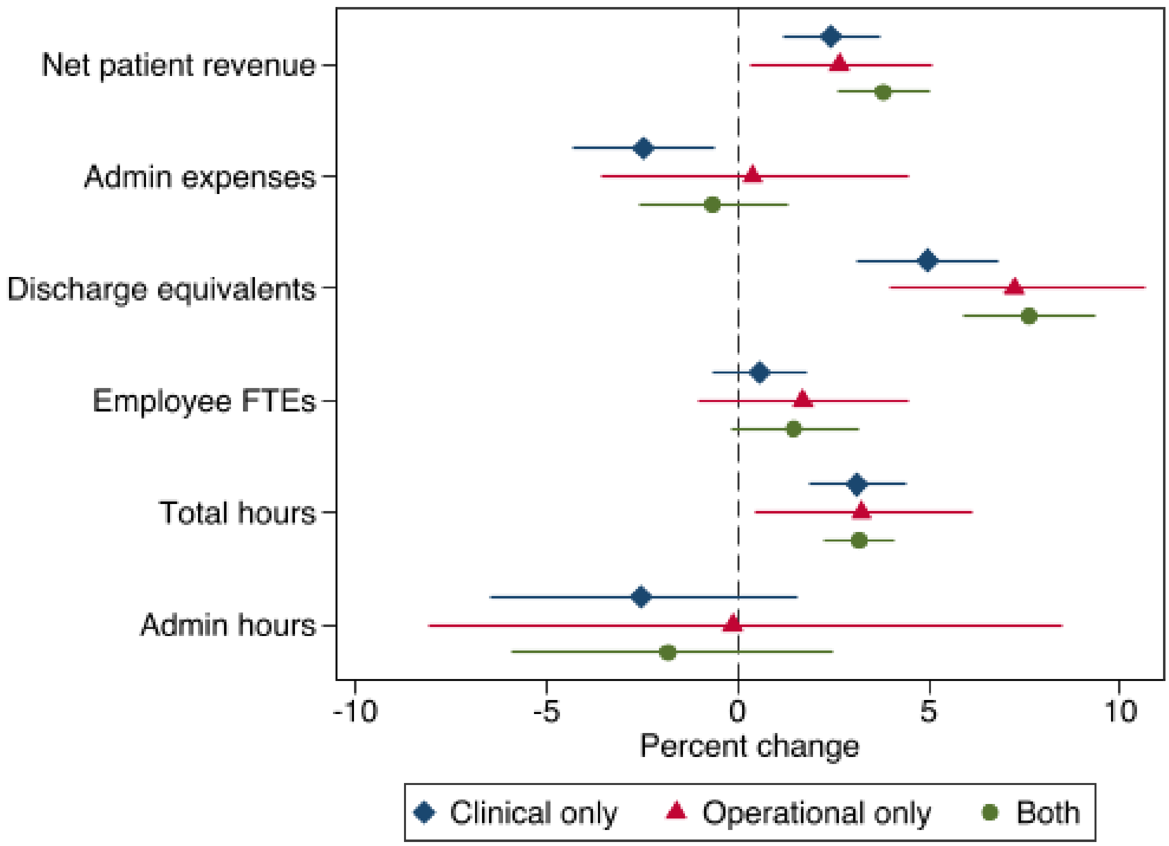
Notes: These tests are limited by the quality of claims coding for granular clinical procedures. Null findings here should therefore be read as absence of evidence in the current claims panel, not necessarily evidence that AI has no clinical effect. p-values are based on 50 nonparametric bootstrap replications with resampling at the hospital level.

Figure 11: Effects of AI adoption on unadjusted mortality and risk coding



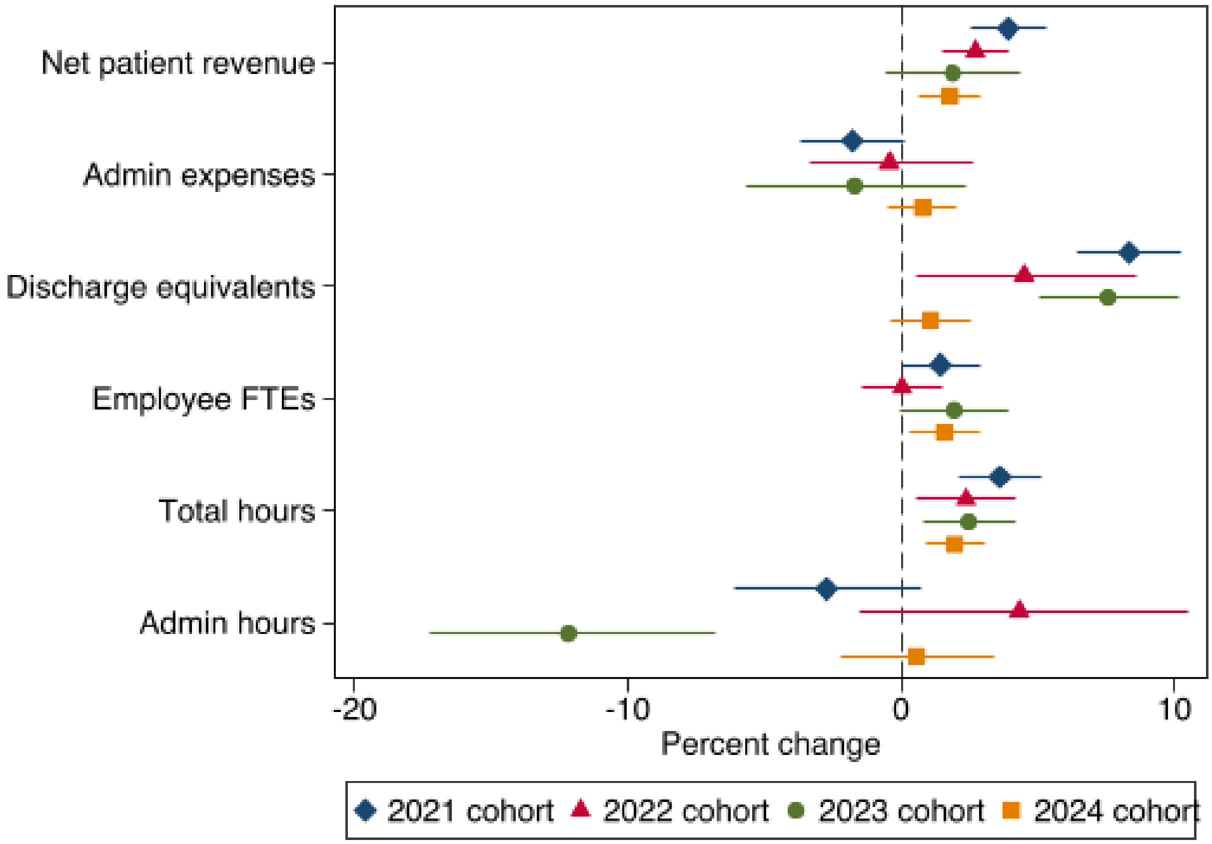
Notes: The figure plots SDID event-study estimates around hospital AI adoption. Diamonds show unadjusted 30-day mortality, measured as percentage-point changes and plotted on the left axis. Circles show severe risk coding, measured as percent changes in the average number of severe risk/comorbidity flags per admission and plotted on the right axis. Inference uses the SDID bootstrap procedure with resampling clustered at the hospital level and 50 bootstrap replications; error bars represent 95% bootstrap confidence intervals.

Figure 12: SDID Estimates by AI Adoption Type



Notes: Each point represents a separate SDID estimate of the effect of AI adoption on the indicated outcome for a single adoption type, with 95% confidence intervals based on 50 bootstrap replications with resampling clustered at the hospital level. Hospitals are classified as clinical only ($N = 566$), operational only ($N = 156$), or both ($N = 789$) based on whether any clinical or operational AI variable in the 2024 AHA survey is "expanding" or "fully integrated." Each adoption type is estimated against the same set of 2,582 never-treated controls. All outcomes are estimated in logs and converted to percent changes via $100 \times (\exp(\hat{\tau}) - 1)$.

Figure 13: Cohort-Specific SDID Estimates for Key Outcomes



Notes: Each point represents a separate SDID estimate of the effect of AI adoption on the indicated outcome for a single adoption cohort, with 95% confidence intervals based on 50 bootstrap replications with resampling clustered at the hospital level. The 2021 cohort (diamonds) has four post-treatment years, 2022 (triangles) has three, 2023 (circles) has two, and 2024 (squares) has one. All outcomes are estimated in logs and converted to percent changes via $100 \times (\exp(\hat{\tau}) - 1)$. Each cohort is estimated against the full set of never-treated controls.

Table 6: Sensitivity of the CARES Act Funding Placebo

Treatment definition	Sample / threshold	Revenue estimate
AI adoption (persistent-adopter baseline)	All hospitals	3.28%
Cumulative CARES dollars	Above median ($8.12 million)	−3.24%
CARES dollars / 2019 revenue	Above median (5.14%)	−1.63%
CARES dollars / 2019 revenue	Above median; never-AI only	−0.27%
CARES dollars / 2019 revenue	Top quartile (9.45%)	0.36%
CARES dollars / 2019 beds	Above median ($63,395)	−1.69%

Notes: This table reports SDID estimates for log net patient revenue, transformed to percentages. CARES treatment begins in 2020. Missing or negative annual funding values are set to zero before cumulative funding is calculated. Normalizing by pre-pandemic revenue or beds addresses the concern that an absolute-dollar threshold primarily captures hospital size. The placebo specifications generally do not reproduce the positive AI-revenue estimate, although they cannot rule out every form of pandemic-era confounding.

Appendix

Figure A1: AHA Annual Survey Artificial Intelligence Questions, FY2021-FY2024

(a) FY2021-FY2023

a. Does your hospital use artificial intelligence (AI) or machine learning in the following? (Check all that apply)

1. ☐ Predicting staffing needs
2. ☐ Predicting patient demand
3. ☐ Staff scheduling
4. ☐ Automating routine tasks
5. ☐ Optimizing administrative and clinical workflows
6. ☐ None of the above

(b) FY2024

a. How would you describe your hospital's current level of AI implementation in the following clinical areas?

	(1) Not Implementing	(2) Exploring	(3) Piloting/Testing	(4) Expanding	(5) Fully Integrated	(0) Don't Know
1. AI-assisted diagnostics (including imaging & early detection)	☐	☐	☐	☐	☐	☐
2. Predictive analytics for patient care (including outcomes & deterioration)	☐	☐	☐	☐	☐	☐
3. Clinical decision support tools	☐	☐	☐	☐	☐	☐
4. AI-assisted surgery	☐	☐	☐	☐	☐	☐
5. AI-powered patient communication and education	☐	☐	☐	☐	☐	☐
6. AI-driven population health management	☐	☐	☐	☐	☐	☐
7. Predictive models for resource allocation during emergencies	☐	☐	☐	☐	☐	☐
8. Other, please specify: ________	☐	☐	☐	☐	☐	☐

b. How would you describe your hospital's current level of AI implementation in the following operational areas?

	(1) Not Implementing	(2) Exploring	(3) Piloting/Testing	(4) Expanding	(5) Fully Integrated	(0) Don't Know
1. Revenue cycle management (e.g., billing, claims processing)	☐	☐	☐	☐	☐	☐
2. Supply chain optimization	☐	☐	☐	☐	☐	☐
3. Staff scheduling and workforce management	☐	☐	☐	☐	☐	☐
4. Patient flow and demand forecasting	☐	☐	☐	☐	☐	☐
5. Optimizing operational efficiency	☐	☐	☐	☐	☐	☐
6. Other, please specify: ________	☐	☐	☐	☐	☐	☐

Notes: The figure shows the artificial intelligence questions in the FY2021-FY2024 versions of the AHA Annual Survey.

Table A1: AI Implementation Status Among General Community Hospitals, 2024

	Not Implementing	Exploring	Piloting/ Testing	Expanding	Fully Integrated	Don't Know	N
Clinical AI Implementation							
AI-Assisted Diagnostics	0.16	0.23	0.16	0.21	0.16	0.08	2,361
Predictive Analytics for Patient Care	0.17	0.22	0.14	0.21	0.18	0.08	2,379
Clinical Decision Support Tools	0.15	0.24	0.10	0.23	0.20	0.08	2,299
AI-Assisted Surgery	0.45	0.23	0.08	0.08	0.03	0.13	2,348
Patient Communication & Education	0.21	0.32	0.20	0.14	0.04	0.09	2,355
Population Health Management	0.26	0.33	0.17	0.06	0.09	0.10	2,380
Resource Allocation During Emergencies	0.45	0.29	0.04	0.03	0.03	0.15	2,348
Other Clinical AI	0.36	0.05	0.06	0.08	0.06	0.38	1,259
Operational AI Implementation							
Revenue Cycle Management	0.19	0.26	0.18	0.23	0.06	0.08	2,356
Supply Chain Optimization	0.28	0.42	0.11	0.05	0.02	0.12	2,343
Staff Scheduling & Workforce Mgmt.	0.26	0.41	0.11	0.09	0.03	0.10	2,287
Patient Flow & Demand Forecasting	0.27	0.36	0.13	0.07	0.06	0.10	2,350
Optimizing Operational Efficiency	0.22	0.31	0.18	0.15	0.05	0.10	2,339
Other Operational AI	0.44	0.05	0.02	0.03	0.03	0.43	1,136

Notes: The table shows how hospitals answered each of 14 AI-related questions in the FY2024 Annual Survey. We classified hospitals as adopters in 2024 if they answered "expanding" or "fully integrated" to any of these questions.

Expanded AHA IT Supplement

To further understand how hospitals are using AI, we consulted the 2024 AHA IT Supplement. Among the 1,511 AI adopters in our main sample, 982 (65%) match to the 2024 IT supplement, so these data provide a 2024 cross-sectional snapshot of current AI use among adopters rather than a historical description of the exact tools used at each hospital's original adoption date. The AHA IT supplement goes beyond whether a hospital adopted AI and asks how hospitals use predictive models and large language models in practice. The picture that emerges is consistent with the adoption-type results above (Table A2). Among matched AI adopters, hospitals overwhelmingly report predictive-model use embedded in the EHR or EHR-launched apps, with especially common applications in inpatient risk prediction, outpatient follow-up, scheduling, and billing automation. Open-text responses likewise emphasize supply chain management, throughput and discharge management, ambient documentation, automated messaging, imaging, natural language processing (NLP), and data extraction. These self-reports line up much more closely with an operational and documentation story than with a narrow story in which AI mainly changes a small set of acute bedside procedures.

Table A2: Hospital-Reported AI Uses in the 2024 AHA IT Supplement

Measure	Share / rate	Interpretation
At least one predictive model in the EHR or an EHR-launched app	88.5%	Predictive analytics is the dominant current AI use case.
Both machine-learning and non-machine-learning predictive models	63.4%	Hospitals often use mixed AI/predictive tool portfolios rather than a single model class.
Inpatient risk / trajectory prediction among predictive-model users	96.8%	Consistent with risk stratification and clinical workflow support.
High-risk outpatient follow-up among predictive-model users	89.6%	Suggests broad workflow and care-management use.
Scheduling among predictive-model users	70.3%	Directly supports an operational throughput mechanism.
Billing automation among predictive-model users	61.6%	Directly supports a documentation / revenue-cycle mechanism.
EHR-integrated large language model currently in use	46.3%	Suggests documentation and messaging support are already relevant.
EHR-integrated large language model planned within one year	29.4%	Points toward continued growth in documentation-oriented AI.
Open-text "other" uses	Descriptive	Supply chain, throughput/discharge/capacity management, ambient documentation, message drafting, imaging, NLP, and data extraction recur frequently.

Notes: These figures come from the 2024 AHA IT supplement matched to the paper's AI-adopter hospitals. The supplement is a 2024 cross-section, so it describes current AI use among adopters rather than the exact tools each hospital used at its original adoption date.

Table A3: SDID Diagnostics for the Eight Primary Outcomes

Outcome	Normalized pre-RMSPE	Effective donors	Maximum donor weight	Maximum absolute change in ATT: Exclude same system	Uniform time weights	Drop top donor
Net patient revenue	0.011	2,467	0.0014	0.0026	0.0008	0.0007
Total expenses	0.006	2,477	0.0023	0.0078	0.0008	0.0015
Clinical expenses	0.005	2,463	0.0020	0.0013	0.0012	0.0002
Administrative expenses	0.017	2,387	0.0024	0.0059	0.0020	0.0011
Discharge equivalents	0.033	2,473	0.0018	0.0093	0.0017	0.0025
Employee FTEs	0.093	2,511	0.0017	0.0043	0.0007	0.0003
Total paid hours	0.108	1,470	0.0032	0.0067	0.0015	0.0015
Administrative paid hours	0.083	1,494	0.0036	0.0157	0.0035	0.0020

Notes: Pre-RMSPE is the median cohort-specific pre-period root mean squared prediction error normalized by the treated pre-period mean; lower values indicate better fit. Effective donors equal the inverse Herfindahl index of donor weights. The maximum donor weight and sensitivity columns report the most demanding cohort-specific value for each outcome. All cohort-outcome cells pass the treated-outcome donor-support check (zero support violations). ATT changes are in log points. Finance and volume outcomes have especially strong pre-period fit and diffuse weights. Labor-hour outcomes have weaker fit, and administrative hours are most sensitive to excluding same-system donors.

Table A4: Sensitivity to Alternative AI Treatment-Timing Definitions

Definition	Revenue	Total exp.	Clinical exp.	Admin. exp.	Discharge eq.	FTEs	Total hours	Admin. hours
Persistent-adopter baseline	3.28	2.94	2.85	−1.19	6.59	1.31	3.17	−1.92
Contemporaneous status	2.82	2.06	2.68	−1.39	7.05	1.06	3.19	−2.58
Exclude 2021 cohort	1.78	1.94	1.21	−0.25	4.70	0.65	2.18	−1.99
Strict observed transition	1.59	3.51	0.42	0.09	1.24	1.63	1.97	−1.35

Notes: Entries are percent changes calculated as $100 \times [\exp(\widehat{\tau}) - 1]$. The baseline uses 624, 165, 100, and 622 adopters in the 2021–2024 cohorts and 2,582 controls. The contemporaneous definition uses 848, 246, 141, and 622 adopters and 2,236 controls; it adds 346 earlier adopters without using 2024 status to screen prior reports. The exclude-2021 row removes the earliest, low-response cohort. The strict-transition row requires an observed nonuse response followed by an observed use response. These are point-estimate sensitivity comparisons; Table 4 reports stronger inference for the contemporaneous definition.

Table A5: Sensitivity to EHR-Vendor-by-Year Controls

	Persistent-adopter definition		Contemporaneous definition	
Outcome	Linked sample	Vendor × year	Linked sample	Vendor × year
Net patient revenue	2.86	2.44	2.11	1.59
Total expenses	2.30	1.88	1.08	0.44
Clinical expenses	2.85	2.32	2.37	1.82
Administrative expenses	−1.01	0.03	−1.43	−0.19
Discharge equivalents	5.46	3.48	5.33	3.55
Employee FTEs	0.47	0.34	−0.06	−0.44
Total paid hours	2.44	1.45	2.47	1.18
Administrative paid hours	−2.59	−2.71	−3.68	−3.93

Notes: Entries are percent changes in the EHR-linked sample. Vendor-by-year interactions allow annual shocks to differ by the hospital's reported primary inpatient EHR vendor. Vendor is first observed in 2022 in the available linked data, so this is not a purely pre-treatment covariate and may absorb both vendor-level confounding and AI implementation delivered through the EHR. The specification is therefore a demanding sensitivity analysis, not the preferred model.

Table A6: Sensitivity to Local COVID-19 Hospitalization Burden

Outcome	Same-sample unadjusted	County COVID adjustment	Leave-one-out county adjustment
Net patient revenue	2.83	2.73	3.09
Total expenses	2.13	2.06	2.26
Clinical expenses	2.78	2.70	2.97
Administrative expenses	−1.27	−1.26	−1.37
Discharge equivalents	7.13	7.01	7.60
Employee FTEs	1.10	1.18	1.16
Total paid hours	3.20	3.29	3.20
Administrative paid hours	−2.62	−2.07	−2.74

Notes: Entries are percent changes in a common sample. The county burden is the annual mean of weekly confirmed COVID-19 admissions per 100,000 residents, constructed from official HHS facility-week reports and entered as $\log(1 + \text{rate})$. The leave-one-out measure subtracts the focal hospital's admissions from its county total to reduce mechanical endogeneity. Facility-report coverage exceeds 99% in every study year. These are point-estimate sensitivity comparisons; they do not rule out unmeasured pandemic shocks.

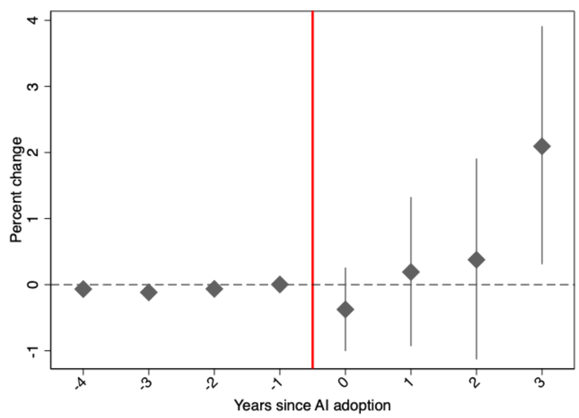
Table A7: Claims-Based Cohort and Mechanism Algorithms

Component	Population / source	Construction
Condition cohorts	Medicare FFS MedPAR acute inpatient stays, 2017–2024	Heart failure, pneumonia, COPD, and stroke cohorts are defined from the principal-diagnosis condition flags in the analytic MedPAR extract. A stay enters one condition cohort based on its principal diagnosis; secondary diagnoses do not determine cohort entry.
Diagnosis density	Same four condition cohorts	Count all diagnosis fields on each stay; secondary diagnoses equal total diagnoses minus the principal diagnosis. Aggregate admission-level counts to hospital-year-condition means, then pool the four condition panels for the summary models.
Present on admission	Same four condition cohorts	Count diagnoses marked present on admission (POA) and not marked POA using the claim-level POA indicators; calculate both counts and the POA share.
New chronic coding	Same four condition cohorts with prior claims history	Flag selected chronic-condition diagnoses on the index stay that are absent from the beneficiary's available prior-year claims; average the indicator at the hospital-year level.
Severe-risk index	Same four condition cohorts	Count serious comorbidity and acuity flags recorded on the inpatient claim and aggregate to the hospital-year-condition level. This is a claims-based severity index and is not the exact CMS mortality risk-adjustment model.
Unadjusted mortality	Same four condition cohorts	Indicator for death within 30 days of admission from Medicare enrollment/vital-status records, without risk adjustment; aggregate to hospital-year-condition rates and pool conditions for the mechanism event study.
Echocardiography	Heart-failure admissions	Identify echocardiography from inpatient and linked professional procedure records; construct any-echo and days from admission to the first observed echo among stays with an echo.
Head CT	Stroke admissions	Identify head CT from inpatient and linked professional procedure records; construct any-CT and days from admission to the first observed scan among stays with a CT.
ICU measures	Four condition cohorts	Use MedPAR intensive-care utilization fields to construct any ICU use and ICU days per admission.
Respiratory measures	COPD and pneumonia admissions	Construct indicators for non-invasive respiratory support and for intubation/respiratory failure from diagnosis and procedure fields, then aggregate to hospital-year rates.
Hospital-wide coding	All Medicare FFS acute inpatient admissions	Calculate diagnoses, submitted charges, and Medicare reimbursement per admission across all acute stays. Screen out hospitals with any nonpositive mean reimbursement or fewer than 10 admissions in any year; repeat with a 25-admission threshold.

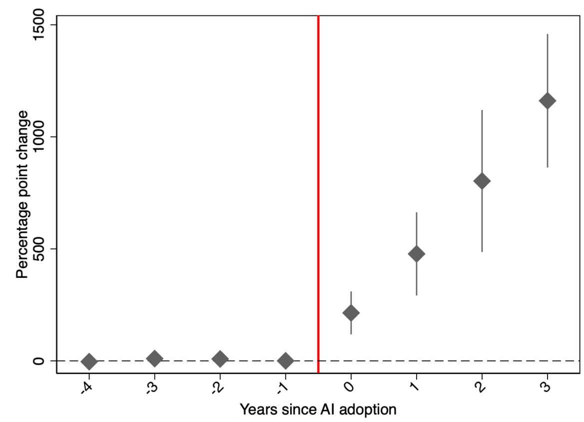
Notes: All hospital-year panels use consistent hospital identifiers and calendar years. The condition-specific pooled models weight hospital-year-condition cells according to the estimation sample rather than treating the hospital-wide Medicare rows as condition-specific. Exact diagnosis and procedure code lists are maintained in the claims-construction programs and should accompany the replication archive; the table documents the analytic algorithms used in this paper.

Figure A2: Additional volume analyses – beds and service type

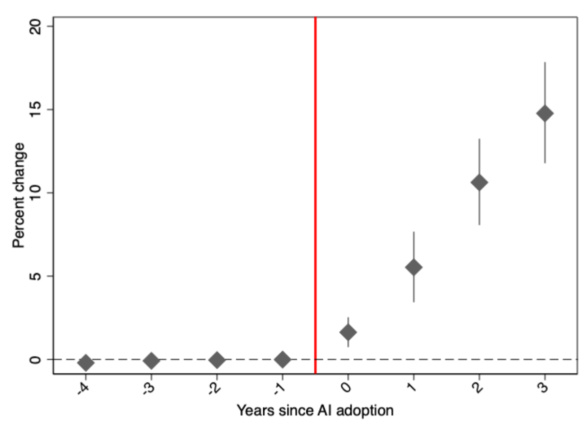
(a) Beds



(b) Inpatient discharges



(c) Outpatient discharge equivalents

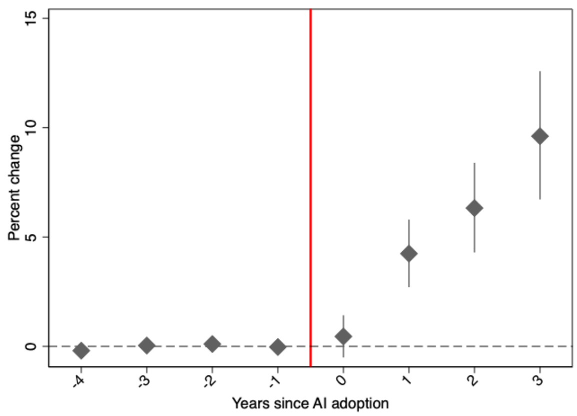


Sources: AHA's Annual Survey (AI adoption) and Hospital Cost Report data (expenses and discharge equivalents).

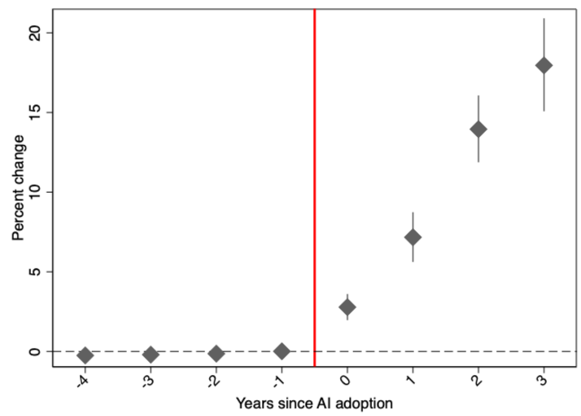
Notes: The figures show estimates from a synthetic difference-in-differences event study where the log of each panel's title is the dependent variable. Coefficient estimates have been transformed using the formula $100 \times (\exp(\text{coef}) - 1)$ to create the percentage effects shown in the figures. Inference uses the SDID bootstrap procedure, with resampling clustered at the hospital level and 50 bootstrap replications. The error bars represent 95% bootstrap confidence intervals.

Figure A3: Additional volume analyses – payer

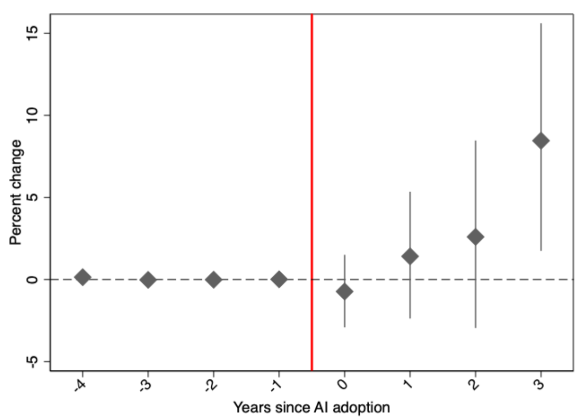
(a) Commercial discharges



(b) Medicare discharges



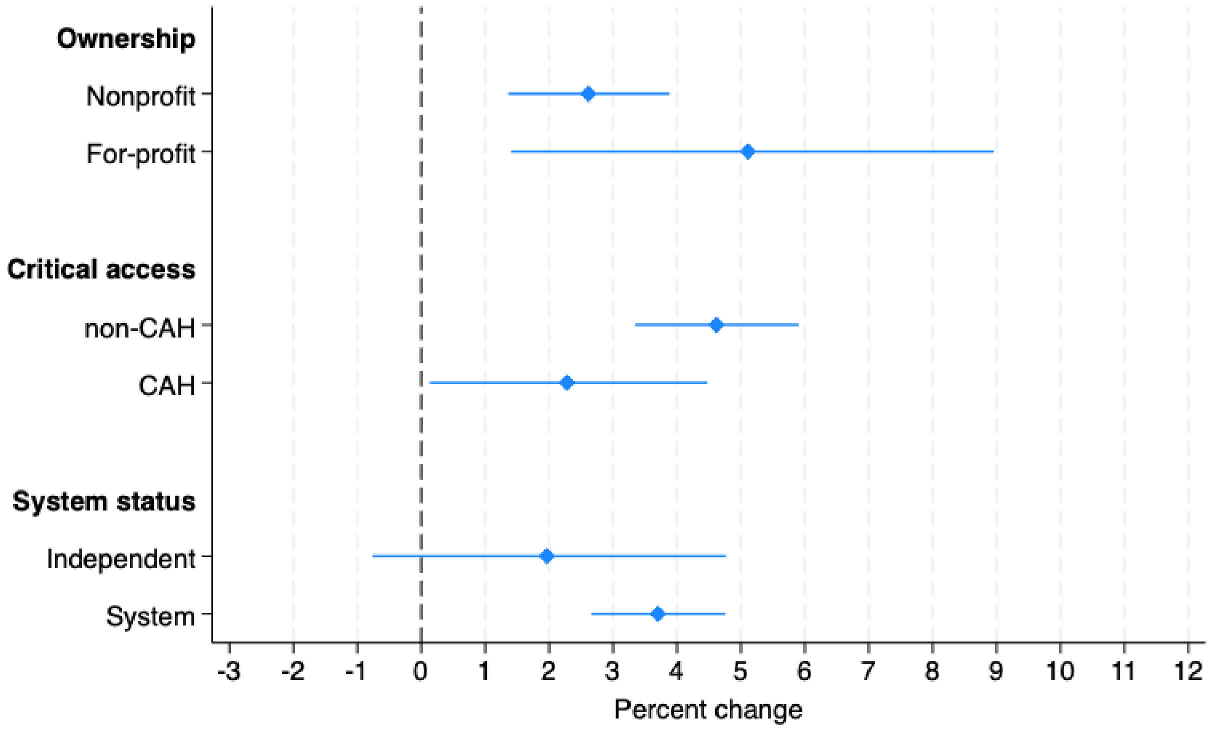
(c) Medicaid discharges



Sources: AHA's Annual Survey (AI adoption) and Hospital Cost Report data (expenses and discharge equivalents).

Notes: The figures show estimates from a synthetic difference-in-differences event study where the log of each panel's title is the dependent variable. Coefficient estimates have been transformed using the formula $100 \times (\exp(\text{coef}) - 1)$ to create the percentage effects shown in the figures. Inference uses the SDID bootstrap procedure, with resampling clustered at the hospital level and 50 bootstrap replications. The error bars represent 95% bootstrap confidence intervals.

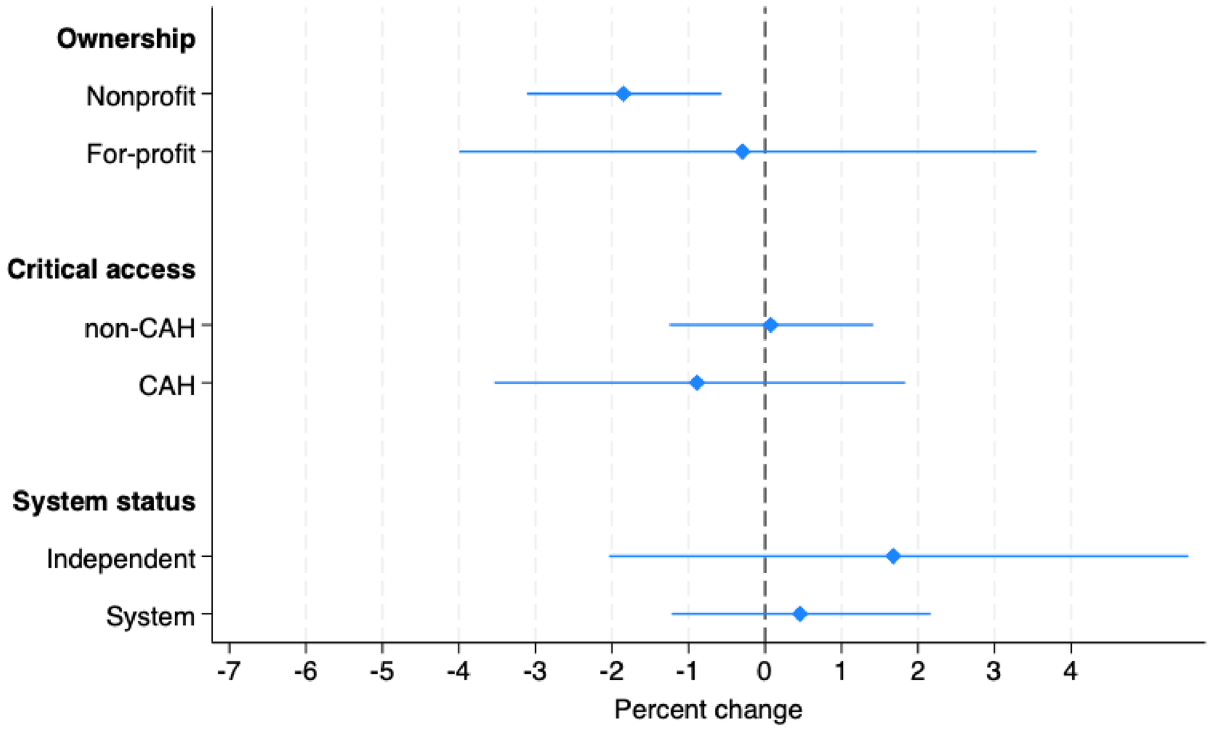
Figure A4: Net patient revenue heterogeneity



Sources: AHA's Annual Survey (AI adoption) and Hospital Cost Report data as collected by RAND (expenses and discharge equivalents).

Notes: The figure shows how the ATT for net patient revenue varies across different subgroups.

Figure A5: Administrative expense heterogeneity

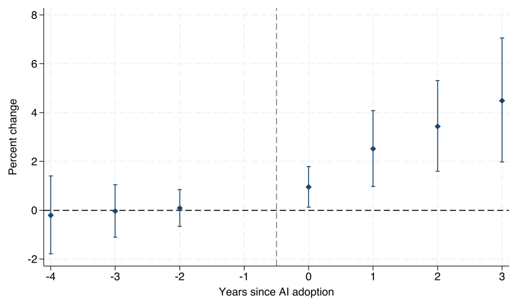


Sources: AHA's Annual Survey (AI adoption) and Hospital Cost Report data (expenses and discharge equivalents).

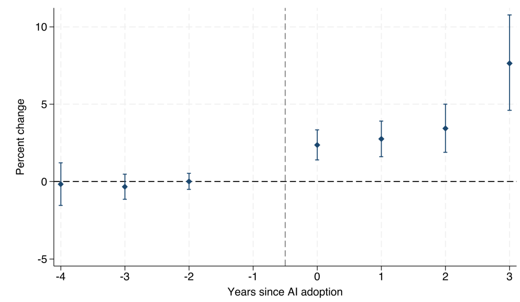
Notes: The figure shows how the ATT for administrative expenses varies across different subgroups.

Figure A6: Finances and volume event studies - propensity score matching

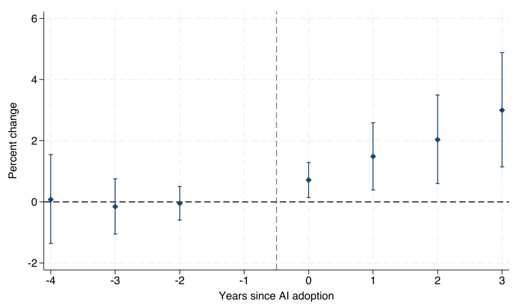
(a) Net patient revenue



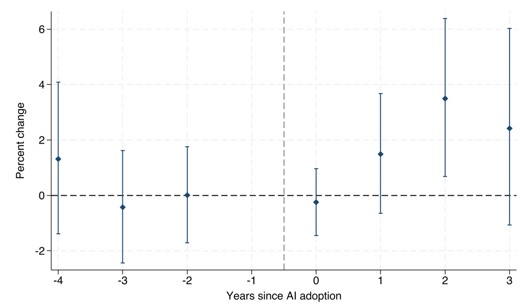
(b) Total expenses



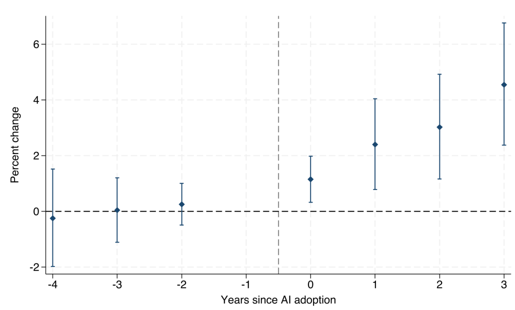
(c) Expenses for clinical activities



(d) Administrative expenses



(e) Discharge equivalents

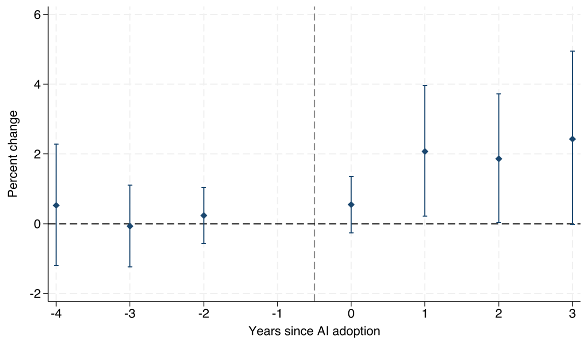


Sources: AHA's Annual Survey (AI adoption) and Hospital Cost Report data (expenses and discharge equivalents).

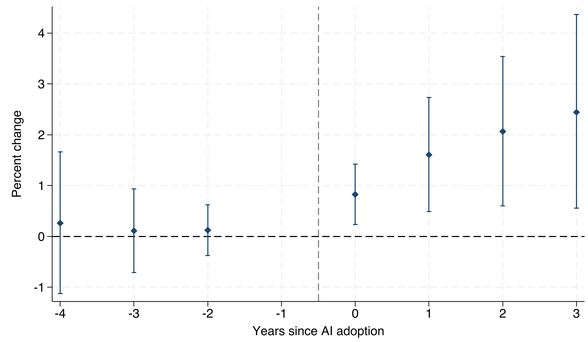
Notes: The figures show estimates from propensity-score-matched difference-in-differences event studies where the log of each panel's title is the dependent variable. Coefficient estimates have been transformed using the formula $100 \times (\exp(\text{coef}) - 1)$ to create the percentage effects shown in the figures. Models included matched cohort-by-hospital and matched cohort-by-year fixed effects, and standard errors were clustered at the hospital level. The error bars represent 95% confidence intervals.

Figure A7: Labor event studies – propensity score matching

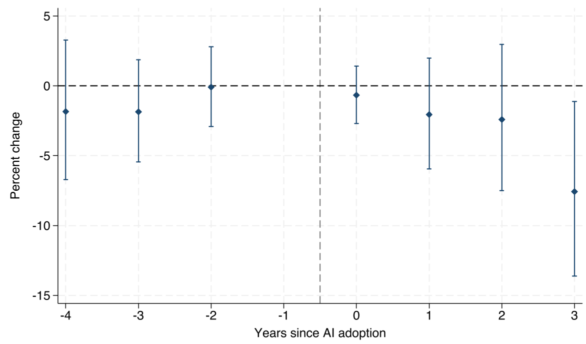
(a) Employees on payroll (FTEs)



(b) Total paid hours



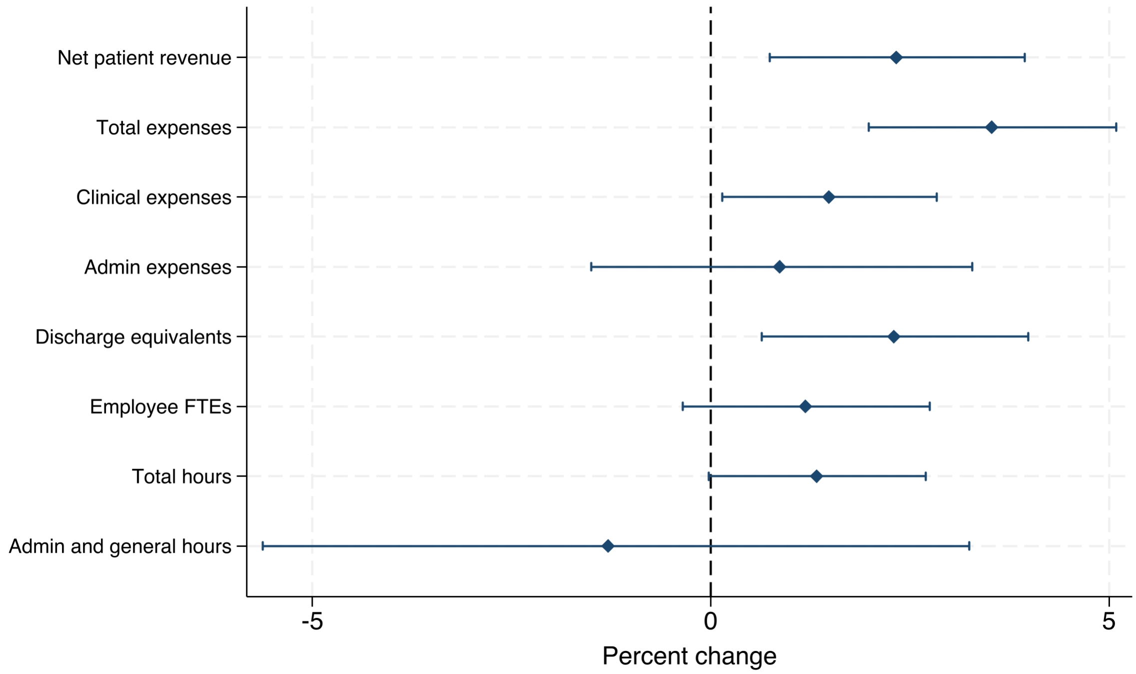
(c) Paid hours, administrative and general



Sources: AHA's Annual Survey (AI adoption) and Hospital Cost Report data.

Notes: The figures show estimates from propensity-score-matched difference-in-differences event studies where the log of each panel's title is the dependent variable. Coefficient estimates have been transformed using the formula $100 \times (\exp(\text{coef}) - 1)$ to create the percentage effects shown in the figures. Models included matched cohort-by-hospital and matched cohort-by-year fixed effects, and standard errors were clustered at the hospital level. The error bars represent 95% confidence intervals.

Figure A8: Summary of AI adoption effects on finances, volume, and employment – propensity score matching



Notes: The figure shows estimates from propensity-score-matched difference-in-differences models, where the dependent variable in each model is the logged version of the variable listed on the vertical axis. Coefficient estimates have been transformed using the formula $(100 \times [\exp(\text{coef})\text{ - }1])$ to express the effects as percentage changes. Models include matched cohort-by-hospital and matched cohort-by-year fixed effects, and standard errors are clustered at the hospital level. Error bars represent 95% confidence intervals.