The Impacts of AI Integration on Efficiency and Employment Patterns in American Hospitals

Daniel R. Arnold, PhD;[1] Jonathan Cantor, PhD;[2] Mallory Go, AB;[1] Arjun Teotia, PhD;[3,4] Christopher M. Whaley, PhD[1]

[1] Brown University
[2] RAND
[3] Henry Ford Health
[4] Michigan State University

Manuscript word count: 3,169
Abstract word count: 165

Corresponding Author
Christopher M. Whaley, Ph.D.
Associate Professor
Brown University School of Public Health
Providence, RI 02903
Phone: 406-431-1239
Email: christopher_whaley@brown.edu

Supported by NIA R01AG073286

Abstract

Since the launch of OpenAI's ChatGPT in November 2022, companies across the global economy have rapidly adopted generative artificial intelligence (AI) tools. AI technologies are believed to improve business efficiencies and operations. Because efficiencies are expected to come from human-performed tasks they may reduce employment. Both aspects of the impacts of AI have little existing empirical evidence. Studying these impacts in the health care sector is especially important, since administrative tasks there appear particularly prone to disruption by AI. However, it remains unclear whether reductions in labor also affect the delivery of care. Combining 2021-2023 data on AI adoption by all U.S. hospitals with measures of hospital performance, we find AI adoption led to a 2.5% increase in net patient revenue, a 1.6% decrease in administrative expenses, and a 6.7% increase in patient volume. We further find that the observed improvements in clinical quality, taken in aggregate, imply that AI adoption in hospitals reduced mortality by roughly 10,000 deaths nationwide.

The use of generative artificial intelligence (GenAI) has rapidly accelerated across the global economy (*1*, *2*). This acceleration was spurred by the release of the Generative Pre-trained Transformer (ChatGPT) by OpenAI in November 2022 (*3*). Since 2021, an estimated $1.1 trillion has been invested in AI technologies (*4*). Across industries, the potential impacts of AI are multifaceted and uncertain (*5*, *6*). AI technologies can improve business efficiencies and operations (*7*). At the same time, these efficiencies often come from human-performed tasks and can thereby reduce human employment (*8–10*). Yet, the impacts of GenAI technologies on both firm performance and labor markets are not well understood.

A particularly important industry to study these impacts is the health care sector (*11–14*). Health care features many tasks where AI technologies are expected to improve performance (e.g., using AI to interpret radiology images). The health care industry also has high levels of administrative tasks (e.g., medical billing and clinical note taking) that are currently manually performed and can be automated or improved using new technologies. One recent report suggested that hospitals spend approximately 56% of their total operating revenue on labor (*15*). As a result, there is considerable potential that GenAI can improve diagnostics and reduce administrative burden and waste in the U.S. health care system (*16*, *17*). In a recent survey by the American Medical Association (AMA), more than half of physicians (57%) said reducing administrative burdens through automation was the biggest area of opportunity for AI (*18*). Other areas of opportunity include improving patient care outcomes by increasing the speed and accuracy of diagnoses (*19*, *20*) and patient engagement and adherence (*21*, *22*). Yet, previous research has primarily been focused on rates of self-reported AI adoption within the healthcare sector (*14*, *23–25*). Despite the potential of GenAI, there exists no quantitative evidence on the associations between adoption of AI technologies and measures of system performance.

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 (*23–25*), 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 (*23*, *26*). One recent study quantified trends in the quality of care delivered by hospitals that adopted AI and found mixed results (*27*).

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 have impacts on patient outcomes or quality of care, then disparities in health outcomes are also likely to be exacerbated (*23*, *28*, *29*).

As AI technologies become prevalent, understanding the impacts of AI adoption on both firm (i.e., hospital) and consumer (i.e., patients) outcomes is critical. AI can do some human jobs, but it is not clear if it complements or substitutes human labor. It is also not clear if new "labor" from AI allows firms to expand capacity or increase throughput. Finally, it is not clear if AI efficiencies are to the firm, or if they impact patient care. To inform policymakers, health care administrators, and health care providers, this paper uses a novel synthetic difference-in-differences quasi-experimental approach to examine the impacts of hospital adoption of AI tools on hospital performance.

Data

We used 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 related to AI in 2021. We used 2021 to 2023 AHA Survey data to track AI adoption by hospitals. The 2023 Annual Survey included hospital characteristics for all 6,093 hospitals in the U.S., of which 5,112 were community hospitals (i.e., nonfederal, short-term general, and other special hospitals) (*30*). For this study, we only considered survey responses from the 4,235 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 question we used to classify a hospital as an AI adopter was "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." Among the short-term general hospitals in the U.S. in 2023, 2,256 (53%) responded to this question. The response rate was similar in 2022 (56%) and lower in 2021 (20%). Among all 4,340 short-term general hospitals that appear in the data between 2021 and 2023, 2,771 (64%) answered the question in at least one year of the survey. We considered a hospital as an AI adopter if it reported it used AI at all (i.e., checked any box 1-5 in the question). Hospitals that never answered the question or answered the question by checking box 6 (none of the above) were considered to be non-adopters.

Healthcare Cost Report Information System (HCRIS) data were the source of the hospital performance measures we analyzed (*31*). The HCRIS data is a comprehensive collection of

financial and statistical information that Medicare-certified institutional health care providers are required to submit annually to the Centers for Medicare & Medicaid Services (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 (*32*, *33*).

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 (*34*, *35*), as are the patient communication and hospital recommendation measures (*36*).

Empirical Methods

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 friendly payer environments. As a solution, we used a Synthetic Difference-in-Differences (SDID) method to estimate the association between AI adoption and

changes in our outcome measures (*37*). SDID calculates a weighted average of various control hospitals to construct a "synthetic" control whose pre-adoption outcome trend is parallel to the trend of the treated hospitals. SDID also applies time weights, giving more importance to pre-adoption years that are better predictors of the post-adoption period, which improves precision and accounts for trends over time.

Once this optimized synthetic control group is built, the SDID method performs a difference-in-differences calculation. It compares the change in the outcome (e.g., net patient revenue) before vs. after AI adoption in the treated hospitals to the corresponding change in the custom-built synthetic control group. This combination of weighting both control units and time periods provides a more robust and credible estimate than traditional methods. When hospitals adopt AI at different times—a "staggered adoption" scenario—the method simply applies to each group of adopting hospitals separately, and the results are then averaged to produce an overall treatment effect. SDID is more appropriate in our setting than the traditional difference-in-differences approach because the group of all possible control hospitals is often on a different pre-adoption trend for several outcomes, which is likely due to the fact that AI adopting hospitals are different than non-adopting hospitals across a number of important dimensions (*23–26*).

We conducted several additional analyses to analyze where AI adoption is having the largest impact. Specifically, we tested whether the impact of AI adoption differed by hospital profit status (for-profit vs. nonprofit), system membership, and critical access status. To be deemed a critical access hospital, a hospital has to have less than 25 beds and be geographically separate from other hospitals. Critical access hospitals have unique financing models and face financial pressures that have contributed to closures or consolidation with health systems (*38*).

We hypothesize that to date for-profit hospitals, hospitals that are part of systems (as opposed to independent hospitals), and non-critical access hospitals have benefited more from AI adoption. We also analyzed how AI adoption impacts patient volume by service type (i.e., inpatient or outpatient) and payer (i.e., commercial, Medicare, or Medicaid). Details on our data and analytic strategy are described further in the appendix.

The Effects of Hospital AI Adoption

We identified 1,221 hospitals that adopted AI over the 2021-2023 period. The majority of AI adopters first reported using AI in 2021 (837 hospitals), while 245 and 139 hospitals adopted in 2022 and 2023, respectively. **Table S1** in the appendix compares characteristics of the AI adopter and non-adopter hospitals at baseline in 2017. AI adopter hospitals were larger (236 beds vs. 117 beds; $P < 0.01$), more likely to be part of a system (0.91 vs. 0.62; $P < 0.01$), less likely to be critical access hospitals (0.16 vs. 0.39; $P < 0.01$), more often located in urban areas (urban continuum code 2.6 vs. 4.1; $P < 0.01$), less likely to be for-profit (0.06 vs. 0.21; $P < 0.01$), more often a major teaching hospital (0.15 vs. 0.05; $P < 0.01$), and had higher casemix indices (1.69 vs. 1.57; $P < 0.01$) than non-adopter hospitals.

In terms of hospital finances and volume, AI adoption was associated with a 2.5% (95% confidence interval [CI], 1.2% to 3.8%) increase in net patient revenue, a 2.4% (95% CI, 1.3% to 3.6%) increase in total expenses, a 2.3% (95% CI, 1.4% to 3.3%) increase in clinical expenses, a 1.6% (95% CI, -2.8% to -0.3%) decrease in administrative expenses, and a 6.7% (95% CI, 4.3% to 9.1%) increase in discharge equivalents (**Fig. 1**). We used log values for the dependent variables and then transformed to percent changes for each of these five results.

The event study plots for net patient revenue and administrative expenses are shown as **Fig. 2** and **Fig. 3**, respectively. **Fig. 2** shows that AI adoption immediately increased net patient revenue (0.8%; 95% CI, -0.1% to 1.7%) and the effect continued to show signs of growing three years post-adoption when the magnitude reached 4.1% (95% CI, 1.9% to 6.4%). 4.1% of the adopting hospital baseline mean net patient revenue ($366 million) equates to $15 million in additional revenue. **Fig. 3** shows that AI adoption also immediately reduced administrative expenses (-1.0%; 95% CI, -2.1% to 0.1%) and the effect also showed signs of growing three years post-adoption when the magnitude reached -2.1% (95% CI, -4.5% to 0.3%). 2.1% of the adopting hospital baseline mean administrative expenses ($65 million) is a $1.4 million reduction in administrative expenses. Event study plots for the other finance and volume outcomes are shown in **Fig. S3** in the appendix.

For labor outcomes, AI adoption was associated with no change in employee FTEs (1.0; 95% CI, -0.3 to 2.2), a 2.7% increase in total paid hours (95% CI, 1.0 to 4.5), and a not statistically significant 2.8% (95% CI, -5.9 to 0.4) decrease in administrative paid hours (**Fig. 1**). Event study plots for the labor outcomes are shown in **Fig. S4** in the appendix.

Association Between Hospital AI Adoption and Care Quality

AI adoption was associated with a 0.14 percentage point (pp) (95% CI, -0.25 to -0.04) decrease in the heart failure mortality rate (**Fig. 4**). The mean heart failure mortality rate at baseline for AI adopters was 11.7% so this corresponds to a 1.2% relative decrease. AI adoption was also associated with a 0.27 pp (95% CI, -0.43 to -0.11) decrease in the pneumonia mortality rate and a 0.14 pp (95% CI, -0.25 to -0.04) decrease in the COPD mortality rate, which both correspond to 1.8% relative decreases. The other four quality outcome measures we analyzed were not

associated with AI adoption. The five quality patient experience measures we analyzed likewise were not associated with AI adoption with "receiving help" being the lone exception. AI adoption was associated with a 0.51 pp (95% CI, -0.89 to -0.13) decrease in the percentage of patients who said they received help when needed. At baseline, 66.3% of patients at AI adopting hospitals reported receiving help when needed so this effect corresponds to a less than 1% relative decrease. Event study plots for quality outcomes and patient experience associations are shown in **Figs. S5 and S6** in the appendix, respectively.

Table 1 places the previously discussed results in an aggregate context. The 1,221 adopting hospitals were treated for 3.6 years on average of the possible four years of treatment (2021-2024). We multiplied 1,221 by 3.6 and the appropriate means and coefficients estimates from Table S1 and Figure 3 to produce the aggregate results in Table 1. For instance, for the aggregate net patient revenue effect we multiplied 1,221 by 3.6 and $366 million (mean net patient revenue at baseline for AI adopters in Table S1) and 0.025 (the percent change in net patient revenue from Figure 3). Using this procedure our estimated aggregate effects are $40.1 billion (95% CI, $19.4 billion to $61.0 billion) in net patient revenue, $40.0 billion (95% CI, $20.5 billion to $59.6 billion) in total expenses, -$4.4 billion (95% CI, -$7.9 billion to -$0.9 billion) in administrative expenses, 6.0 million (95% CI, 3.8 million to 8.1 million) discharge equivalents, and 65,000 (95% CI, 24,000 to 106,000) additional employees. Additionally, we estimated nearly 10,000 fewer in-hospital deaths (95% CI, 3,253 to 14,706) across the three mortality measures that improved following AI adoption.

Additional analyses

We found AI adoption was associated with larger increases in revenues for 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 found little heterogeneity across the same three dimensions in terms of effects on administrative expenses (**Figs. S1 and S2** in the appendix).

We also analyzed whether the volume increases we identified were contained to a particular service type (e.g., inpatient) or payer (e.g., commercial) or applied more broadly. **Figs. S7 and S8** in the appendix shows volume increased across the board – by type (inpatient, outpatient) and payer (commercial, Medicare, and Medicaid). Notably, 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%; 95% CI, 0.3% to 3.7%) than the increase in patient volume.

Discussion

The launch of ChatGPT and other AI applications has created potentials to transform the health care industry by improving patient interaction, supporting clinical diagnoses, enhancing health education, providing health advice, promoting wellness, and increasing efficiencies (*39*). This study shows that the adoption of AI by hospitals is associated with increases in revenue, clinical expenses, and volume and decreases in administrative costs. Additionally, we show evidence that hospital AI adoption is associated with 10,000 fewer in-hospital deaths. Finally, this paper finds that AI adoption increases patient volume across service and payer type and that this additional volume is largely a result of improved efficiency rather than physical expansion (e.g., more hospital beds).

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 a major teaching hospital, and treat more complex patient populations, which is consistent with previous studies (*23–25*, *40*). 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 (*28*, *41*). These differences in AI adoption beg 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 these benefits.

This study's findings should be interpreted while cognizant of its several limitations. First, a number of the hospitals we labeled as non-adopters could actually be adopters who simply chose not to answer the AHA's survey question on the topic. In this case they could be receiving weight in our synthetic control and would in turn bias our results toward the null. Second, the most recently available data on AI adoption is as of 2023 and we are measuring outcomes through 2024. As a result, some hospitals may have adopted in 2024 and we again considered them non-adopters making our results again more conservative. Finally, the SDID method used in this study hinges on the assumption that the constructed synthetic control is a valid counterfactual, that is, it would have continued on the same trend as the treated group absent the treatment.

Initial findings suggest that the integration of AI into hospital administrative workflows holds substantial promise for mitigating systemic operational burdens. Emerging AI applications in areas such as predictive scheduling algorithms, Natural Language Processing for automated clinical documentation and billing, and streamlined revenue cycle management are showing a

measurable impact on efficiency. A direct corollary of these gains is the potential for increased patient throughput, which logically translates to enhanced revenue streams. However, while these preliminary results are encouraging, they necessitate a deeper, more systematic inquiry. Future research must be directed toward elucidating the specific causal mechanisms by which AI interventions generate these improvements. A critical challenge moving forward will be developing scalable and equitable implementation strategies to ensure these advancements are not confined to a few large medical centers but can be disseminated throughout the diverse landscape of the American health care system, thereby maximizing their public health impact.

References

1. E. Brynjolfsson, D. Li, L. Raymond, Generative AI at Work. *Q. J. Econ.* **140**, 889–942 (2025).

2. McKinsey & Company, The state of AI in 2025: Agents, innovation, and transformation (2025). https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai.

3. A. Jackson, ChatGPT one year on: adoption, transformation, regulation (2023). https://aimagazine.com/machine-learning/chatgpt-one-year-on-adoption-transformation-regulation.

4. Stanford University Human-Centered Artificial Intelligence, "Artificial Intelligence Index Report 2025" (2025).

5. E. Brynjolfsson, B. Chandar, R. Chen, "Canaries in the Coal Mine? Six Facts about the Recent Employment Effects of Artificial Intelligence" (2025).

6. M. Buenaventura, A. Shenk, A. Nergui, S. Mann, F. C. Dolan, K. V. Patel, J. Mignano, J. W. Welburn, "Artificial Intelligence Adoption and Sectoral Transformation: Implications for Health Care, Financial Services, Climate and Energy, and Transportation" (2025); https://www.rand.org/pubs/research_reports/RRA3888-1.html.

7. American Hospital Association, 4 Actions to Close Hospitals' Predictive AI Gap (2025). https://www.aha.org/aha-center-health-innovation-market-scan/2025-11-04-4-actions-close-hospitals-predictive-ai-gap.

8. M. R. Frank, D. Autor, J. E. Bessen, E. Brynjolfsson, M. Cebrian, D. J. Deming, M. Feldman, M. Groh, J. Lobo, E. Moro, D. Wang, H. Youn, I. Rahwan, Toward understanding the impact of artificial intelligence on labor. *Proc. Natl. Acad. Sci.* **116**, 6531–6539 (2019).

9. D. Autor, Applying AI to Rebuild Middle Class Jobs. National Bureau of Economic Research 32140 [Preprint] (2024). https://doi.org/10.3386/w32140.

10. D. Acemoglu, D. Autor, S. Johnson, "Can we have Pro-Worker AI?" (Center for Economic Policy Research).

11. R. M. Wachter, E. Brynjolfsson, Will Generative Artificial Intelligence Deliver on Its Promise in Health Care? *JAMA* **331**, 65–69 (2024).

12. I. Y. Chen, A. Zink, AI Tools in Human Hands: Measuring Real-World Impact in the Emergency Department. *NEJM AI* **2**, AIe2500083 (2025).

13. A. Zink, S. Morriss, A. Gangopadhyaya, Z. Obermeyer, "Building Equitable Artificial Intelligence in Health Care" (Urban Institute, 2023).

14. T. D. Nguyen, C. M. Whaley, K. Simon, N. Mehta, H. Yu, R. K. McBain, A. Mehrotra, J. H. Cantor, Adoption of Artificial Intelligence in the Health Care Sector. *JAMA Health Forum* **6**, e255029 (2025).

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

16. M. D. Howell, G. S. Corrado, K. B. DeSalvo, Three Epochs of Artificial Intelligence in Health Care. *JAMA* **331**, 242–244 (2024).

17. A. Shuaib, Transforming Healthcare with AI: Promises, Pitfalls, and Pathways Forward. *Int. J. Gen. Med.* **17**, 1765–1771 (2024).

18. American Medical Association, Physician enthusiasm grows for health care AI (2025). https://www.ama-assn.org/press-center/ama-press-releases/ama-physician-enthusiasm-grows-health-care-ai.

19. Y. Jiang, A. V. Edwards, G. M. Newstead, Artificial Intelligence Applied to Breast MRI for Improved Diagnosis. *Radiology* **298**, 38–46 (2021).

20. M. Khalifa, M. Albadawy, AI in diagnostic imaging: Revolutionising accuracy and efficiency. *Comput. Methods Programs Biomed. Update* **5**, 100146 (2024).

21. A. Babel, R. Taneja, F. Mondello Malvestiti, A. Monaco, S. Donde, Artificial Intelligence Solutions to Increase Medication Adherence in Patients With Non-communicable Diseases. *Front. Digit. Health* **3** (2021).

22. J. N. Sekandi, W. Shi, R. Zhu, P. Kaggwa, E. Mwebaze, S. Li, Application of Artificial Intelligence to the Monitoring of Medication Adherence for Tuberculosis Treatment in Africa: Algorithm Development and Validation. *JMIR AI* **2**, e40167 (2023).

23. P. Nong, J. Adler-Milstein, N. C. Apathy, A. J. Holmgren, J. Everson, Current Use And Evaluation Of Artificial Intelligence And Predictive Models In US Hospitals. *Health Aff. (Millwood)* **44**, 90–98 (2025).

24. R. Bin Abdul Baten, How are US hospitals adopting artificial intelligence? Early evidence from 2022. *Health Aff. Sch.* **2**, qxae123 (2024).

25. A. Goldfarb, X. (Henry) He, F. Teodoridis, Patterns of Artificial Intelligence Adoption by Hospitals. *AEA Pap. Proc.* **115**, 40–45 (2025).

26. J. Chen, A. S. Yan, Hospital Artificial Intelligence/Machine Learning Adoption by Neighborhood Deprivation. *Med. Care* **63**, 227 (2025).

27. Y.-M. Hwang, M. Y. Ng, M. Pillai, M. P. Sahai, T. Hernandez-Boussard, AI Implementation in U.S. Hospitals: Regional Disparities and System-Level Implications. medRxiv [Preprint] (2025). https://doi.org/10.1101/2025.06.27.25330441.

28. P. Nong, E. Maurer, R. Dwivedi, The urgency of centering safety-net organizations in AI governance. *Npj Digit. Med.* **8**, 117 (2025).

29. D. Khullar, A.I. Could Worsen Health Disparities, *The New York Times* (2019). https://www.nytimes.com/2019/01/31/opinion/ai-bias-healthcare.html.

30. American Hospital Association, Fast Facts on U.S. Hospitals, 2025 (2025). https://www.aha.org/statistics/fast-facts-us-hospitals.

31. RAND Corporation, RAND Hospital Data: Web-Based Tool, (2018); https://www.rand.org/pubs/tools/TL303.html.

32. E. Prager, M. Schmitt, Employer consolidation and wages: Evidence from hospitals. *Am. Econ. Rev.* **111**, 397–427 (2021).

33. M. Schmitt, Do hospital mergers reduce costs? *J. Health Econ.* **52**, 74–94 (2017).

34. A. S. Oseran, R. K. Wadhera, E. J. Orav, J. F. Figueroa, Effect of Medicare Advantage on Hospital Readmission and Mortality Rankings. *Ann. Intern. Med.* **176**, 480–488 (2023).

35. R. M. Werner, E. T. Bradlow, Relationship Between Medicare's Hospital Compare Performance Measures and Mortality Rates. *JAMA* **296**, 2694–2702 (2006).

36. A. K. Jha, E. J. Orav, J. Zheng, A. M. Epstein, Patients' Perception of Hospital Care in the United States. *N. Engl. J. Med.* **359**, 1921–1931 (2008).

37. D. Arkhangelsky, S. Athey, D. A. Hirshberg, G. W. Imbens, S. Wager, Synthetic Difference-in-Differences. *Am. Econ. Rev.* **111**, 4088–4118 (2021).

38. C. Whaley, M. Bartlett, G. Bai, Financial Performance Gaps Between Critical Access Hospitals and Other Acute Care Hospitals. *JAMA Health Forum* **5**, e243959 (2024).

39. F. Fui-Hoon Nah, R. Zheng, J. Cai, K. Siau, L. Chen, Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration. *J. Inf. Technol. Case Appl. Res.* **25**, 277–304 (2023).

40. E. G. Poon, C. H. Lemak, J. C. Rojas, J. Guptill, D. Classen, Adoption of artificial intelligence in healthcare: survey of health system priorities, successes, and challenges. *J. Am. Med. Inform. Assoc.* **32**, 1093–1100 (2025).

41. M. L. Burns, S.-Y. Chen, C.-A. Tsai, J. Vandervest, B. Pandian, P. Nong, D. A. Hanauer, A. Rosenberg, J. Platt, Generative AI costs in large healthcare systems, an example in revenue cycle. *Npj Digit. Med.* **8**, 579 (2025).

Tables and Figures

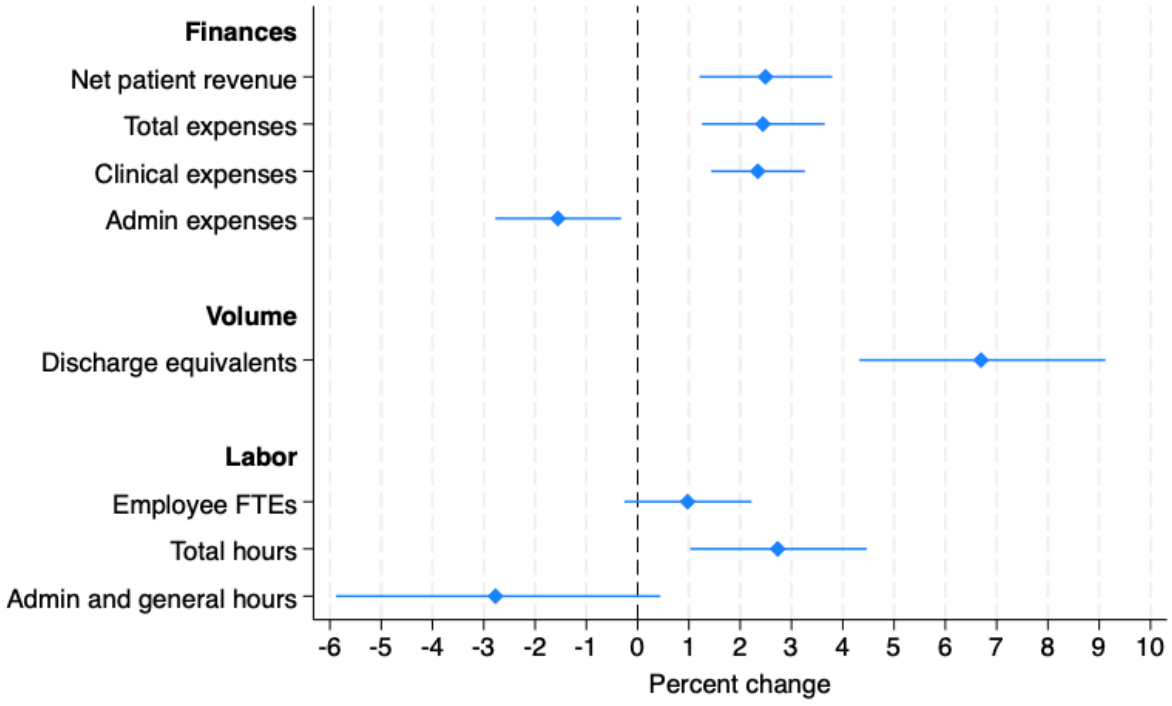


Fig. 1. Summary of AI adoption effects on revenues, expenses, volume, employment, efficiency. Shown are estimates from a synthetic difference-in-differences model where the dependent variable for each model corresponds to the logged version of the variable listed on the vertical axis. Coefficient estimates have been transformed using the formula exp(coef - 1)*100 to create the percentage effects shown in the figure. The regression model included hospital and year fixed and standard errors were clustered at the hospital level. The error bars represent 95% confidence intervals.

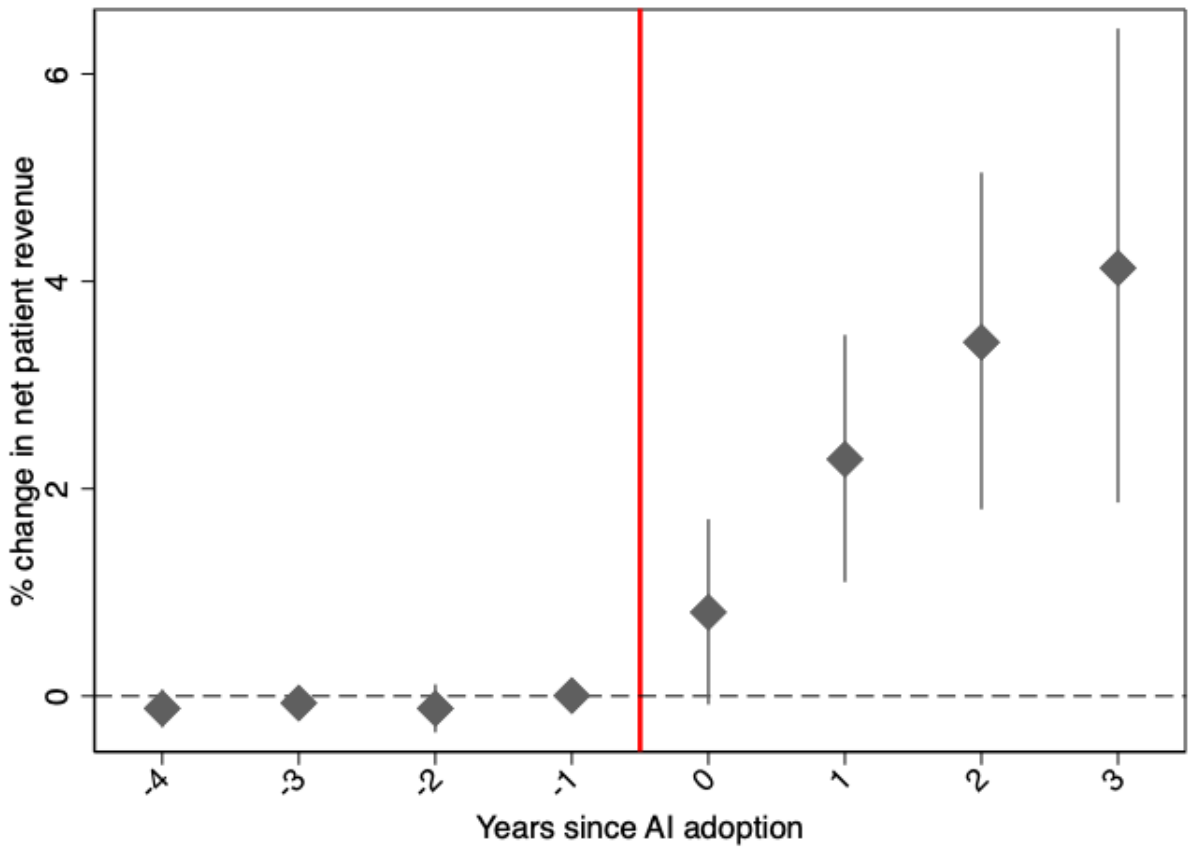


Fig. 2. Percentage change in net patient revenue after AI adoption. Shown are estimates from a synthetic difference-in-differences event study where log(net patient revenue) is the dependent variable. Coefficient estimates have been transformed using the formula exp(coef - 1)*100 to create the percentage effects shown in the figure. The regression model included hospital and year fixed and standard errors were clustered at the hospital level. The error bars represent 95% confidence intervals.

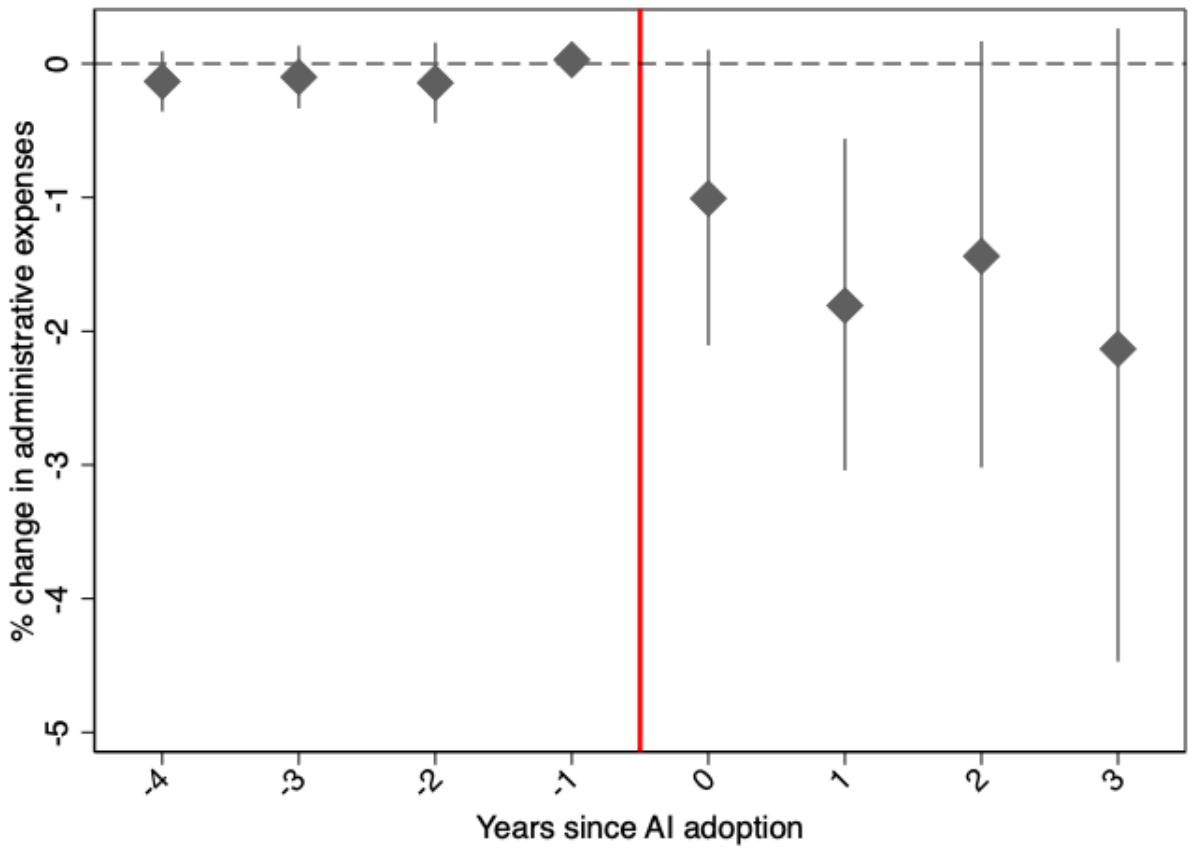


Fig. 3. Percentage change in administrative expenses after AI adoption. Shown are estimates from a synthetic difference-in-differences event study where log(administrative expenses) is the dependent variable. Coefficient estimates have been transformed using the formula exp(coef - 1)*100 to create the percentage effects shown in the figure. The regression model included hospital and year fixed and standard errors were clustered at the hospital level. The error bars represent 95% confidence intervals.

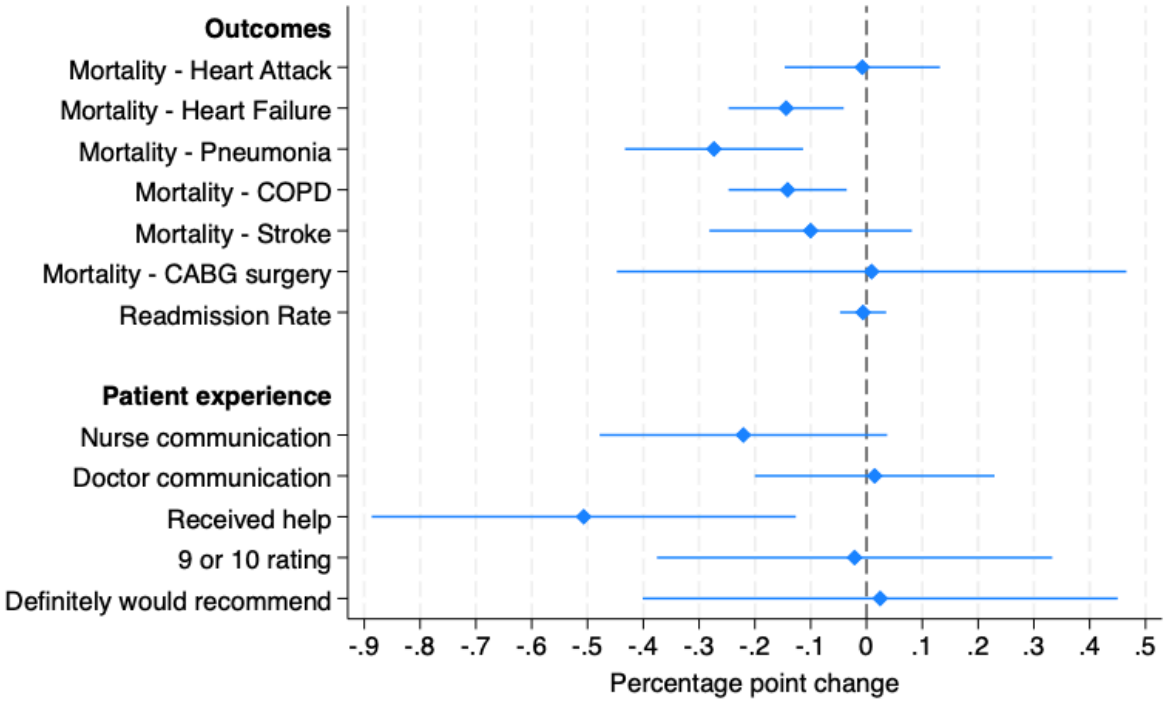


Fig. 4. Summary of AI adoption effects on quality. Shown are estimates from a synthetic difference-in-differences model where the variable listed on the vertical axis is the dependent variable for each model. The regression model included hospital and year fixed and standard errors were clustered at the hospital level. The error bars represent 95% confidence intervals.

	Estimate (95% CI)
Finances	
Net patient revenue	$40.1 billion ($19.4 billion to $61.0 billion)
Total expenses (millions of $)	$40.0 billion ($20.5 billion to $59.6 billion)
Clinical expenses (millions of $)	*$19.9 billion ($12.2 billion to $27.7 billion)*
Admin. expenses (millions of $)	*-$4.4 billion (-$7.9 billion to -$0.9 billion)*
Volume	
Discharge equivalents	6.0 million (3.8 million to 8.1 million)
Labor	
Total hours	519 billion (195 billion to 849 billion) ~65k employees (24k to 106k employees)*
Quality	
Mortality - heart failure	1,867 fewer in-hospital deaths (533 to 3,200)
Mortality - pneumonia	5,760 fewer in-hospital deaths (2,346 to 9,173)
Mortality - COPD	1,370 fewer in-hospital deaths (374 to 2,333)

Table 1. Aggregate effects of hospital AI adoption, 2021-2024. Shown are aggregated effects calculated by multiplying the SDID estimates by 1,221 (the number of adopting hospitals), 3.6 (the average number of adoption years), and the relevant adopting hospital sample means from Table S1 in the appendix. Full calculation details are included in the appendix.

The Impacts of AI Integration on Efficiency and Employment Patterns in American Hospitals

Daniel R. Arnold, PhD;[1] Jonathan Cantor, PhD;[2] Mallory Go, AB;[1] Arjun Teotia, PhD;[3,4] Christopher M. Whaley, PhD[1]

Supplemental Appendix

Appendix Table of Contents

Materials and Methods p. 2

Tables S1 p. 7

Figs. S1 to S8 p. 9

References p. 26

Materials and Methods

A. Empirical Strategy

The Synthetic Difference-in-Differences (SDID) method is a statistical technique used for causal inference with panel data, designed to estimate the effects of a policy or event (*3*). It combines features from two widely used methods: Difference-in-Differences (DID) and Synthetic Control (SC).

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 the treatment. From SC, SDID borrows the idea of reweighting untreated (control) units to create a "synthetic" control group. This synthetic control is constructed so that its outcome trends in the pre-treatment period closely match the trends of the treated group. SDID incorporates the core logic of DID by allowing for time-invariant differences between the treated and control groups (i.e., unit fixed effects). This means the method focuses on matching pre-treatment trends rather than requiring a perfect match in pre-treatment outcome levels, 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 SDID estimate process involves three main steps:

1. Compute Unit Weight ($\hat{\omega}$) : Find weights for the control units to make the weighted average trend of their pre-treatment outcomes parallel to the average trend of the treated units.
2. Compute Time Weights ($\hat{\lambda}$): Find weights for the pre-treatment time periods that best balance the pre- and post-treatment periods for the control units.
3. Estimate the Effect ($\hat{\tau}$): Use these unit and time weights in a weighted two-way fixed-effects regression to estimate the treatment effect. The core estimating equation is:

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

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. In Fig SXX we show the raw pre-treatment trends for treated and control hospitals for all our outcome variables. The trends are frequently different across treated and control groups and there are sizable level differences given treated hospitals are generally larger. The following lays out the preliminaries of our SDID implementation.

Data and Groups:

- Data: A balanced panel of U.S. hospitals from 2017-2024.
- Units: 4,022 general acute care hospitals
- Treated Group: 1,221 hospitals that adopted AI between 2021 and 2023.
- Control Group: The remaining 2,801 hospitals that did not adopt.

Treatment and Timing:

- Treatment: The adoption of AI tools for administrative or clinical workflows.
- Timing: Staggered adoption as hospitals adopting in 2021 (XX hospitals), 2021 (), and 2023 ().

Estimation proceeded as follows.

1. **Cohort-specific estimation:** The SDID algorithm is applied separately for each adoption cohort (e.g., hospitals that first adopted AI in 2021, those that adopted in 2022, etc.). For each cohort:
    a. The pre-treatment period is defined as the years from 2017 up to the year before adoption.
    b. The control group consists of all 2,801 hospitals that had not yet adopted AI.
    c. Unit weights ($\hat{\omega}$) are calculated to create a synthetic control from the non-adopting hospitals whose average pre-adoption outcome trends are parallel to those of the specific adoption cohort.
    d. Time weights ($\hat{\lambda}$) are calculated for the pre-adoption years to improve the precision of the counterfactual prediction.
    e. A cohort-specific treatment effect is estimated using the weighted DID regression.
2. **Aggregate Treatment Effect:** The individual treatment effects from each adoption cohort are then averaged, weighted by the size of each cohort, to produce a single, overall Average Treatment Effect on the Treated (ATT) for AI adoption. We estimated the effects described in #1 and #2 using the sdid package in Stata (*4*).

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. This is particularly useful for assessing when an effect emerges and whether it grows or fades over time.

3. **The Dynamic Cohort-Specific Estimator ($\hat{\tau}^{sdid}_{a,l}$):** The core of the event-study extension is the dynamic estimator, which measures the treatment effect for a specific adoption cohort (a) at a specific time lag (l) after treatment begins. Instead of averaging over all post-treatment periods, this estimator compares the outcome difference between a treated cohort and its synthetic control at a single point in time ($a-1+l$) to the average difference observed across the entire pre-treatment period. The estimator for the effect l periods after treatment for cohort a is defined as:

$$\hat{\tau}^{sdid}_{a,l} = \left(\frac{1}{N^a_{tr}}\sum_{i\in I^a} Y_{i,a-1+l} - \sum_{i=1}^{N_{co}} \omega_i Y_{i,a-1+l}\right) - \sum_{t=1}^{a-1}\lambda_t\left(\frac{1}{N^a_{tr}}\sum_{i\in I^a} Y_{i,t} - \sum_{i=1}^{N_{co}} \omega_i Y_{i,t}\right)$$

   By construction, the standard cohort-specific SDID estimate ($\hat{\tau}^{sdid}_{\alpha}$) is simply the average of these dynamic estimators over all post-treatment periods for that cohort.

4. **The Aggregated Dynamic Estimator ($\hat{\tau}^{sdid}_{l}$):** In a staggered adoption setting where different cohorts receive treatment at different times, it's useful to have a single estimate for the effect at a specific lag l (e.g., the average effect two years post-treatment). This can be achieved by calculating a weighted average of the cohort-specific dynamic effects ($\hat{\tau}^{sdid}_{\alpha,l}$) across all cohorts that can be 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 (i.e., the 2021 cohort in our context). The formula for this aggregated dynamic estimator is:

$$\hat{\tau}^{sdid}_{l} = \sum_{a\in A_l} \frac{N^a_{tr}}{N^l_{tr}}\hat{\tau}^{sdid}_{a,l}$$

   Where N^l_{tr} is the total number of treated units across all cohorts observed at lag l.

5. **Connection to the Overall ATT:** This framework is internally consistent as the overall ATT from the standard SDID procedure can be recovered from these dynamic estimates. The overall ATT is a weighted average of the aggregated dynamic estimators ($\hat{\tau}^{sdid}_{l}$), with weights proportional to the total number of units for which the effect at lag l can be calculated. This relationship is shown by the equation:

$$\hat{ATT} = \frac{1}{T_{post}}\sum_{l=1}^{T_{tr}} N^l_{tr}\hat{\tau}^{sdid}_{l}$$

   This equation demonstrates how the overall average effect is composed of the period-by-period dynamic effects. We estimated the event study effects described in #3-#5 using the sdid_event package in Stata, an extension of the sdid package (*5*).

B. Details on the aggregate effects calculations

To estimate the aggregate impact of adoption across the U.S. healthcare system, we extrapolated our SDID estimates to the full population of 1,221 adopting hospitals over the 2021–2024 study period. We calculated the average duration of adoption within this four-year window to be 3.6 years.

Finances and Volume Estimates. For financial and volume-based outcomes, we calculated cumulative effects by applying the estimated percentage change (SDID coefficient) to the baseline sample mean (Table S1), scaled by the number of adopters and the duration of adoption.

- *Revenue & Expenses*: Applying the 2.5% increase in net patient revenue to the baseline mean of $366 million yielded an aggregate increase of $40.1 billion. Total expenses followed a similar trajectory, rising by $40.0 billion (based on a 2.4% estimate). Decomposing these costs revealed a shift in resource allocation: while clinical expenses increased by $19.9 billion, administrative expenses decreased by $4.4 billion.
- *Volume*: Operational throughput also expanded; applying the 6.7% estimate to the baseline of 20,233 discharge equivalents suggested a total increase of 6.0 million discharge equivalents across the adopting cohort.

Labor Estimates. To contextualize the increase in labor hours, we converted aggregate hours into employee counts. We estimated a total increase of approximately 519 million labor hours (2.7% effect size × 4.3 million baseline hours × 1,221 hospitals × 3.6 years). Using a standard full-time equivalent (FTE) benchmark of 2,000 hours per year (40 hours/week × 50 weeks) over the 4-year period, this increase corresponds to approximately 65,000 additional employees.

Mortality Estimates. For mortality outcomes, we employed a population-based market share approach. We first derived annual national hospitalization volumes for Heart Failure, Pneumonia, and COPD by applying historical hospitalization rates (per 100,000) to the average U.S. adult population for 2021–2024. We assumed that adopting hospitals—which account for one-third of total U.S. hospital beds—treated a proportional one-third of these national cases. We then calculated in-hospital death reductions by applying the SDID percentage-point reduction to this specific patient volume.

- *Heart Failure*: With an estimated 4 million national hospitalizations over the study period, the adopting hospitals managed approximately 1.33 million cases. Applying the 0.14 percentage-point reduction in mortality yielded an estimated 1,867 fewer in-hospital deaths.
- *Pneumonia*: Based on 6.4 million national hospitalizations, the adopting hospitals managed approximately 2.13 million cases. The 0.27 percentage-point reduction corresponded to 5,760 fewer in-hospital deaths.

- *COPD*: Based on 2.8 million national hospitalizations, the adopting hospitals managed approximately 0.93 million cases. The 0.14 percentage-point reduction corresponded to 1,307 fewer in-hospital deaths.

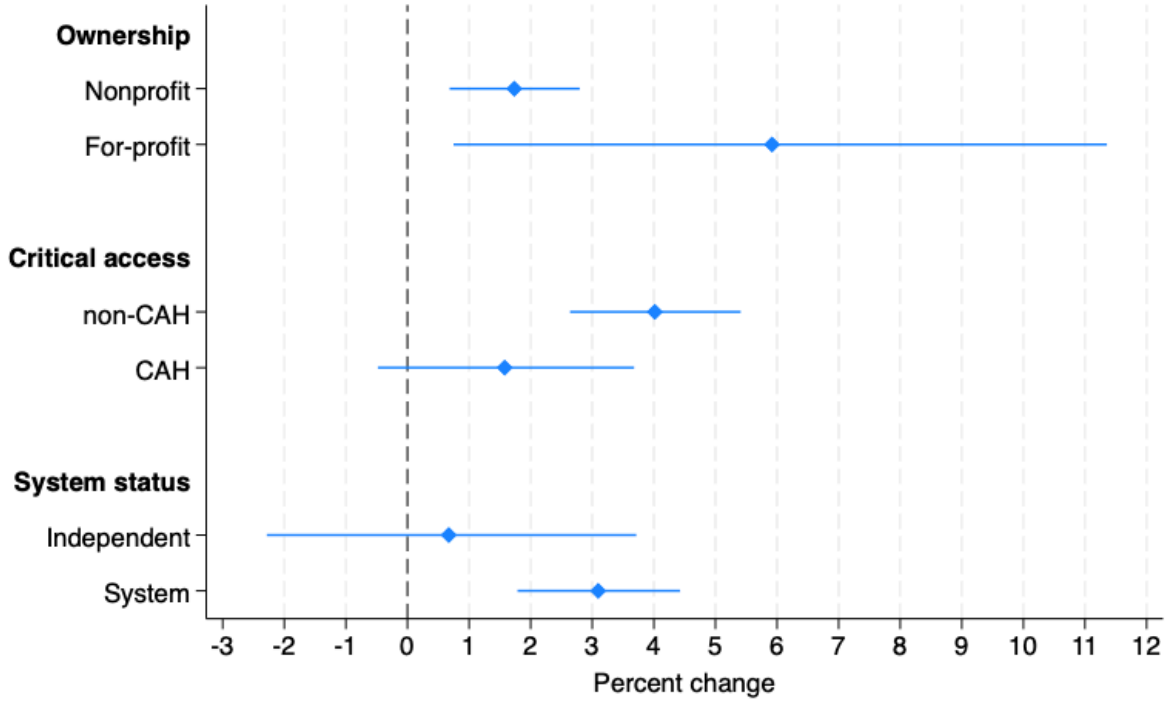
Table S1. Baseline (2017) characteristics of Hospital AI adopters and non-adopters

	AI adopters	Non-adopters	Standardized difference
N	1,221	2,801	
Finances and Volume			
Net patient revenue (millions of $)	366 (517)	146 (217)	0.56
Total expenses (millions of $)	372 (530)	144 (218)	0.56
Expenses for clinical activities (millions of $)	193 (270)	75 (117)	0.56
Expenses for administrative activities (millions of $)	65 (87)	26 (39)	0.57
Discharge equivalents	20,233 (20,503)	9,319 (12,302)	0.64
Labor			
Employees on payroll (full-time equivalents)	1,729 (2,388)	722 (1,038)	0.55
Total paid hours (thousands)	4,322 (5,233)	2,313 (2,600)	0.49
Paid hours, administrative and general (thousands)	430 (670)	250 (321)	0.34
Quality			
Outcomes			
30-day mortality rate for patients admitted to the hospital with a heart attack (%)	13.4 (1.2)	13.6 (1.20)	-0.16
30-day mortality rate for patients admitted to the hospital with a heart failure (%)	11.7 (1.6)	12.1 (1.6)	-0.23
30-day mortality rate for patients admitted to the hospital with pneumonia (%)	15.7 (2.0)	16.2 (1.9)	-0.24
30-day mortality rate for patients admitted to the hospital with COPD (%)	8.0 (1.1)	8.1 (1.1)	-0.11
30-day mortality rate for patients admitted to the hospital with a stroke (%)	14.4 (1.7)	14.7 (1.6)	-0.15
30-day mortality rate for patients who have undergone CABG surgery at the hospital (%)	3.1 (0.8)	3.4 (0.9)	-0.30
30-day all-cause readmission rate (%)	15.2 (0.8)	15.3 (0.8)	-0.12
Patient Experience			
Patients who reported that their nurses "Always" communicated well (%)	79.9 (4.3)	80.1 (5.5)	-0.05
Patients who reported that their doctors "Always" communicated well (%)	80.8 (4.1)	82.1 (5.6)	-0.26
Patients who reported that they "Always" received help as soon as they wanted (%)	66.3 (7.6)	68.5 (9.1)	-0.26
Patients who gave their hospital a rating of 9 or 10 on a	73.0 (7.0)	72.2 (8.4)	0.09

scale from 0 (lowest) to 10 (highest) (%)			
Patients who reported YES, they would definitely recommend the hospital (%)	73.2 (8.0)	71.1 (9.1)	0.24
Hospital characteristics			
Beds	236 (243)	117 (147)	0.59
Part of a system	0.91 (0.29)	0.62 (0.49)	0.73
Rural-urban continuum code (1=most urban, 10=most rural)	2.6 (2.1)	4.1 (2.7)	-0.60
Critical Access Hospital	0.16 (0.37)	0.39 (0.49)	-0.53
For-profit	0.06 (0.24)	0.21 (0.40)	-0.44
Major teaching hospital	0.15 (0.35)	0.05 (0.22)	0.33
Medicare casemix index (CMI)	1.69 (0.31)	1.57 (0.31)	0.39
Medicare inpatient discharge share	0.35 (0.14)	0.44 (0.17)	-0.55
Medicaid inpatient discharge share	0.09 (0.09)	0.10 (0.10)	-0.10
Operating margin	0.031 (0.149)	0.003 (0.227)	0.15

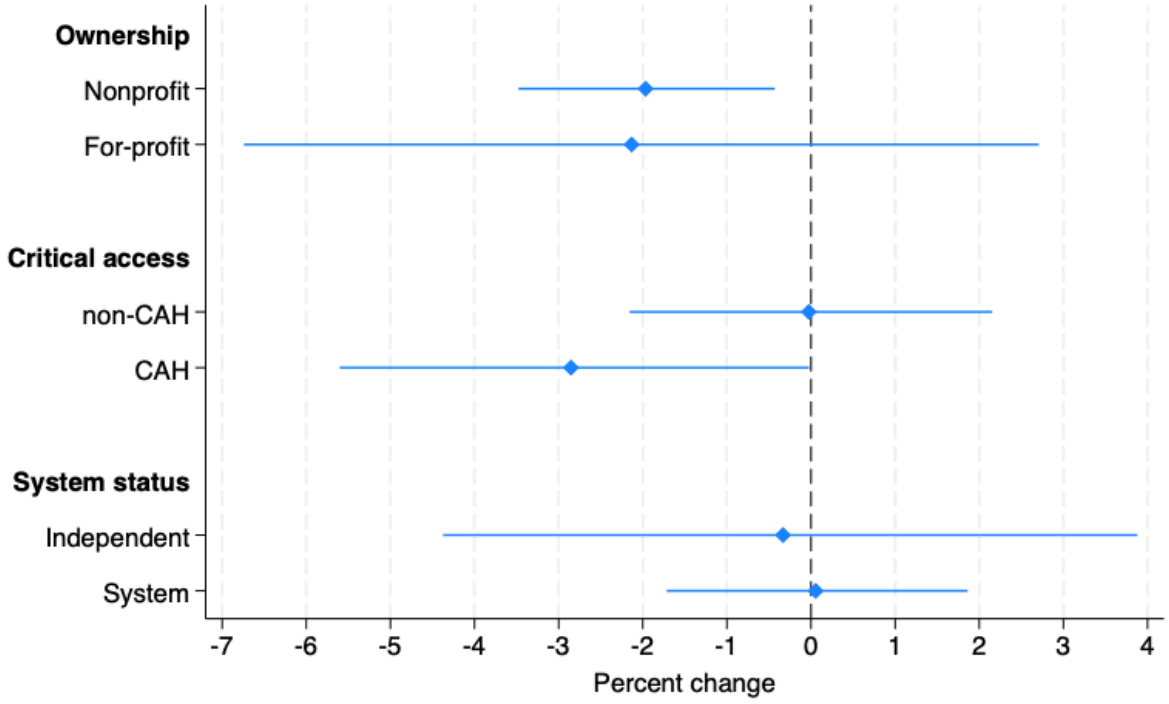
Sources: AHA's Annual Survey (AI adoption), CMS's Hospital Compare (quality measures), Hospital Cost Report data as collected by RAND (all other measures).

Fig. S1. 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: Every event study shown was estimated using the sdid_event command in Stata. The dependent variable for each event study was a logged version of the panel's title. Placebo standard errors clustered at the hospital level.

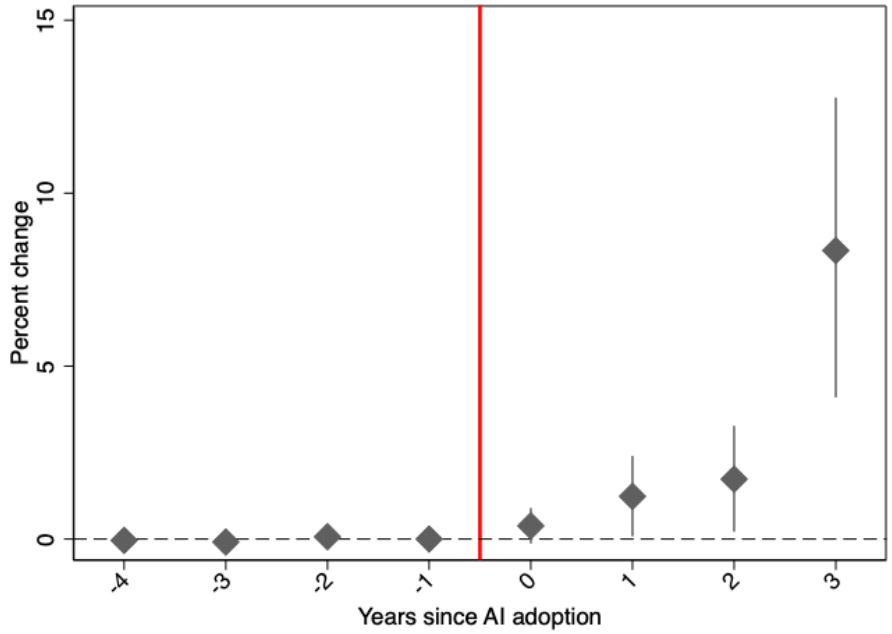
Fig. S2. Administrative expenses heterogeneity



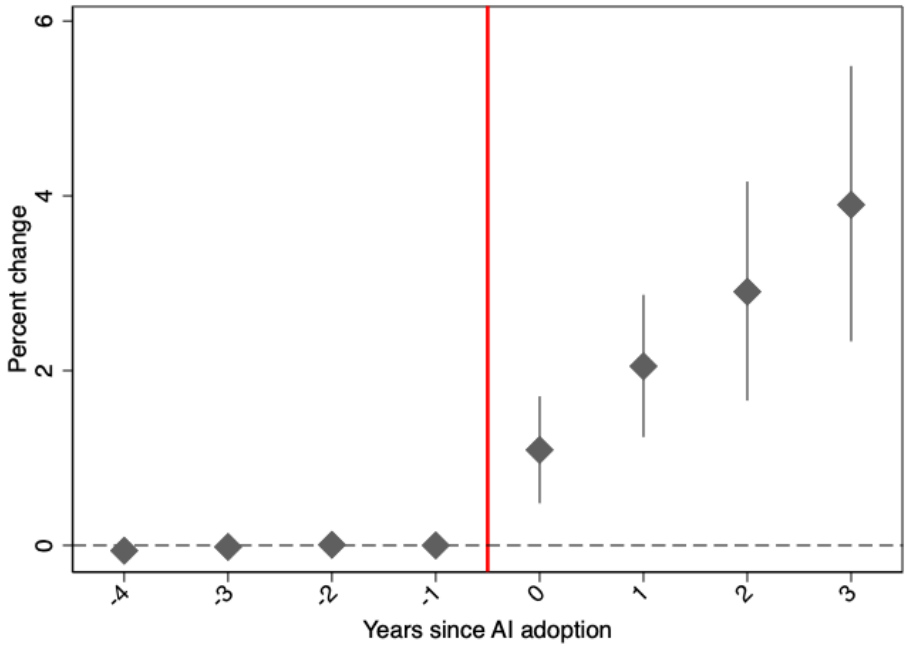
Sources: AHA's Annual Survey (AI adoption) and Hospital Cost Report data as collected by RAND (expenses and discharge equivalents).
Notes: Every event study shown was estimated using the sdid_event command in Stata. The dependent variable for each event study was a logged version of the panel's title. Placebo standard errors clustered at the hospital level.

Fig. S3. Additional finance and volume event studies

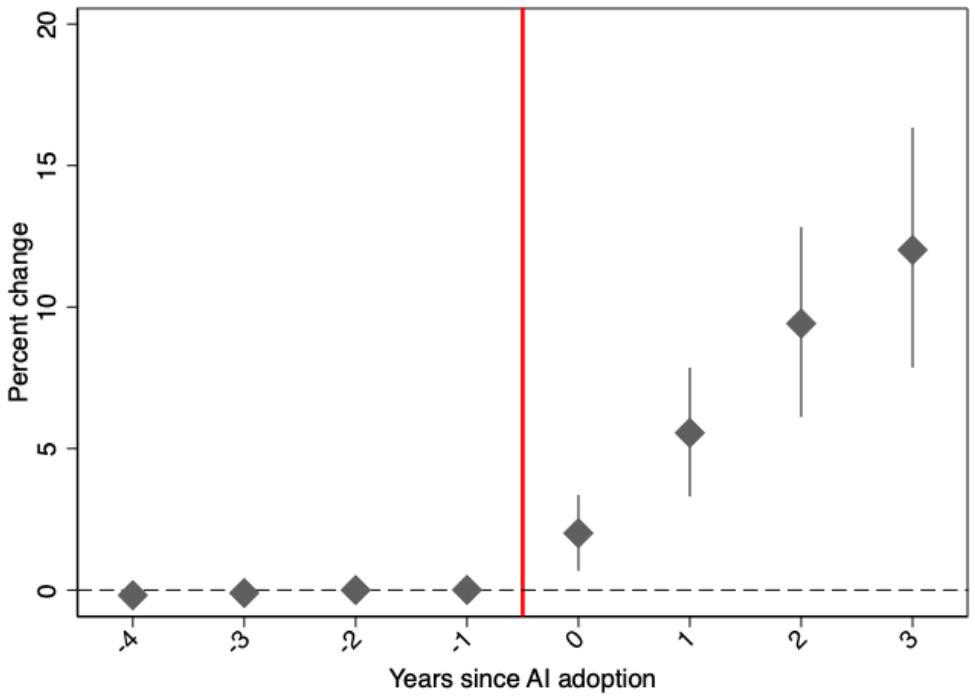
(A) Total expenses



(B) Expenses for clinical activities



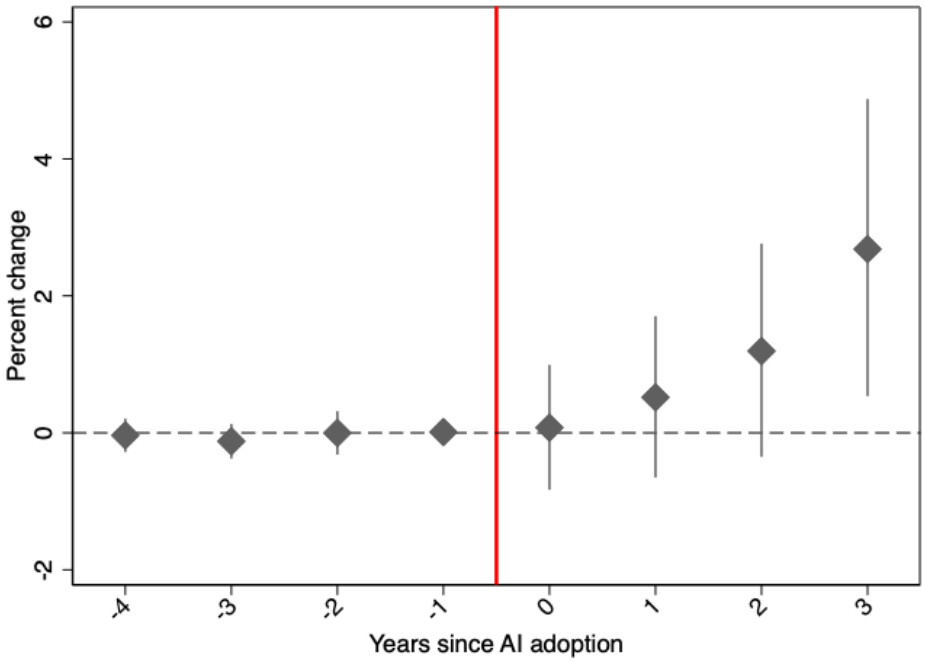
(C) Discharge equivalents



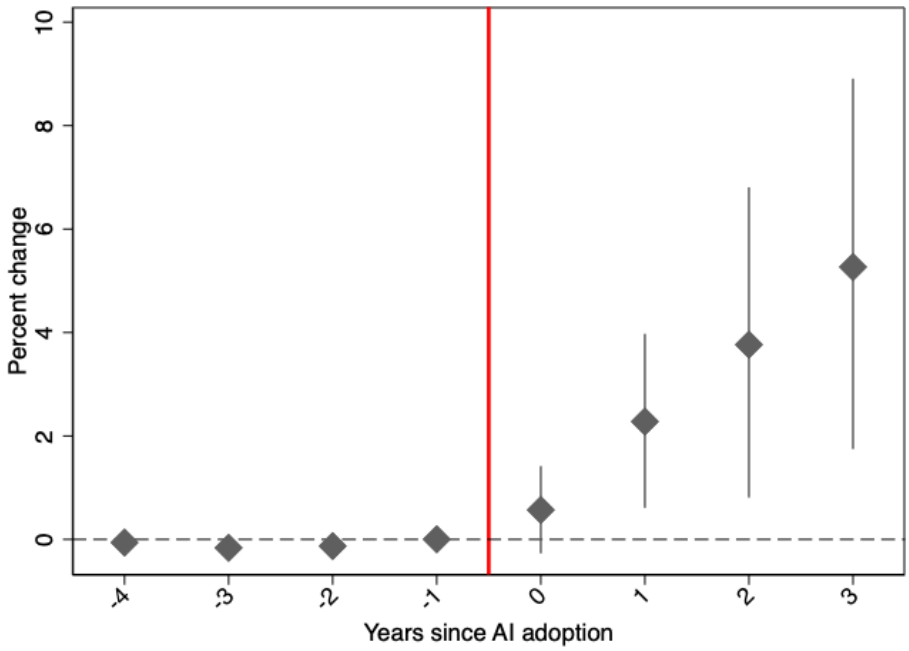
Sources: AHA's Annual Survey (AI adoption) and Hospital Cost Report data as collected by RAND (expenses and discharge equivalents).
Notes: Every event study shown was estimated using the sdid_event command in Stata. The dependent variable for each event study was a logged version of the panel's title. Placebo standard errors clustered at the hospital level.

Fig. S4. Labor market event studies

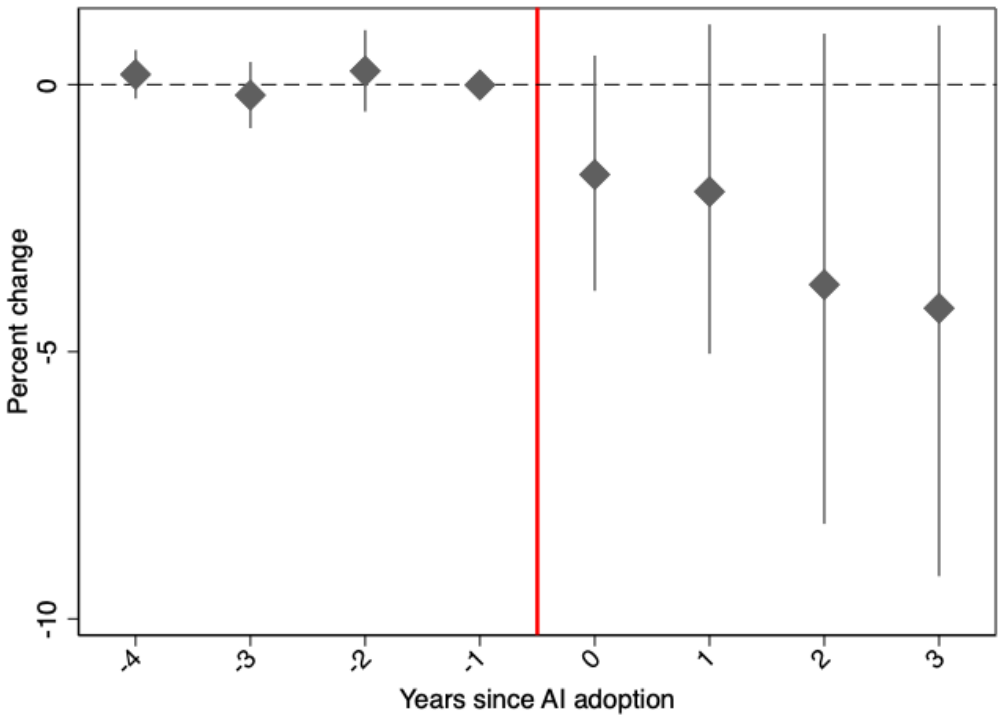
(A) Employee FTEs



(B) Total hours



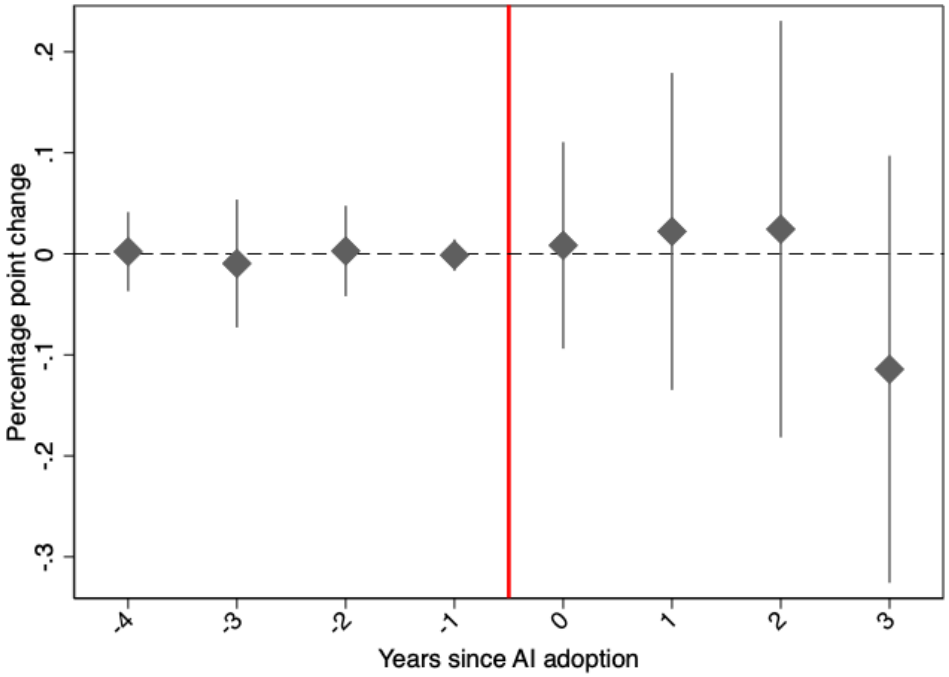
(C) Administrative and general hours



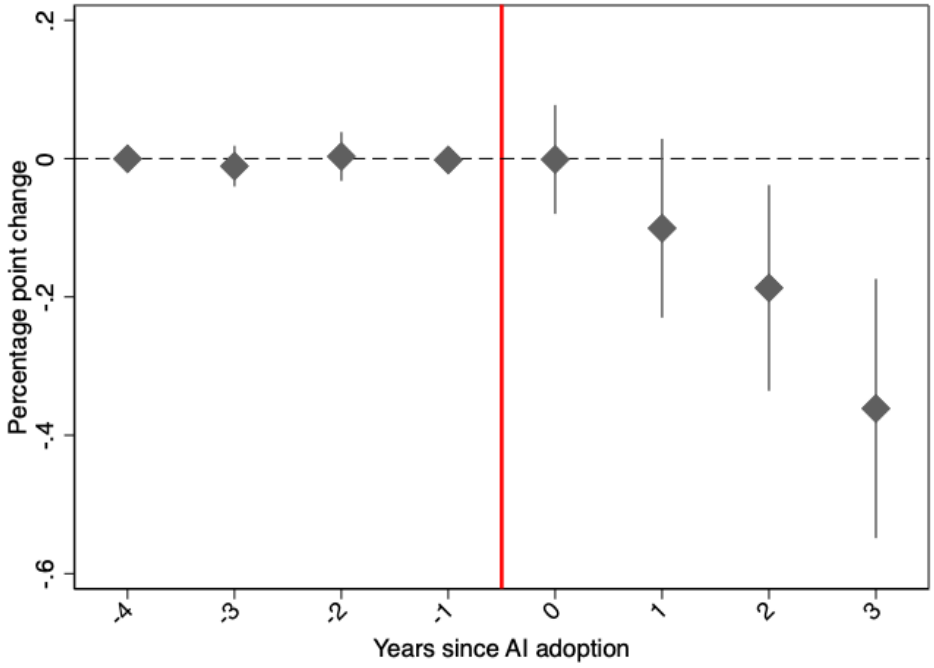
Sources: AHA's Annual Survey (AI adoption) and Hospital Cost Report data as collected by RAND (FTEs and hours).
Notes: Every event study shown was estimated using the sdid_event command in Stata. The dependent variable for each event study was a logged version of the panel's title. Placebo standard errors clustered at the hospital level.

Fig. S5. Quality event studies - outcomes

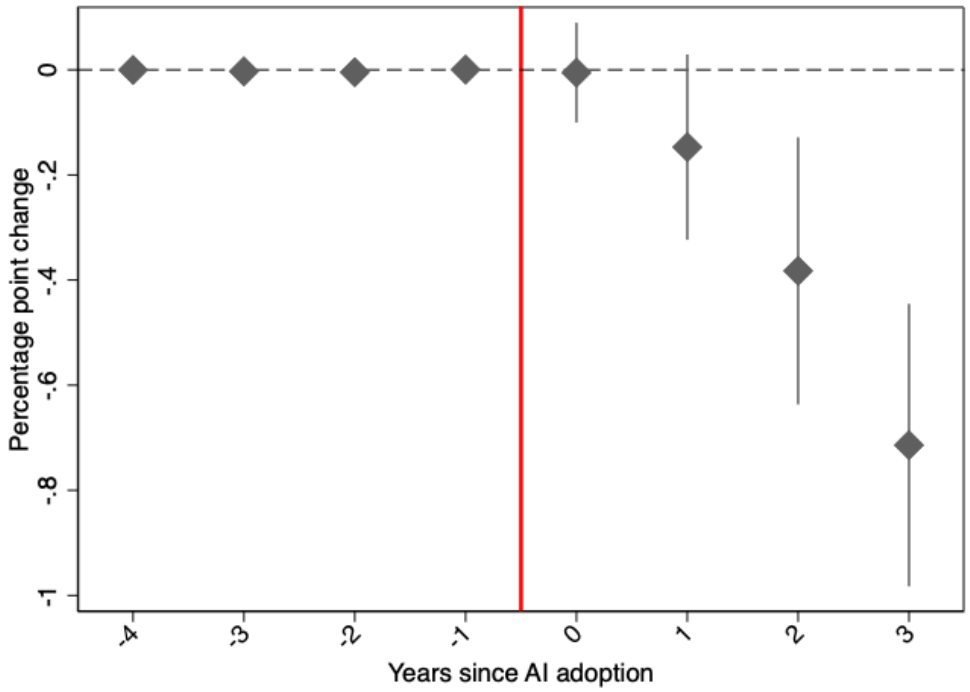
(A) Mortality rate - heart attack



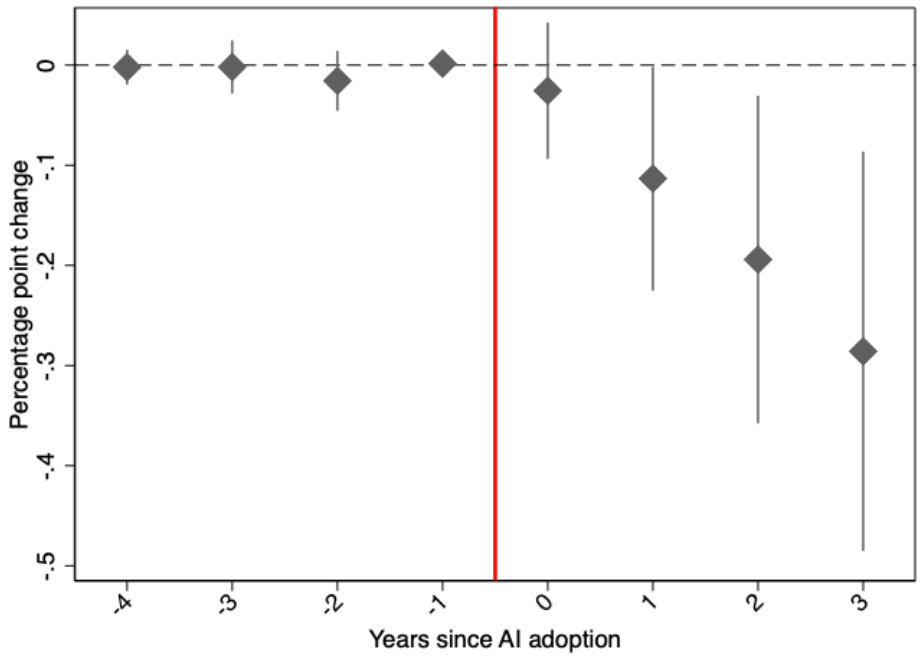
(B) Mortality rate - heart failure



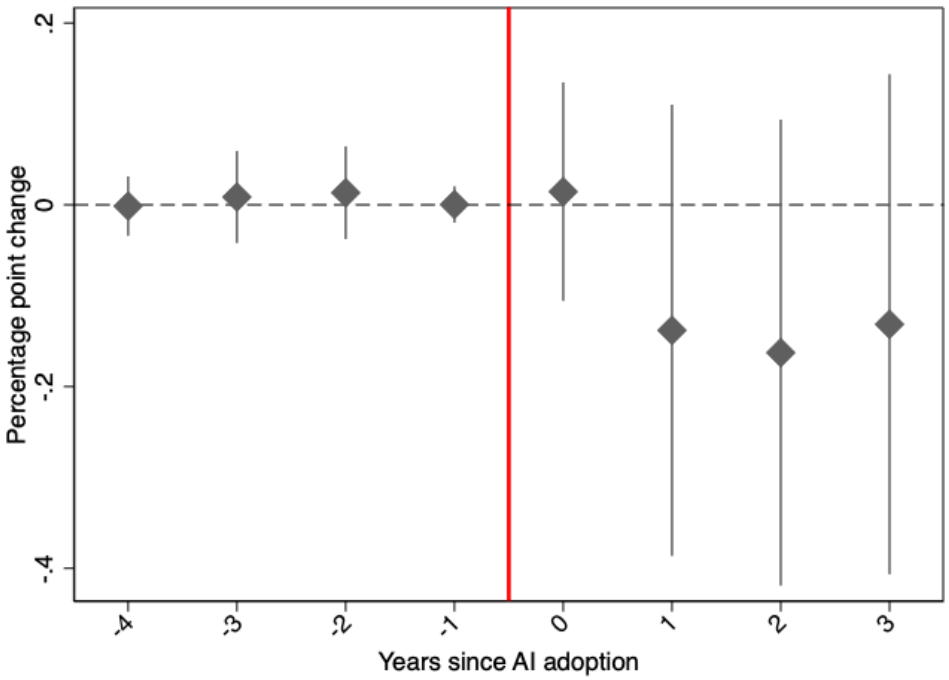
(C) Mortality rate - pneumonia



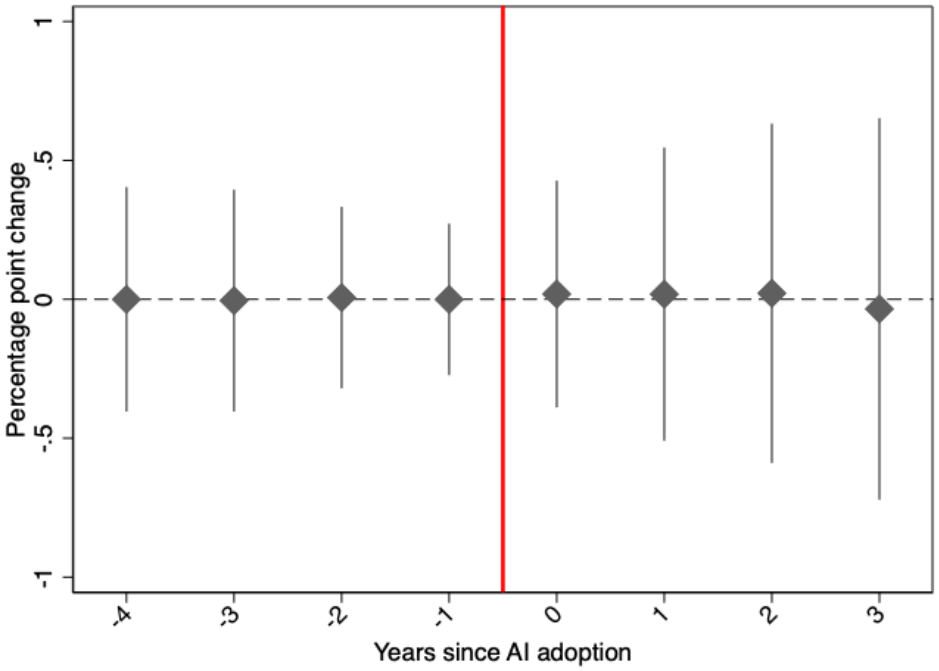
(D) Mortality rate - COPD



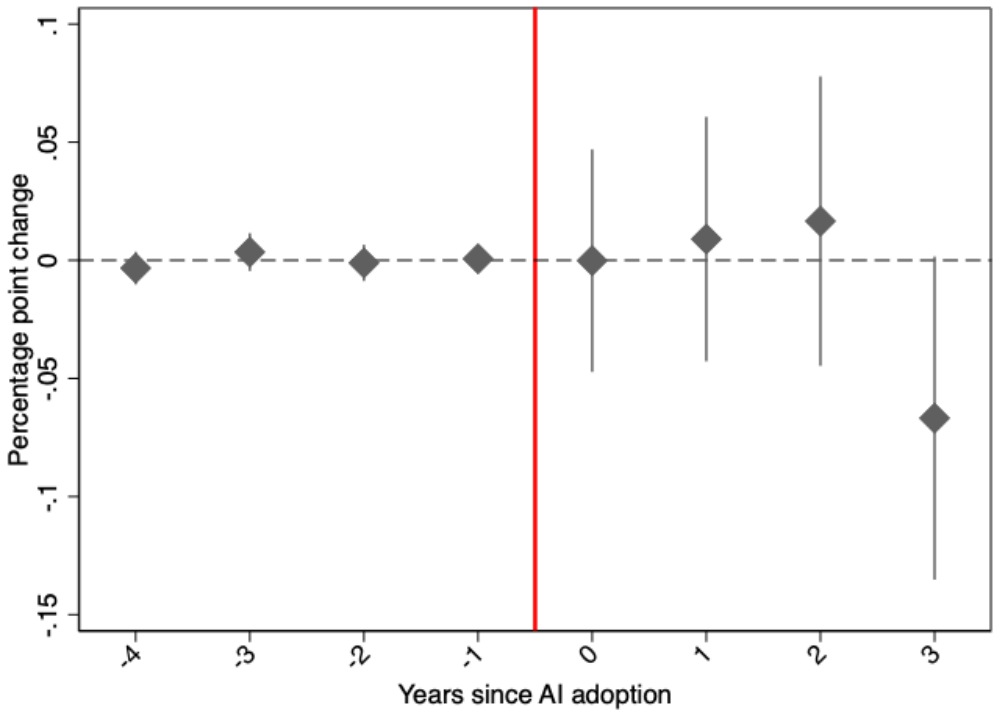
(E) Mortality rate - stroke



(F) Mortality rate - CABG surgery



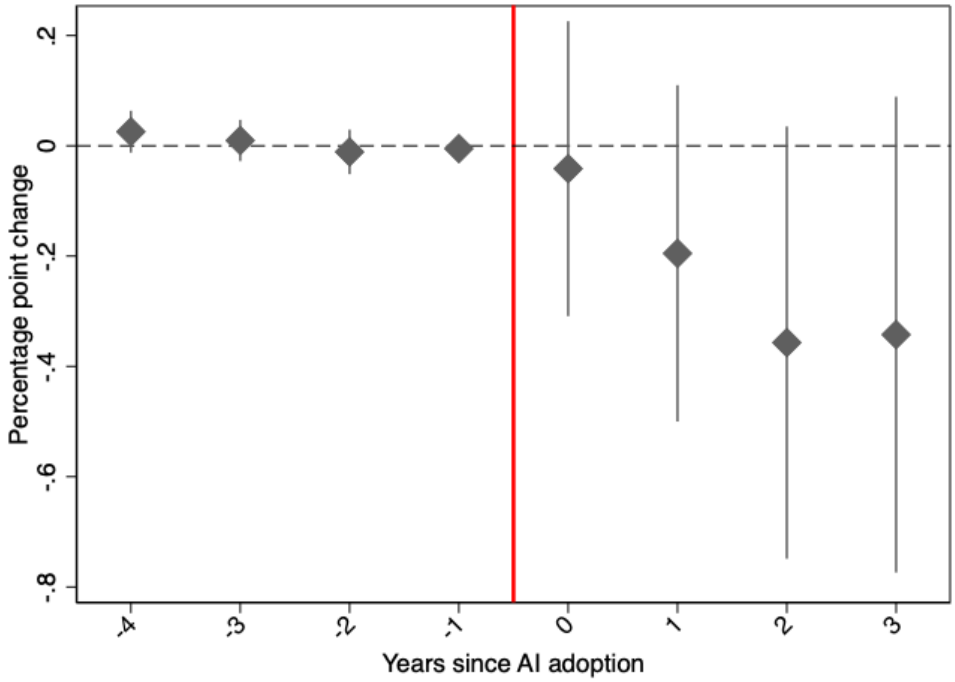
(G) All-cause readmission rate



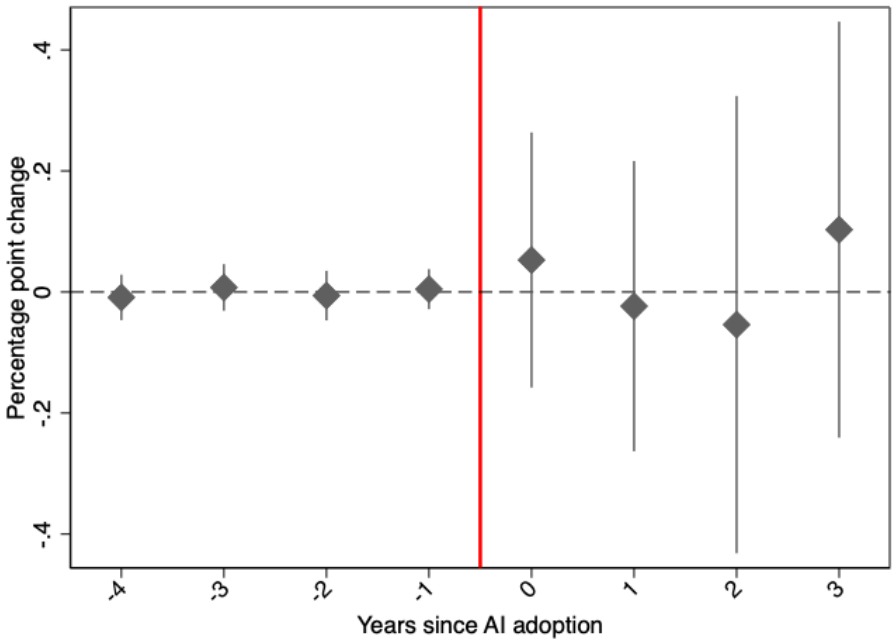
Sources: AHA's Annual Survey (AI adoption) and CMS Hospital Compare (quality).
Notes: Every event study shown was estimated using the sdid_event command in Stata. The dependent variable for each event study was each panel's title. Placebo standard errors clustered at the hospital level.

Fig. S6. Quality event studies - patient experience

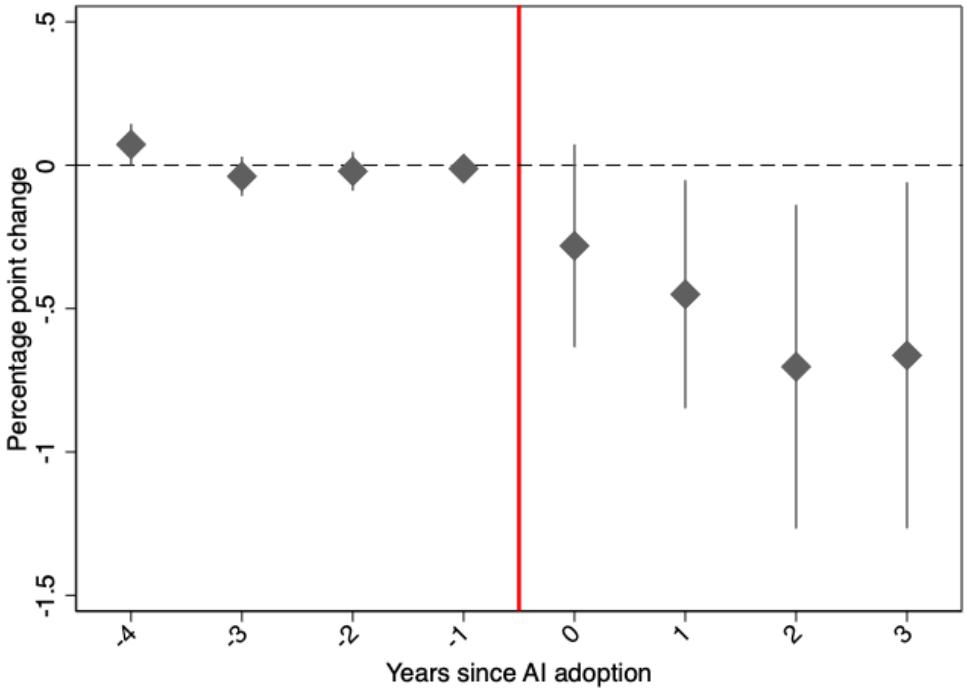
(A) Nurse communication



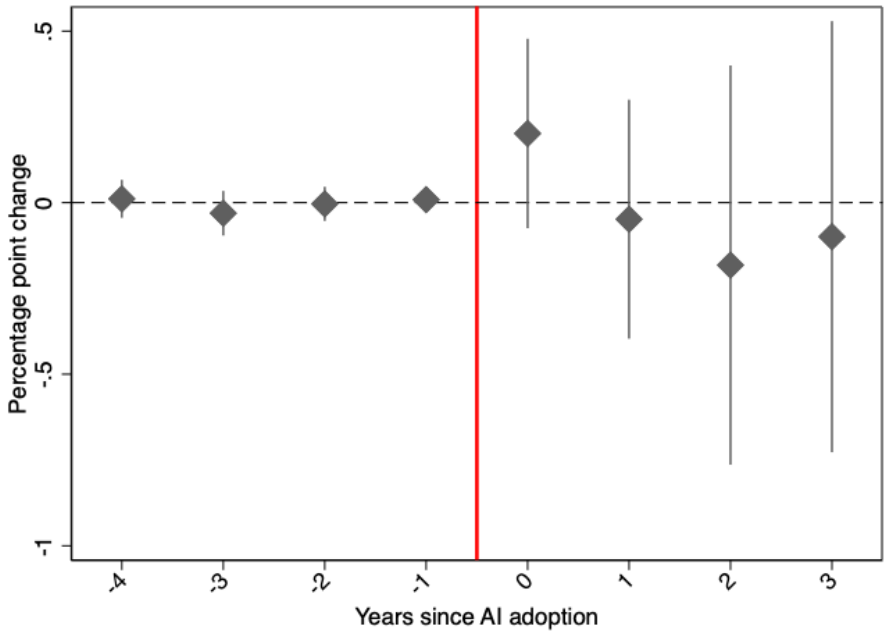
(B) Doctor communication



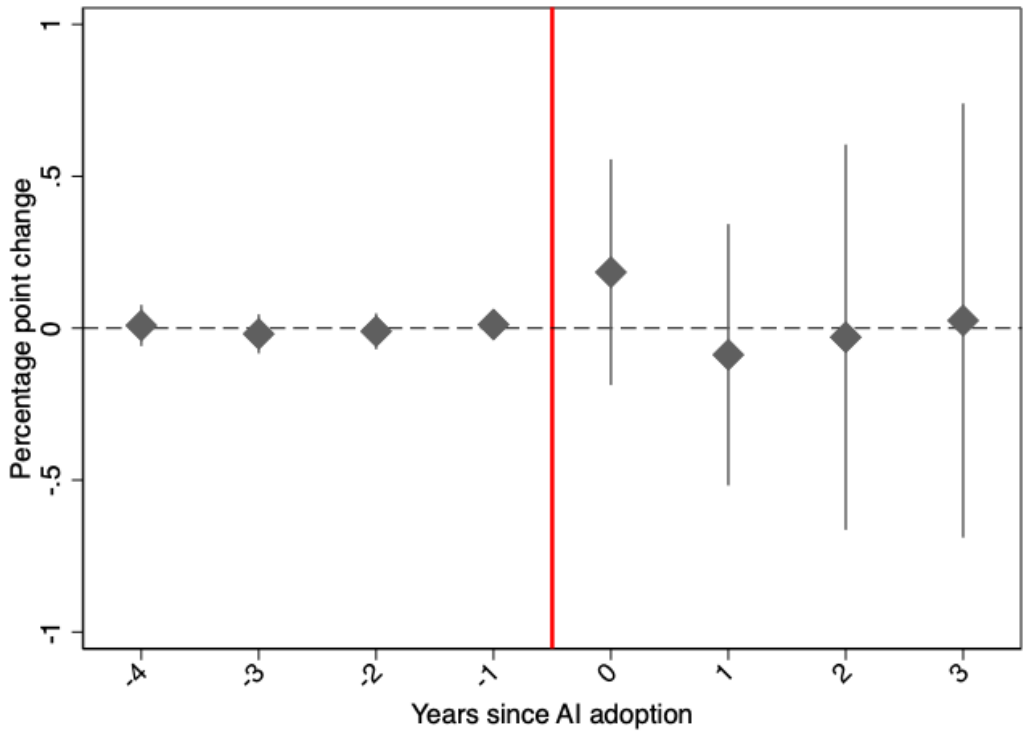
(C) Receiving help



(D) 9 or 10 rating



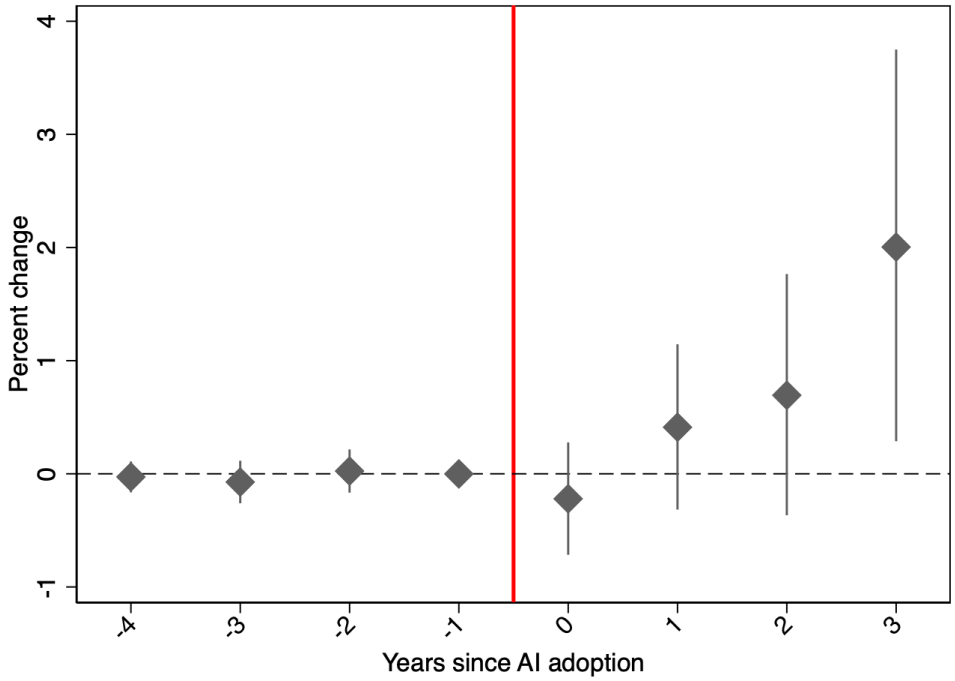
(E) Recommend hospital



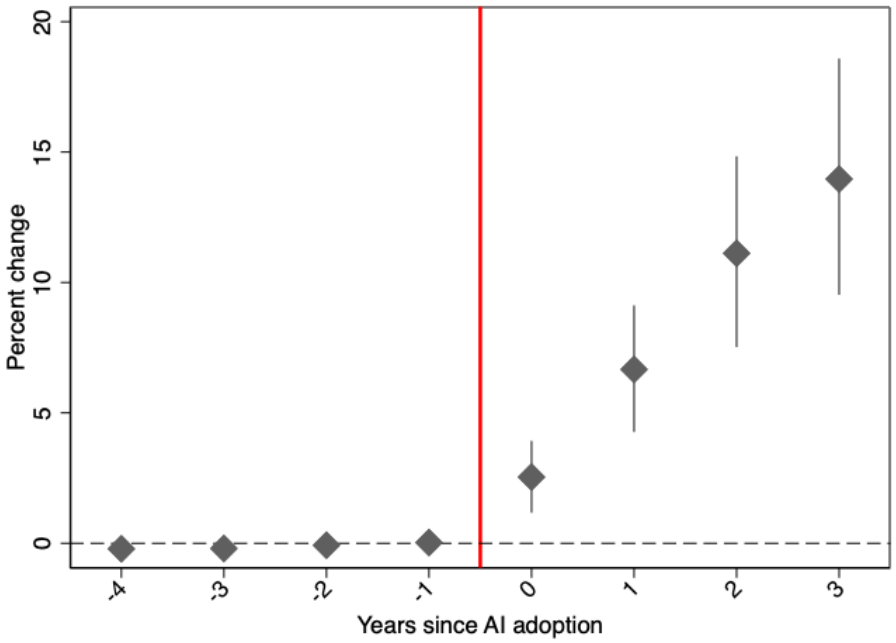
Sources: AHA's Annual Survey (AI adoption) and CMS Hospital Compare (quality).
Notes: Every event study shown was estimated using the sdid_event command in Stata. The dependent variable for each event study was each panel's title. Placebo standard errors clustered at the hospital level.

Fig. S7. Additional volume analyses - beds and service type

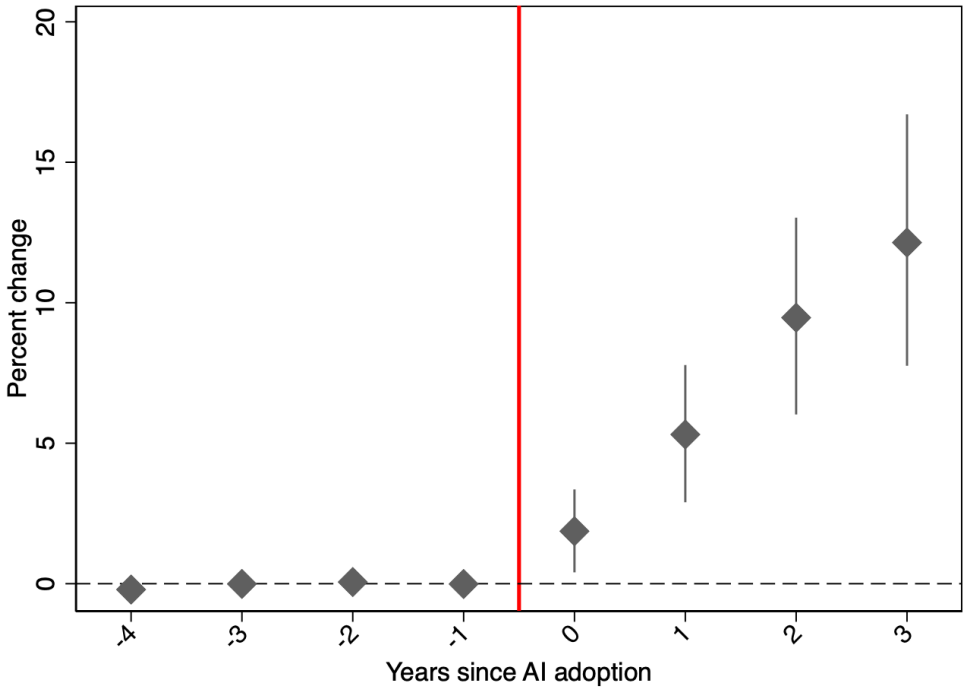
(A) Beds



(B) Inpatient discharges



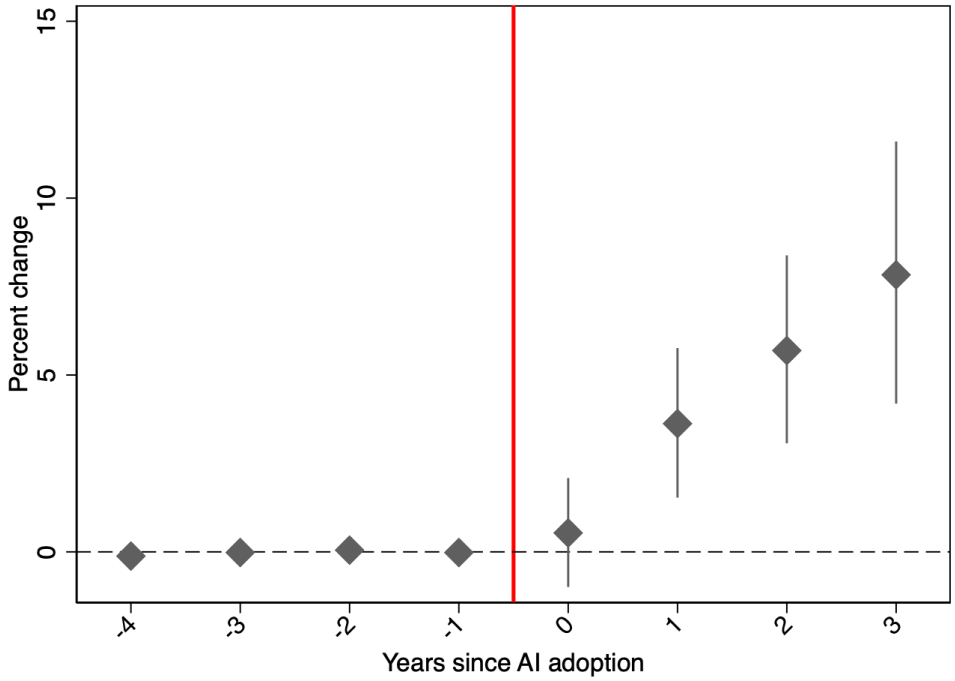
(C) Outpatient discharge equivalents



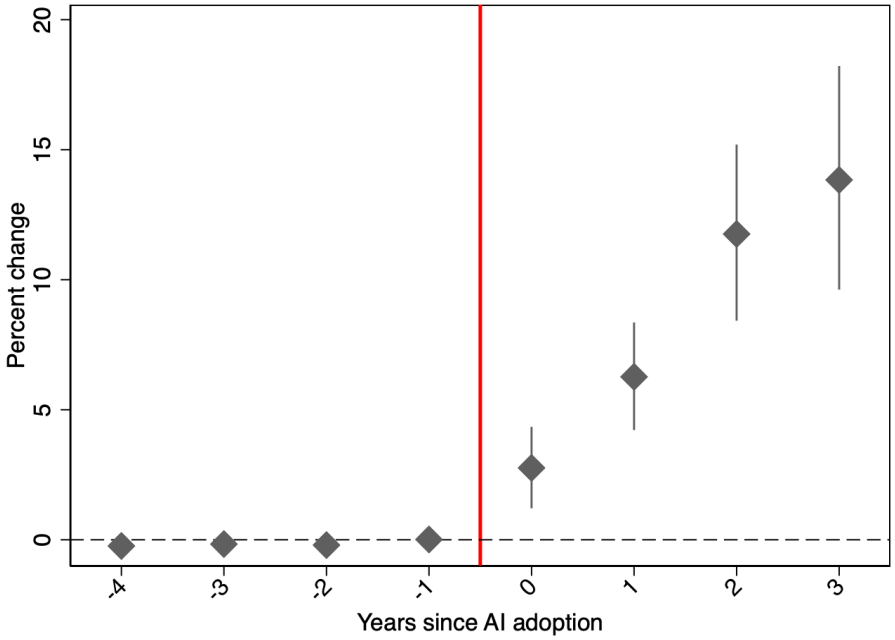
Sources: AHA's Annual Survey (AI adoption) and CMS Hospital Compare (quality).
Notes: Every event study shown was estimated using the sdid_event command in Stata. The dependent variable for each event study was each panel's title. Placebo standard errors clustered at the hospital level.

Fig. S8. Additional volume analyses - payer

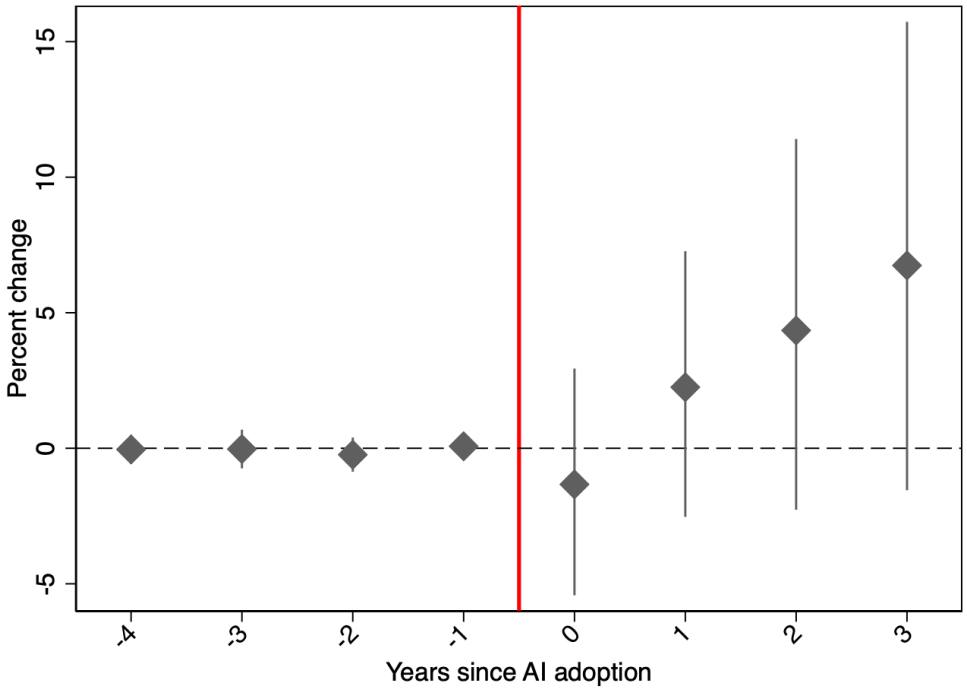
(A) Commercial inpatient discharges



(B) Medicare inpatient discharges



(C) Medicaid inpatient discharges



Sources: AHA's Annual Survey (AI adoption) and CMS Hospital Compare (quality).
Notes: Every event study shown was estimated using the sdid_event command in Stata. The dependent variable for each event study was each panel's title. Placebo standard errors clustered at the hospital level.

Appendix References

1. American Hospital Association, Fast Facts on U.S. Hospitals, 2025 (2025). https://www.aha.org/statistics/fast-facts-us-hospitals.

2. RAND Corporation, RAND Hospital Data: Web-Based Tool, (2018); https://www.rand.org/pubs/tools/TL303.html.

3. D. Arkhangelsky, S. Athey, D. A. Hirshberg, G. W. Imbens, S. Wager, Synthetic Difference-in-Differences. *Am. Econ. Rev.* **111**, 4088–4118 (2021).

4. D. Clarke, D. Pailañir, S. Athey, G. Imbens, On synthetic difference-in-differences and related estimation methods in Stata. *Stata J.* **24**, 557–598 (2024).

5. D. Ciccia, D. Clarke, D. Pailañir, SDID_EVENT: Stata module providing Synthetic Difference-in-Differences (SDID) event-study estimators. *Stat. Softw. Compon.* (2024).