Variation in AI Exposure Across Physician Specialties: A Medicare Claims Analysis

PRELIMINARY – PLEASE DO NOT CITE OR CIRCULATE

Dan Zeltzer, PhD
School of Public Health, University of California, Berkeley, CA

Tyler Tam
Stanford University, Stanford, CA

Benji Bhattacharya
Stanford University, Stanford, CA

Victoria Udalova, PhD
U.S. Census Bureau, Washington, D.C.

Maria Polyakova, PhD
School of Medicine, Stanford University, Stanford, CA and National Bureau of Economic Research (NBER), Cambridge, Massachusetts

Corresponding author: Dan Zeltzer, PhD, University of California, Berkeley, CA. E-mail: dzeltzer@berkeley.edu.

Any opinions and conclusions expressed herein are those of the authors and do not reflect the views of the U.S. Census Bureau. The U.S. Census Bureau reviewed this data product for unauthorized disclosure of confidential information and approved the disclosure avoidance practices applied to this release under numbers CBDRB-FY2020-CES005–035, CBDRB-FY21-009, CBDRB-FY2021-CES010–015, CBDRB-FY2021-CES010-01, CBDRB-FY23-0319. All numeric values were rounded according to U.S. Census Bureau disclosure protocols to preserve data privacy.

ABSTRACT

Importance: Artificial intelligence (AI) capabilities in healthcare are expanding rapidly, raising many questions about the future of the physician workforce. Yet many physician services require direct, in-person physical interaction that cannot be performed by current AI systems. How much the need for physical interactions is likely to constrain the AI substitutability of physician services is poorly understood.

Objective: To estimate the share of physician services that require in-person physical interaction with a patient—and are therefore less exposed to substitution by software-based AI in the near term—overall and across medical specialties.

Design, Setting, and Participants: Cross-sectional analysis of 2023 Medicare Fee-for-Service Part B billing records for 589,854 physicians across 62 specialties in the United States.

Main Outcomes and Measures: The primary outcome was the Medicare payment-weighted proportion of services classified as requiring in-person physical interaction between physician and patient, overall and by specialty. Healthcare Common Procedure Coding System (HCPCS) service codes were classified manually or with large language models (LLM) using Restructured Berenson-Eggers Type of Service (BETOS) category rules and code descriptions. Evaluation and Management (E&M) codes were excluded from classification due to heterogeneous delivery modalities and reported separately.

Results: Of 4,929 unique HCPCS codes representing $44.9 billion in 2023 Medicare Part B payments, 4,683 were non-E&M. Of these, 68% were classified as requiring a physical interaction, accounting for 55% of non-E&M payments ($11 of $20 billion). The physical interaction share varied substantially across specialties. A higher specialty-level share of physical services was positively associated with average years of training, weekly work hours, and annual income in a specialty.

Conclusions and Relevance: The requirement for in-person physical interaction constrains the near-term potential of software-based AI to substitute for physician services, but this constraint differs markedly across specialties. Procedural and surgical specialties have the lowest estimated AI exposure; specialties dominated by imaging interpretation and cognitive services have the highest. Most of Medicare spending on physician services is concentrated in specialties where physical services dominate, suggesting lower near-term AI exposure. However, physicians are disproportionately concentrated in specialties with a higher share of E&M and non-physical services, increasing their potential exposure to AI substitution.

INTRODUCTION

Artificial intelligence (AI) systems can now interpret medical images (Rajpurkar et al. 2022; Najjar, 2023; Esteva et al. 2017), generate text-based diagnostic assessment (Singhal et al. 2023, Zeltzer et al. 2025), and conduct patient-facing communication with increasing proficiency (Tu et al. 2025). Yet a fundamental constraint separates these capabilities from direct substitution of physician work: many physician services require in-person physical interaction - examination, procedural intervention, or hands-on treatment - that software-based AI can inform but not carry out.

A growing body of research in economics and related fields has examined the likely impacts of AI on different occupations using broad task taxonomies (Autor et al. 2003; Brynjolfsson et al. 2017; Acemoglu et al. 2022; Acemoglu 2025) or LLM-specific exposure scoring (Eloundou et al. 2024; Felten et al. 2021). These studies fail to capture how physician services are organized, delivered, and reimbursed. Administrative billing records, which document the specific services each physician performs, offer a more precise basis for estimating AI exposure. To our knowledge, no study has used service-level data to systematically quantify physicians' potential AI exposure across specialties.

This study classified each physician service billed to Medicare Fee-for-Service Part B in 2023 according to whether it required in-person physical interaction with a patient. Services requiring physical interaction were considered to have low near-term exposure to substitution by software-based AI. Using a reproducible classification of each HCPCS code, we derived Medicare payment-weighted shares of physical services overall and by specialty. We hypothesized that the prevalence of physical services varies substantially across specialties, with procedural and surgical fields - which require prolonged training and account for a smaller share of the physician workforce (Gottlieb et al. 2025) - having the lowest AI exposure, and the largest specialties facing greater exposure and greater uncertainty.

METHODS

Participants

The study used two data sources: a national provider registry to establish the physician universe and specialty-level workforce counts, and Medicare billing records to characterize the services each specialty provides.

The physician universe was drawn from the December 2023 National Provider and Enumeration System (NPPES), which records a specialty for each active National Provider Identifier (NPI) in

the U.S., using the NUCC taxonomy. Records classified as allopathic or osteopathic physicians (primary taxonomy codes beginning with "20") were selected. NUCC specialties were mapped to Medicare Provider Enrollment, Chain, and Ownership System (PECOS) specialty categories using the December 2023 Medicare Provider and Supplier Taxonomy Crosswalk. When codes mapped to multiple categories, the subspecialty was retained to create one-to-one mappings. NPI records whose NUCC taxonomy lacked a corresponding PECOS specialty (e.g., medical examiners) were excluded.

PECOS specialty-level data on average income, weekly work hours, and years of training were obtained from Earnings and Labor Supply of U.S. Physicians dataset of the U.S. Census Bureau that contains machine-readable files for Gottlieb et al. (2025).

Billing data were obtained from the 2023 Centers for Medicare and Medicaid Services (CMS) Medicare Physician and Other Practitioners - by Provider and Service file. This file contains all fee-for-service Medicare Part B billing records aggregated at the provider-service level. Providers were identified by NPI and services by Healthcare Common Procedure Coding System (HCPCS) code. Restructured Berenson-Eggers Type of Service (BETOS) categories were assigned to each HCPCS using the 2024 BETOS classification release. Records at different places of service were aggregated to the provider-service level, summarizing service volume and total Medicare standardized payment amounts. Line items representing products rather than services were excluded: codes in the Durable Medical Equipment, Prosthetics, Orthotics, and Supplies (DMEPOS) BETOS category; vaccine product codes (HCPCS 90476–90759; 91300–91311); and HCPCS Level II codes for drugs, biologicals, and temporary supplies (A, J, and Q codes).

The two data sources were merged using NPI. NPI records present in billing data but absent from the NPPES registry were excluded. Physicians with at least one billing record comprised the analytic sample for the service-level analysis. Physicians registered in NPPES but absent from billing data - including those serving non-Medicare populations or in non-billing roles - were retained to compute the number of actively registered physicians in each specialty.

Outcomes

The primary outcome was the Medicare payment-weighted proportion of non-E&M services classified as requiring in-person physical interaction between physician and patient, overall and by specialty. We focus on payment-weighted shares because they better approximate the economic significance of each service category than unweighted code counts.

Each HCPCS code was classified as requiring physical interaction (“physical”) or not requiring physical interaction ("non-physical") using a three-step procedure. First, Evaluation and Management (E&M) codes were set aside due to their heterogeneous delivery modalities and reported separately. Second, codes in the Procedure and Anesthesia BETOS categories were

classified as physical; codes in Imaging and selected Tests categories (General Lab, Anatomic Pathology, Molecular) were classified as non-physical, with the exception of three specific HCPCS codes (36415, 92240, and 93313) classified as physical as they require direct invasive patient contact. Third, remaining codes were classified based on code description using Anthropic Claude Opus 4.6, a Large Language Model (LLM). The model was prompted as follows: system prompt, "You are a medical expert"; user prompt: "Task: [HCPCS short description]. Question: Does the task described require a direct, in-person physical interaction/exam between patient and physician? Answer only 'Yes' or 'No.'" LLM analysis was performed in batch mode using Python 3.

Statistical Analysis

Code-level classification was merged into the provider-service billing data. Each non-E&M HCPCS code was classified as physical or non-physical. Records were then aggregated to the specialty level, computing for each specialty the share of codes in each category (physical, not physical, E&M), weighted by Medicare payments. Payment-weighted shares reflect the contribution of physical codes to Medicare's Part B spending and physician Medicare revenue. The number of physicians in each specialty was calculated from the full NPPES-derived study population, including physicians without Medicare billing records.

Descriptive univariable linear regression models were used to characterize cross-sectional associations between the specialty-level share of physical services and specialty characteristics (average income, weekly work hours, years of training), weighted by the number of physicians per specialty. Models were estimated with and without E&M codes. Analyses were performed in R version 4.5.2.

RESULTS

Study participants' characteristics

The analytic sample included 1,458,629 unique physicians in 62 specialties. Of these, 589,854 (40%) had at least one billing record and were included in the service-level analysis. These providers had billing records for 4,929 unique HCPCS codes, accounting for $44.9 billion in Medicare Part B payments.

HCPCS code classification

Figure 1 illustrates the total number of HCPCS codes by classification method and status. Of these, 5% (249 codes; $24.9 billion in Medicare spending) were in the Evaluation and Management (E&M) category. Among non-E&M codes, 68% by code count (3,174 codes; $11 billion in Medicare spending) were classified as requiring physical interaction between patient and

physician. Weighted by Medicare payments, 56% of services were in the E&M category and were not classified, and 54% of the remaining non-E&M services were classified as requiring physical interaction. eTable in the Supplement reports the classification method, status, and Medicare spending for each code.

Specialty-level analysis

The share of unclassified and physical services differed substantially across the 62 specialties examined. Figure 2 reports the share of unclassified services and services classified as physical for 15 specialties that had the highest Medicare billing volume in the sample. Among these, diagnostic radiology and pathology stood out: nearly all billed services were classified, and of those, close to 100% were classified as not requiring physical interaction. Radiation oncology similarly had few unclassified services, but 45% of classified services were deemed to require physical interaction. Hospitalists, family practice, internal medicine, and emergency medicine had the highest shares (approximately 90% or above) of codes in the unclassified E&M category. Specialties with the highest proportion of physical services were procedural and surgical specialties, including dermatology, nephrology, orthopedic surgery, ophthalmology, and urology.

Figure 3 shows the associations between the estimated share of physical services for a specialty and several specialty characteristics. Larger specialties, measured by number of providers, tended to have a larger share of unclassified E&M codes and a lower share of services classified as physical. The share of physical services was also positively associated with average weekly hours of work, average length of training in a specialty, and average income. Regression results are reported in eFigure in the Supplement and include specifications with E&M services alternative weighting.

DISCUSSION AND CONCLUSION

Physical interaction requirements impose substantial constraints on the near-term potential for software-based AI to substitute for physician services. However, physical services remain exposed to AI augmentation—for example, AI-assisted image recognition during endoscopy or AI-guided surgical navigation—meaning that even low-substitution specialties may experience significant productivity changes. The degree of AI impact is likely to markedly differ across specialties that require varying levels of physical interactions with patients.

Our findings reveal a tension in the distribution of AI exposure across the physician workforce. A substantial share of Medicare spending on physician services is concentrated in specialties where physical services dominate, suggesting that the economic footprint of physician work most resistant to AI substitution is large. At the same time, by count, most physicians are concentrated in specialties that provide a higher share of E&M and non-physical services, increasing their

potential exposure to AI substitution. The positive associations between a specialty's share of physical services and its average training length, work hours, and income suggest that the services most shielded from substitution are also those with the highest barriers to entry and the greatest per-physician investment. Conversely, the largest specialties - primary care, internal medicine, hospitalist medicine - combine high E&M shares with relatively lower training requirements, placing the broadest segment of the physician workforce in the path of greatest near-term AI disruption. These patterns have implications for workforce planning, training pipelines, and the distribution of returns to medical education as AI capabilities expand.

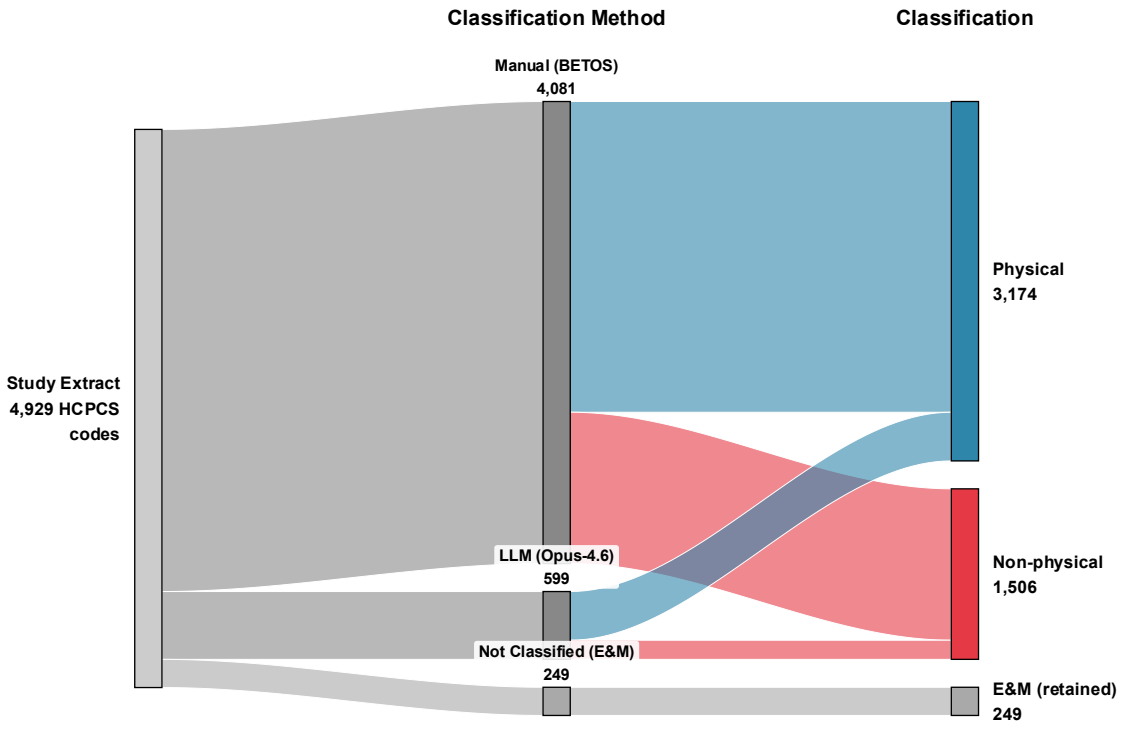
This study has several limitations. Our classification focuses on software-based AI in its current form. Emerging technologies - including robotic surgical systems, AI-guided interventional platforms, and AI-enabled diagnostic devices - are beginning to blur the boundary between physical and non-physical services. As these hardware-software integrations mature, specialties currently classified as low-exposure may face increasing AI exposure over longer time horizons, and our estimates should be interpreted as a near-term lower bound. Additionally, the analysis is limited to Medicare Fee-for-Service Part B, which underrepresents specialties primarily serving non-Medicare populations. E&M codes, which account for 56% of Medicare Part B physician spending, were not classified owing to heterogeneous delivery modalities. However, E&M services are predominantly cognitive (typically encompassing history-taking, clinical assessment, and care planning - tasks where LLMs have demonstrated increasing proficiency). Therefore, their exclusion likely makes our estimates of AI exposure conservative, particularly for primary care and hospital-based specialties where E&M dominates. As AI capabilities continue to advance, tracking how exposure translates into actual changes in physician workflow, productivity, and workforce demand across specialties will be an important priority for health policy research.

REFERENCES

1. Rajpurkar, Pranav, Emma Chen, Oishi Banerjee, and Eric J. Topol. "AI in health and medicine." *Nature medicine* 28, no. 1 (2022): 31-38.
2. Najjar, Reabal. "Redefining radiology: a review of artificial intelligence integration in medical imaging." *Diagnostics* 13, no. 17 (2023): 2760.
3. Esteva, Andre, Brett Kuprel, Roberto A. Novoa, Justin Ko, Susan M. Swetter, Helen M. Blau, and Sebastian Thrun. "Dermatologist-level classification of skin cancer with deep neural networks." *Nature* 542, no. 7639 (2017): 115-118.
4. Singhal, Karan, Shekoofeh Azizi, Tao Tu, S. Sara Mahdavi, Jason Wei, Hyung Won Chung, Nathan Scales et al. "Large language models encode clinical knowledge." *Nature* 620, no. 7972 (2023): 172-180.
5. Zeltzer, Dan, Zehavi Kugler, Lior Hayat, Tamar Brufman, Ran Ilan Ber, Keren Leibovich, Tom Beer et al. "Comparison of initial artificial intelligence (AI) and final physician

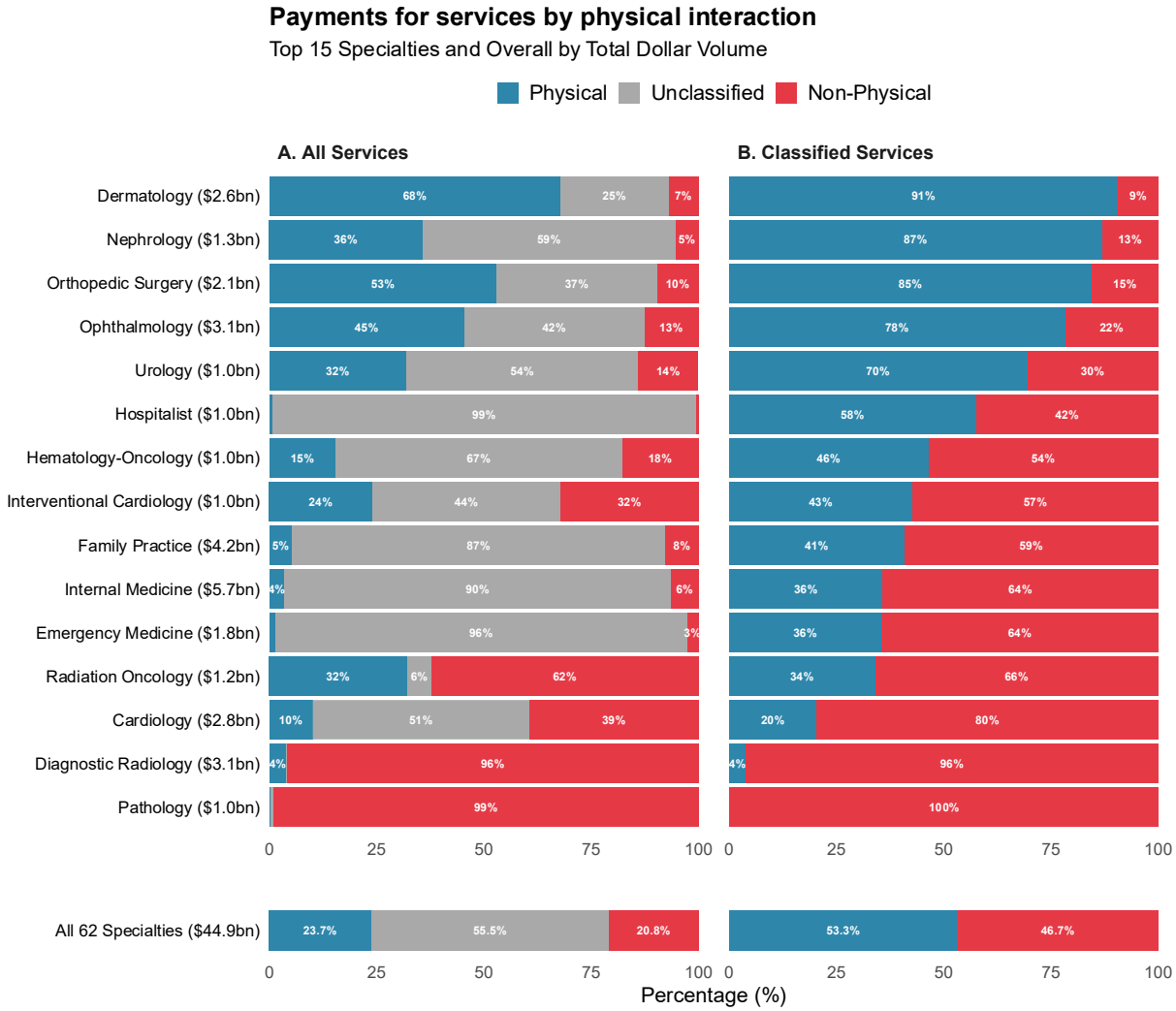
recommendations in AI-assisted virtual urgent care visits." *Annals of Internal Medicine* 178, no. 4 (2025): 498-506.
6. Tu, Tao, Mike Schaekermann, Anil Palepu, Khaled Saab, Jan Freyberg, Ryutaro Tanno, Amy Wang et al. "Towards conversational diagnostic artificial intelligence." *Nature* 642, no. 8067 (2025): 442-450.
7. Acemoglu, Daron, David Autor, Jonathon Hazell, and Pascual Restrepo. "Artificial intelligence and jobs: Evidence from online vacancies." *Journal of Labor Economics* 40, no. S1 (2022): S293-S340.
8. Acemoglu, Daron. "The simple macroeconomics of AI." *Economic Policy* 40, no. 121 (2025): 13-58.
9. Autor, David H., Frank Levy, and Richard J. Murnane. "The skill content of recent technological change: An empirical exploration." *The Quarterly Journal of Economics* 118, no. 4 (2003): 1279-1333.
10. Brynjolfsson, Erik, and Tom Mitchell. "What can machine learning do? Workforce implications." *Science* 358, no. 6370 (2017): 1530-1534.
11. Eloundou, Tyna, Sam Manning, Pamela Mishkin, and Daniel Rock. "GPTs are GPTs: Labor market impact potential of LLMs." *Science* 384, no. 6702 (2024): 1306-1308.
12. Felten, Edward, Manav Raj, and Robert Seamans. "Occupational, industry, and geographic exposure to artificial intelligence: A novel dataset and its potential uses." *Strategic Management Journal* 42, no. 12 (2021): 2195-2217.
13. Gottlieb, Joshua D., Maria Polyakova, Kevin Rinz, Hugh Shiplett, and Victoria Udalova. "The earnings and labor supply of U.S. physicians." *The Quarterly Journal of Economics* 140, no. 2 (2025): 1243-1298.

Figure 1 HCPCS code classification



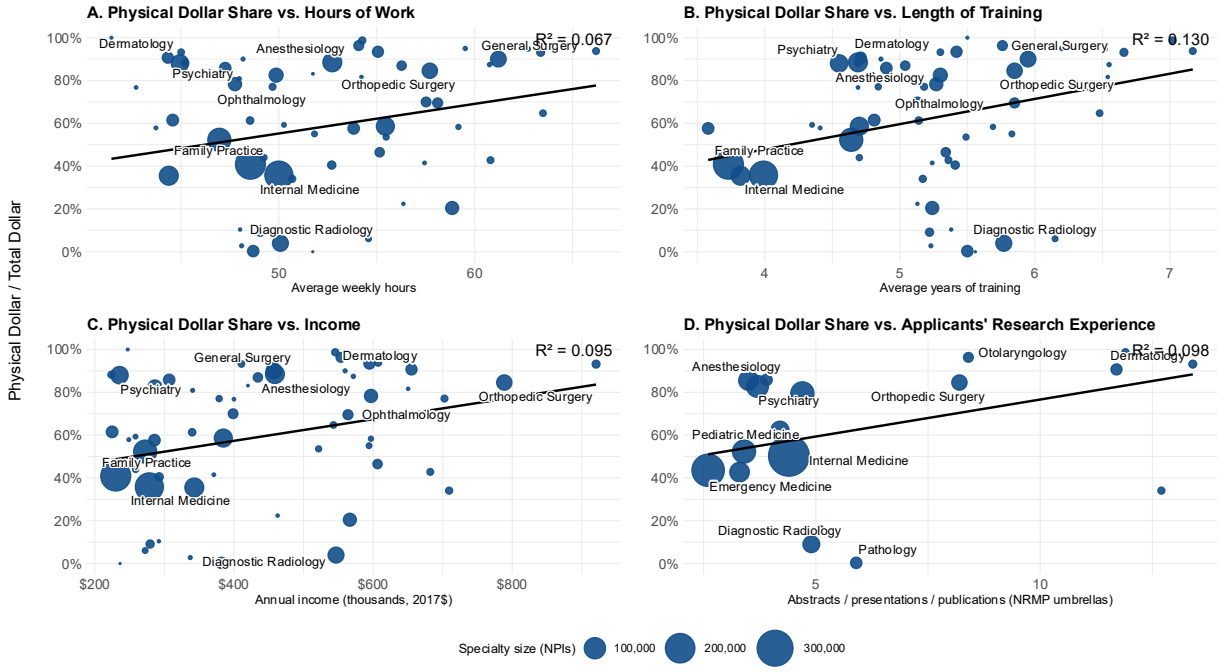
Notes: Figure shows a Sankey-style flow diagram describing how unique HCPCS codes in the study extract (N = 4,929) were classified as requiring in-person physical interaction ("physical") or not ("non-physical") using a three-step procedure. First, Evaluation and Management (E&M) codes were set aside due to heterogeneous delivery modalities and reported separately (N = 249). Second, codes were classified using a rule-based BETOS approach (see Methods for details), which classified N = 4,081 codes. Third, remaining codes (N = 599) were classified based on HCPCS code descriptions using an LLM (Anthropic Claude Opus 4.6), prompted to answer whether the task requires a direct, in-person physical interaction/exam. The final classification comprises Physical (N = 3,174) and Non-physical (N = 1,506) codes, plus E&M codes retained (N = 249). Link widths represent counts of unique HCPCS codes; node labels report code totals.

Figure 2 Specialty-level payment-weighted shares of E&M services and services that were classified as requiring physical interaction. Top 15 specialties by total payment**.**



Notes: Figure presents horizontal stacked bar charts of 2023 Medicare Part B payment-weighted service shares by specialty (N = 62); the top 15 specialties are shown individually and ordered by the share of physical services among classified (non–E&M) services. The left panel displays total payments divided into three mutually exclusive categories: Physical (blue), Non-Physical (red), and Evaluation and Management (E&M; gray), with E&M reported separately due to heterogeneous delivery modalities. The right panel restricts to non–E&M services and shows the payment-weighted share classified as Physical versus Non-Physical; percentages are displayed within bars. Total Medicare payments (billions of dollars) are reported in parentheses next to specialty names. The bottom bars summarize all 62 specialties. Payment-weighted shares were calculated by aggregating provider-service level Medicare payments to the specialty level and dividing category payments by total specialty payments.

Figure 3 Association between specialty-level payment-weighted share of services that require physical interaction and specialty characteristics



Notes: Figure shows four bubble scatterplots relating the specialty-level Medicare Part B payment-weighted share of services classified as requiring in-person physical interaction (y-axis; "Physical $ / Total $") to specialty characteristics: average weekly hours worked (Panel A), average years of training (Panel B), annual income (thousands of 2017 dollars; Panel C), and applicants' research experience (abstracts/presentations/publications from NRMP "umbrella" summaries; Panel D). Each point represents a specialty (N = 62 specialties);); bubble size is proportional to the number of actively registered physicians (NPIs) in that specialty. Solid lines show fitted linear regression trends, with panel-specific R^2 reported.