Digital Transformation in Healthcare Delivery: From Telemedicine to AI Evidence and Open Questions

Dan Zeltzer
UC Berkeley Public Health
Health Policy and Management

Finnish Health Economics Society
Helsinki, January 2026

Motivation: A Decade of Healthcare's Digital Transformation

Two shocks pushed healthcare into digital delivery:

1. **COVID** (2020-) lead to a massive surge in telehealth adoption and use.
2. **Generative AI** (2023-) provides new tools that can transform clinical workflows.

This talk:

Share evidence from several of my projects on telehealth and AI, and discuss related evidence, open questions, and opportunities.

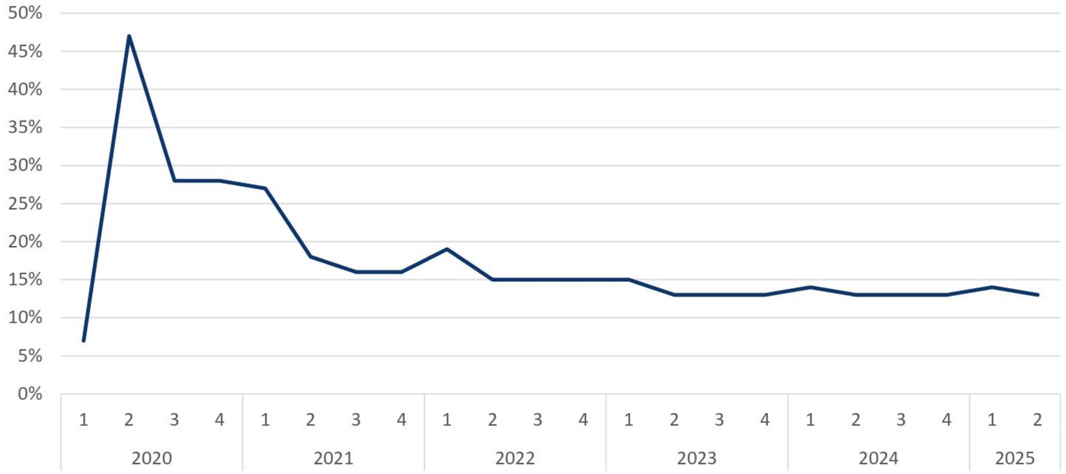
Telehealth

The pandemic compressed years of digital transformation into months

COVID-19 accelerated telehealth adoption, removing legal, reimbursement, and infrastructure barriers.

- **In 2020,** Telehealth went from <10% of consultations globally to **30-50%** almost overnight.
- **Post-pandemic,** in most high-income countries telehealth settled at **10-30%** of consultations.

Medicare users with at least one telehealth visit, %

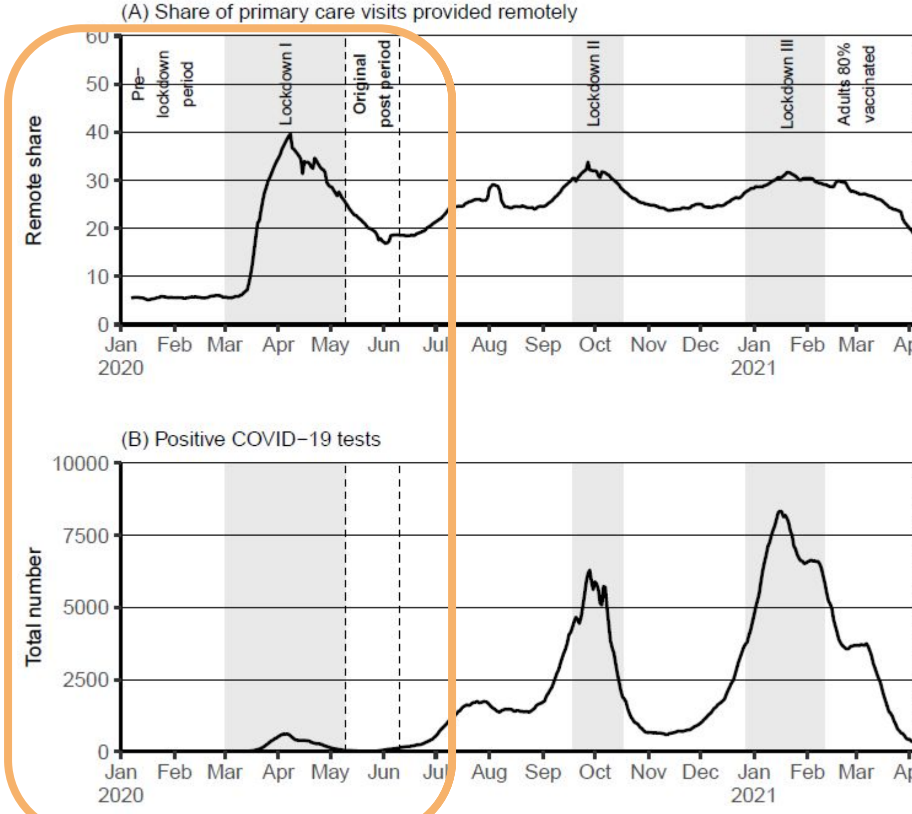


Source: CMD Medicare Telehealth Trends Report

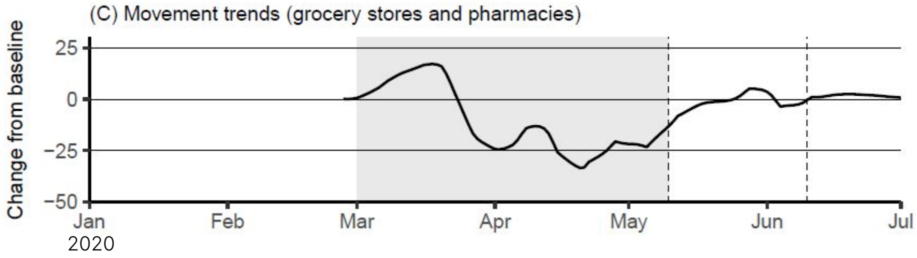
What is the impact of increased access to telemedicine?

Early 2020 in Israel provided a unique “natural experiment”:

- sharp rise in telehealth visits
- rapid return to normalcy after very tight first lockdown.



Source: Clalit Health Services, Israel Ministry of Health; au

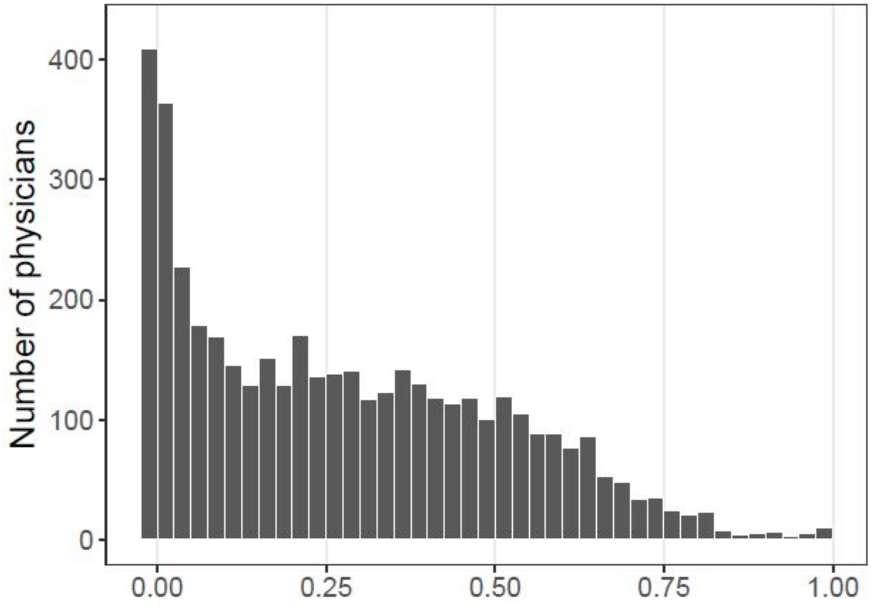


Source: Google Mobility Trends, authors’ calculations

Physician Heterogeneous Adoption

Physicians varied widely in telehealth adoption: from 0% to nearly 100% remote.

We exploit this variation as quasi-random assignment of patient access.



Physician Remote Share of Visits (March-April 2020)

The impact of increased access to telemedicine
(Zeltzer Einav et al., J Eur Econ Assoc 2023)

Does telehealth lead to duplicate visits, unnecessary tests, reduced care quality?

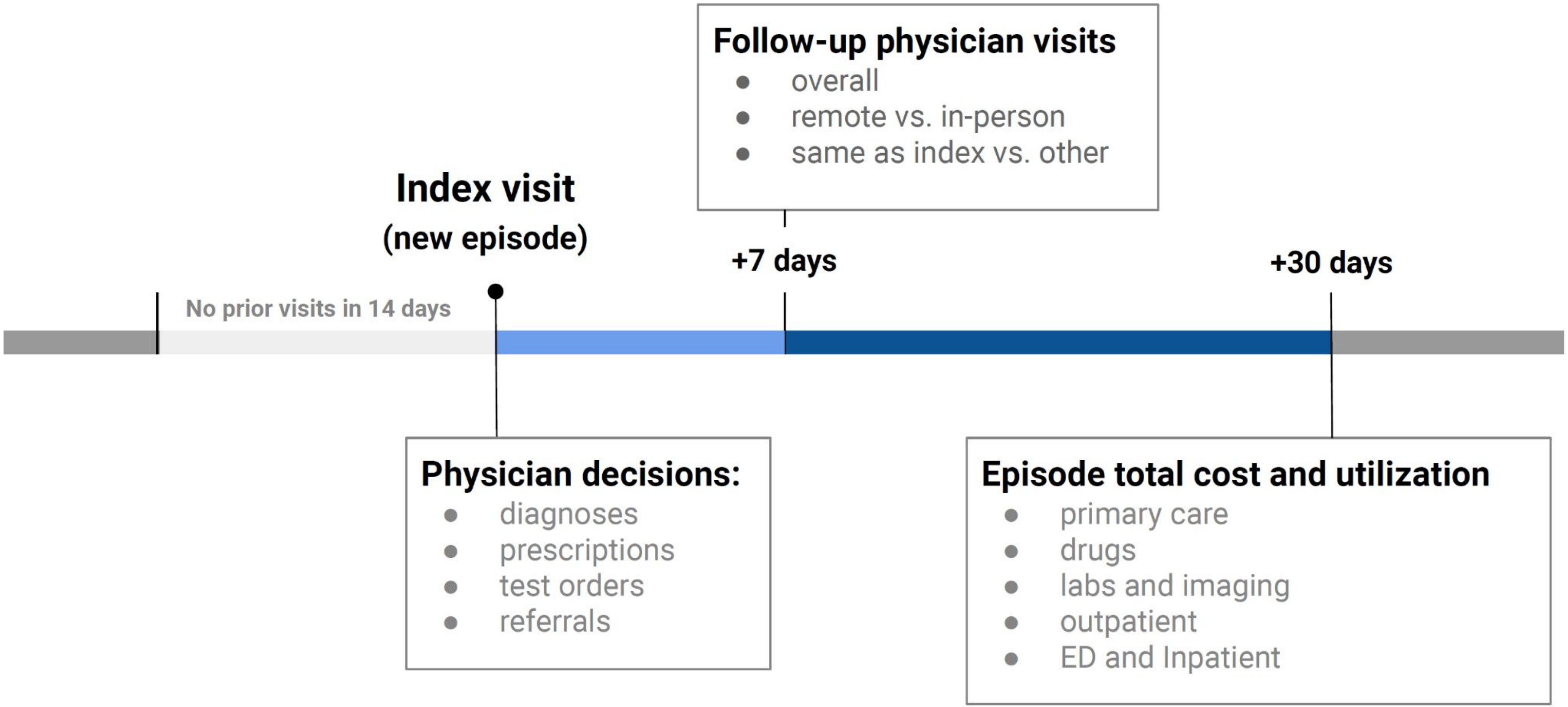
Data: 11 million primary care episodes in 2019-2020 from Israel's largest health fund covering >50% of the population.

Challenge: Modality (telehealth vs in-person) choice likely driven by patient/case characteristics, so we cannot just compare in-person and remote visits.

Empirical Approach (Diff-in-Diff):

- **Classify physicians as high/low adopters** (adjusting for time, place, and case covariates) during April-March 2020 lockdown period.
- **Compare patients of high- and low adopters, pre- and post-lockdown**
- Measure outcomes across *all* primary care visits (in-person + remote).

Unit of Observation: A Care Episode Starting with a Primary Care Visit

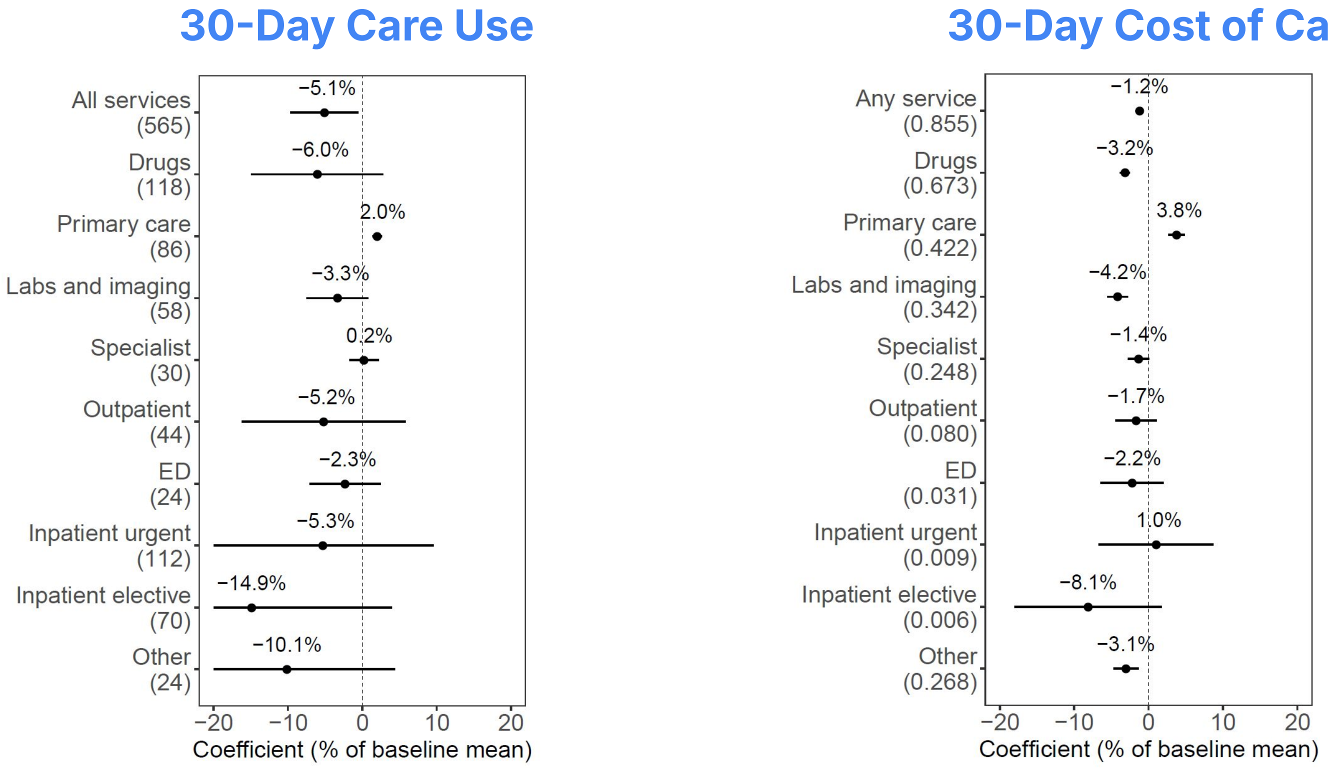


Results: The Impacts of Increased Access to Telemedicine

Telehealth access is associated with:

1. **PCP visit initiation increases by 3.5%** (relative to the pre-COVID baseline); episode cost down **6%.**
2. **Less intensive index visits**: fewer prescriptions, imaging referrals, and specialist referrals (down ~5-10%).
3. **More 7-day follow-ups (+8.2%)**, but mostly with the same physician—suggesting continuity rather than fragmentation.

Results: The Impacts of Increased Access to Telemedicine



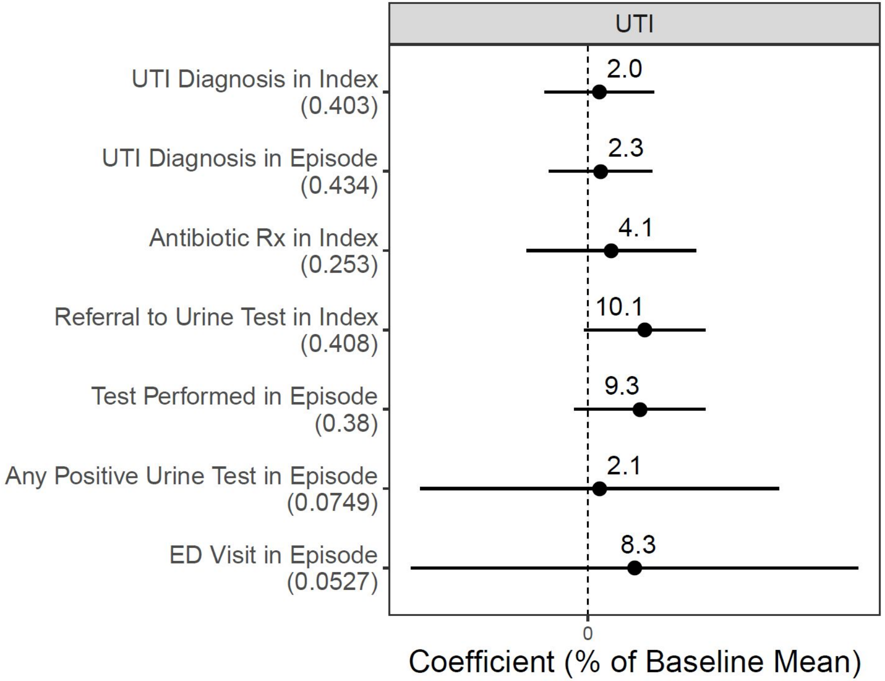
Does telehealth access impact diagnostic quality?

Approach:

- Focus on specific conditions (UTI, AMI, fractures) that are common, distinct from COVID-19, and with missed diagnoses likely observed.
- To capture diagnostic accuracy, analyzed episodes with target condition + all related presentations.

Result: No change in diagnostic accuracy or follow-up.

Episodes with Urinary Conditions



The Impact of Increased Access to Telehealth

Increased access to telehealth results in:

- A small increase in PCP episode initiation, but decreases in episode-level and total costs
- More follow-ups, mostly as remote check-ins with the same provider (suggesting convenience/ongoing management rather than fragmentation)
- No evidence of missed diagnoses or adverse outcomes for specific conditions

Takeaway:

- Telehealth can productively complement in-person care.
- Similar evidence appears in other contexts (e.g., Nakamoto et al., 2024 for Medicare).

From Telehealth to Devices

Telehealth expansion spurred growth in complementary device markets (>$100 Billion in 2024)

- Home diagnostics
- Wearable medical devices
- Remote Patient Monitoring (RPM) for cardiology, diabetes, respiratory conditions

Medicare gradually expanded RPM coverage; multiple hospital-at-home (HaH) programs operate across most US states.

Adoption and Utilization of Device-Assisted Telemedicine
(Zeltzer Einav et al., J of Health Econ 2023)

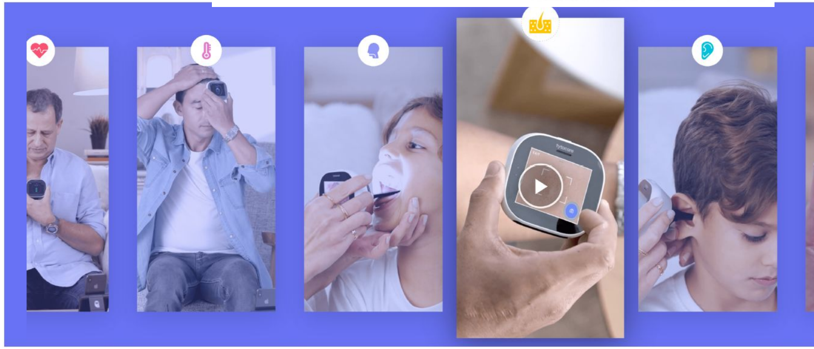
Context: bluetooth-connected home diagnostic device to support remote diagnosis of minor acute conditions.

Aim: estimate impact on care use

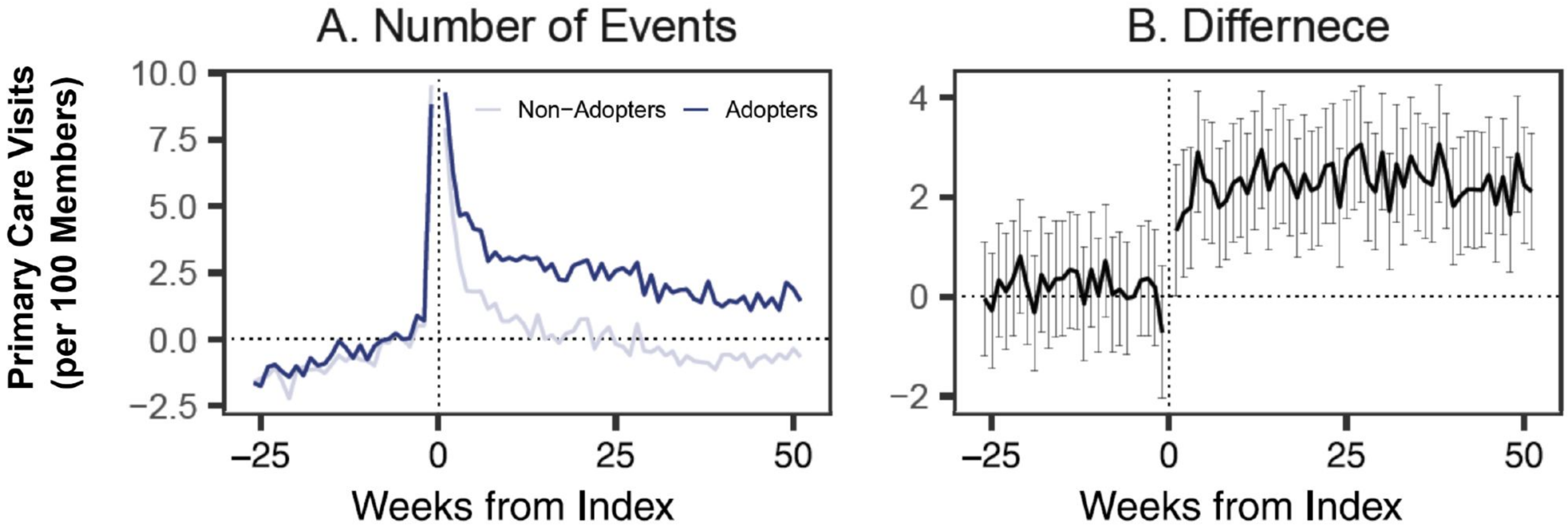
Approach:

- Define adoption as first-ever device-assisted telehealth visit
- Match adopters and never-adopters who visited the *same* virtual clinic

Sample: 30k adopters, 2019-2021; same number of non-adopters.

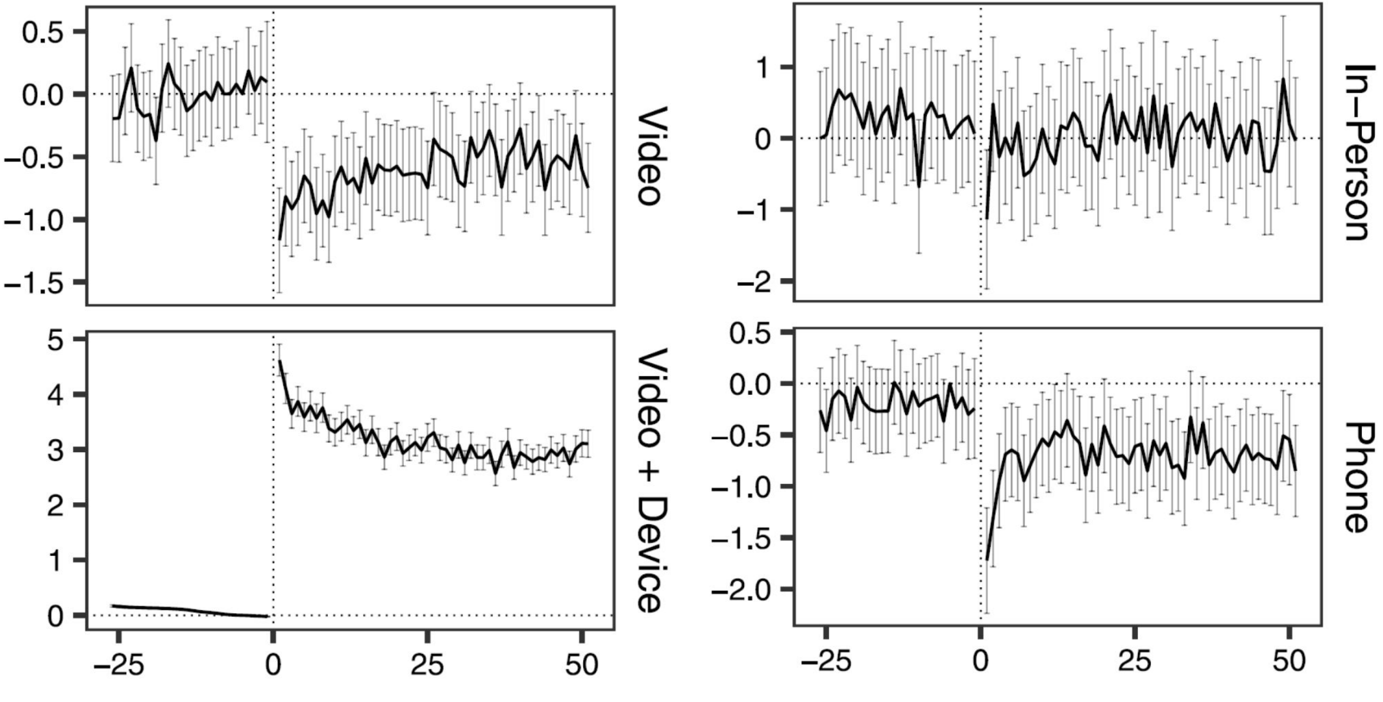


Device adoption led to increase in primary care use

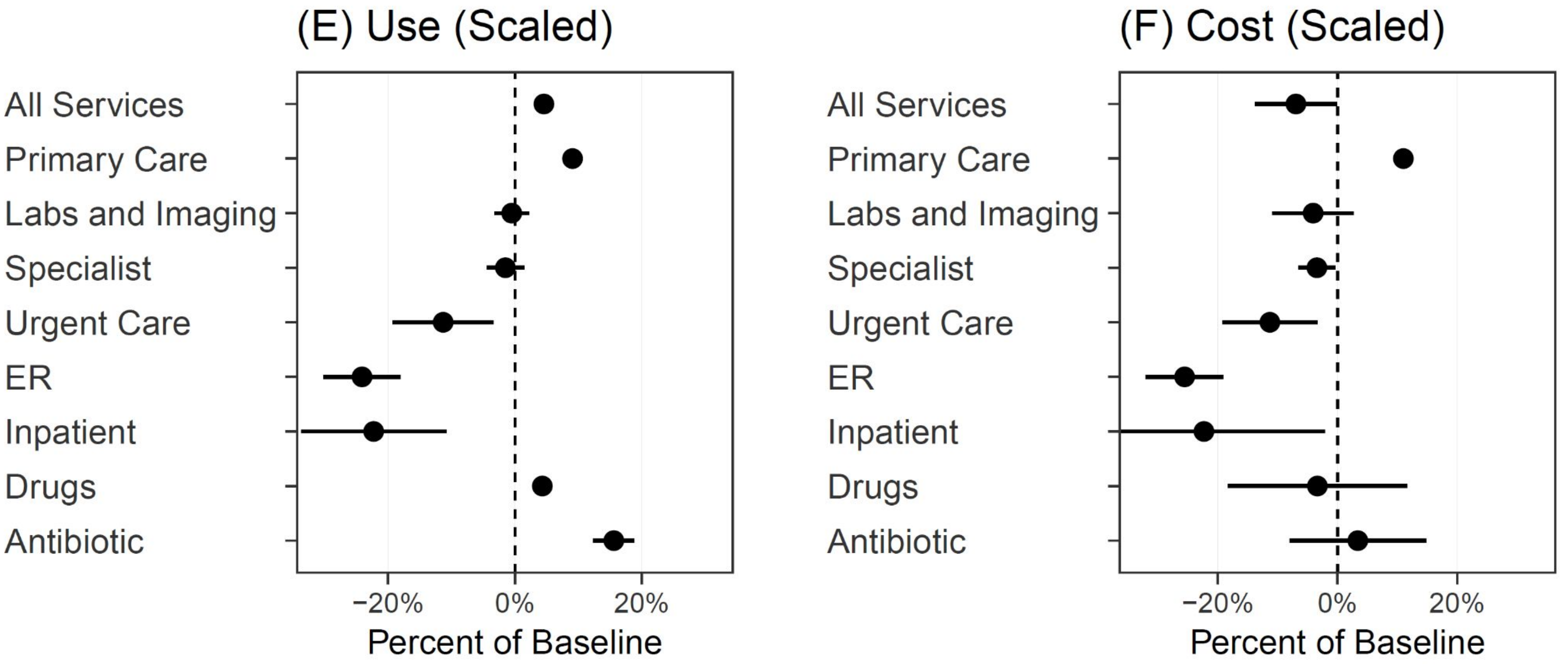


Device-assisted telehealth visits mainly substituted for other telehealth modalities

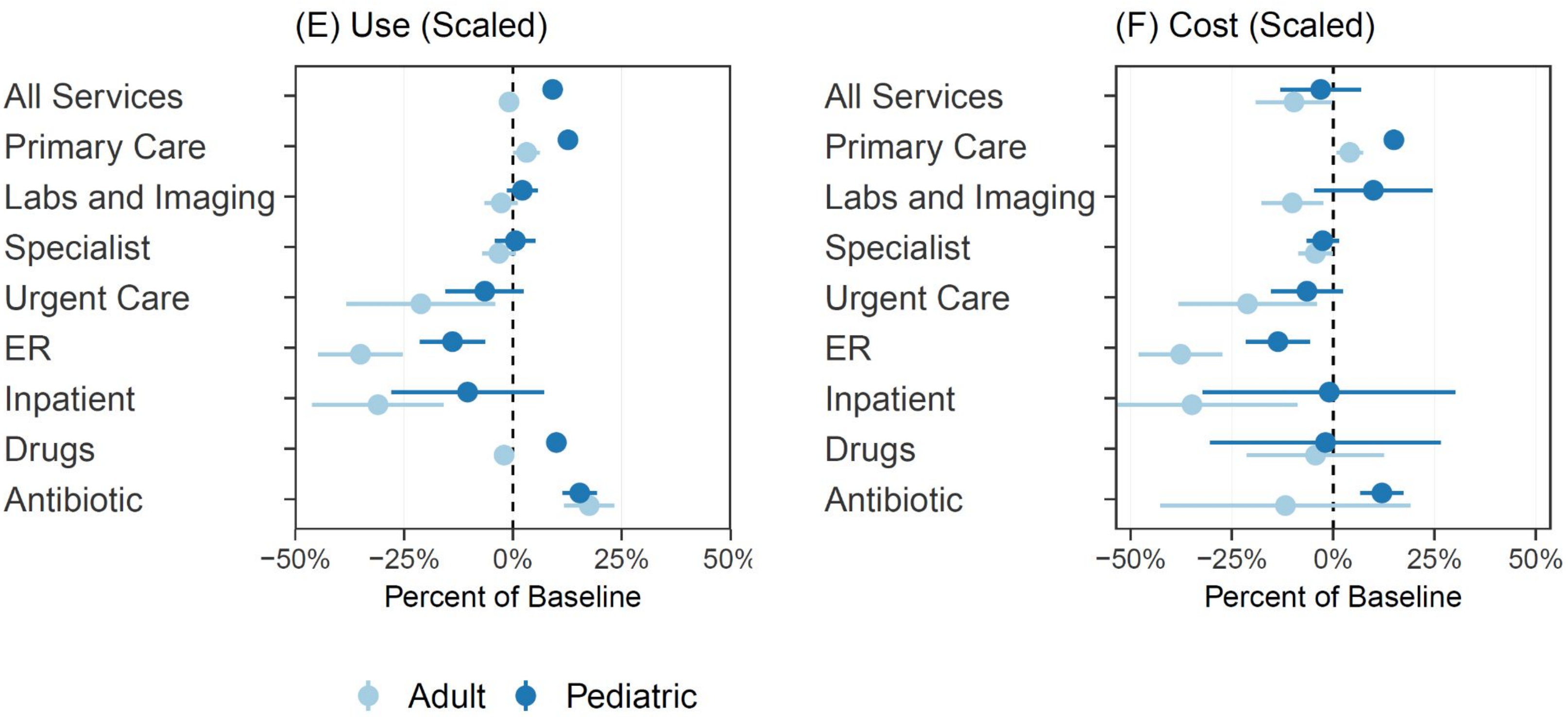
Weekly visits per 100 patients



Offsets of more intensive modalities, but more antibiotics



Larger impacts on adult patients



Telehealth: Challenges and Opportunities

- The evidence suggests telehealth will remain a fixture of healthcare delivery. certain specialties (e.g., psychiatry) are becoming primarily remote
- **The new outpatient care is "multimodal":** audio/video visits, RPM, e-visits, ...

Challenges:

- **Payments/Incentives:** How to reimburse providers to induce proper use of many modalities?
- **Ensuring equity:** How to ensure proper access for everyone?

Opportunities:

- **Flexibility and convenience for patients and providers:** UK NHS estimated savings of 530 years of travel/wait time, and about £40 million in travel costs in 2020-2021
- **Expanded reach:** E-curbsides; Cross-borders.
- Improved matching and **allocative efficiency.**
- Shifting from intermittent to **continuous care.**

Artificial Intelligence (AI)

Increasing Scope for Using Algorithms in Diagnosis

- **Radiology:** Algorithms can interpret mammograms as well as humans (McKinney et al., *Nature* 2020)
- **Dermatology:** Algorithms can accurately classify common skin cancers from images (Esteva et al., *Nature* 2017).
- **Pathology:** Algorithms can identify cancer from biopsy and surgical resection slides (Xu et al., *Nature* 2024).
- **Emergency Medicine:** Compared to benchmark predictions, emergency physicians overtest low-risk patients and undertest high-risk patients, leading to adverse outcomes (Mullainathan Obermeyer, *QJE 2022*).
- **Internal Medicine:** AI models exhibit high diagnostic accuracy compared with human physicians (Google AMIE, Nature 2025)

Algorithms for Clinical Assistance?

Key Questions:

- Can algorithms perform in real-world primary care?
- Where do they hold advantages over physicians?

Evidence Gaps:

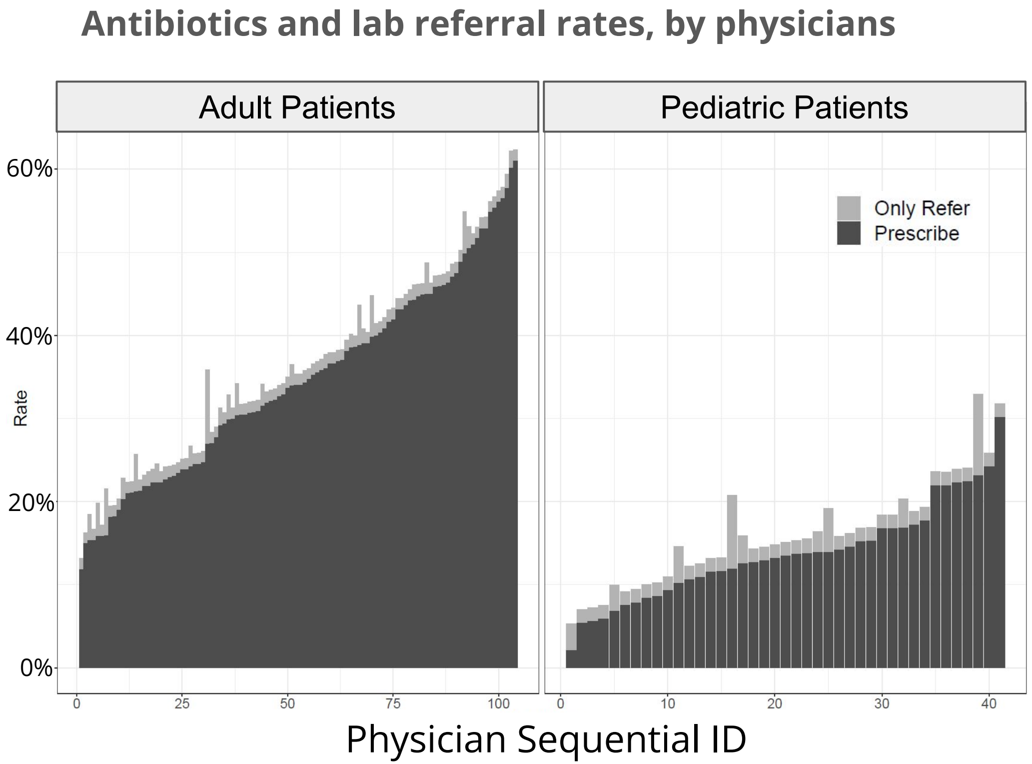
- Much evidence comes from retrospective, curated datasets.
- In primary care, patients do not present as neatly defined case descriptions, and intake is an essential part of the process.

Physician Decisions Are Noisy

- **Practice variation is well documented across settings and outcomes** (Wennberg & Gittelsohn, Science 1973; Chandra & Staiger, JPE 2007; Cutler et al., AER 2013; Molitor, AEJ: Applied 2018; Card et al., AEJP 2023; Currie et al., JEP, forthcoming).
- Evidence from quasi-random designs show **a lot of the variation is unwarranted** (i.e., not driven by patient need or preference).

Between-Physician Variability in Antibiotics Prescription

- Antibiotic prescribing varies widely across physicians.
- Patients are routed to the first-available clinician, so this reflects practice style, not patient differences.
- Clear scope for improving consistency.
- Can AI help?

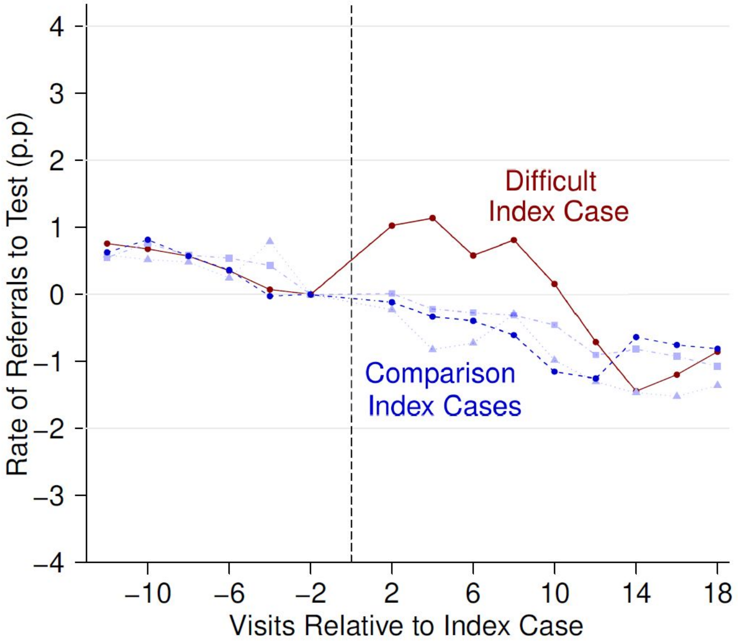


Source: Joseph Rashba (M.A. thesis, 2023). Sample of upper respiratory virtual clinic visits where physicians are assigned on a first-available basis.

Within-Provider Variability in Diagnostic Testing
(Chodick et al., Rev. Econ. Stat. 2024)

Within-physician variation in testing:
Comparing visits around a PCP's "difficult" encounter (new cancer diagnosis) to matched weeks/years for the *same* physician, **testing rises ~5% for subsequent unrelated patients**.
The increase is only for ~1 hour, consistent with attention/emotion rather than learning.

Lab and imaging referral rates around difficult cases (compared with other days of same physician)



Source: Chodick et al. 1.6 million in-person primary care visits with 747 physicians in Israel, 2012–2015.

Broader Evidence for Variability in Physician Decision-Making

Between Providers:

- C-section rates vary significantly across physicians and hospitals, impacting maternal and infant outcomes (Currie & MacLeod, *JOLE* 2017).
- Prescribing practices vary with diagnostic skill (Currie & MacLeod, *ECMA* 2020).
- Radiologists differ in diagnostic skill, beyond preferences (Chan et al., *QJE* 2022).

Within Providers:

- C-section decisions are influenced by complications in the *prior patient*'s delivery (Singh, *Science* 2021); Military physicians exert more effort for higher-ranked patients (Schwab and Singh, *Science* 2023).
- Autocorrelation (Jin et al. 2024) and cognitive fatigue (Kolstad et al, in progress) in ER decisions.

Variation matters because it drives outcomes, costs, and equity

ORIGINAL ARTICLE

MAYO CLINIC PROCEEDINGS: DIGITAL HEALTH

MAYO CLINIC

Diagnostic Accuracy of Artificial Intelligence in Virtual Primary Care

Dan Zeltzer, PhD; Lee Herzog, MD; Yishai Pickman, PhD; Yael Steuerman, PhD; Ran Ilan Ber, PhD; Zehavi Kugler, MD; Ran Shaul, BASc; and Jon O. Ebbert, MD MSc

Abstract

Objective: To evaluate the diagnostic accuracy of artificial intelligence (AI)-generated clinical diagnoses.
Patients and Methods: A retrospective chart review of 102,059 virtual primary care clinical encounters from October 1, 2022, to January 31, 2023 was conducted. Patients underwent an AI medical interview, after which virtual care providers reviewed the interview summary and AI-provided differential diagnoses, communicated with patients, and finalized diagnoses and treatment plans. Our accuracy measures were agreement between AI diagnoses, virtual care providers, and blind adjudicators. We analyzed AI diagnostic agreement across different diagnoses, presenting symptoms, patient demographic characteristics such as race, and provider levels of experience. We also evaluated model performance improvement with retraining.

Diagnostic Accuracy in Virtual Primary Care

Zeltzer et al. (Mayo Clinic Proc Digit Health, 2025)

Design: Retrospective chart review of virtual primary care encounters

Sample: 102,059 single-diagnosis visits Oct 2023-Jan 2024 with 133 primary care providers. Mostly younger adults.

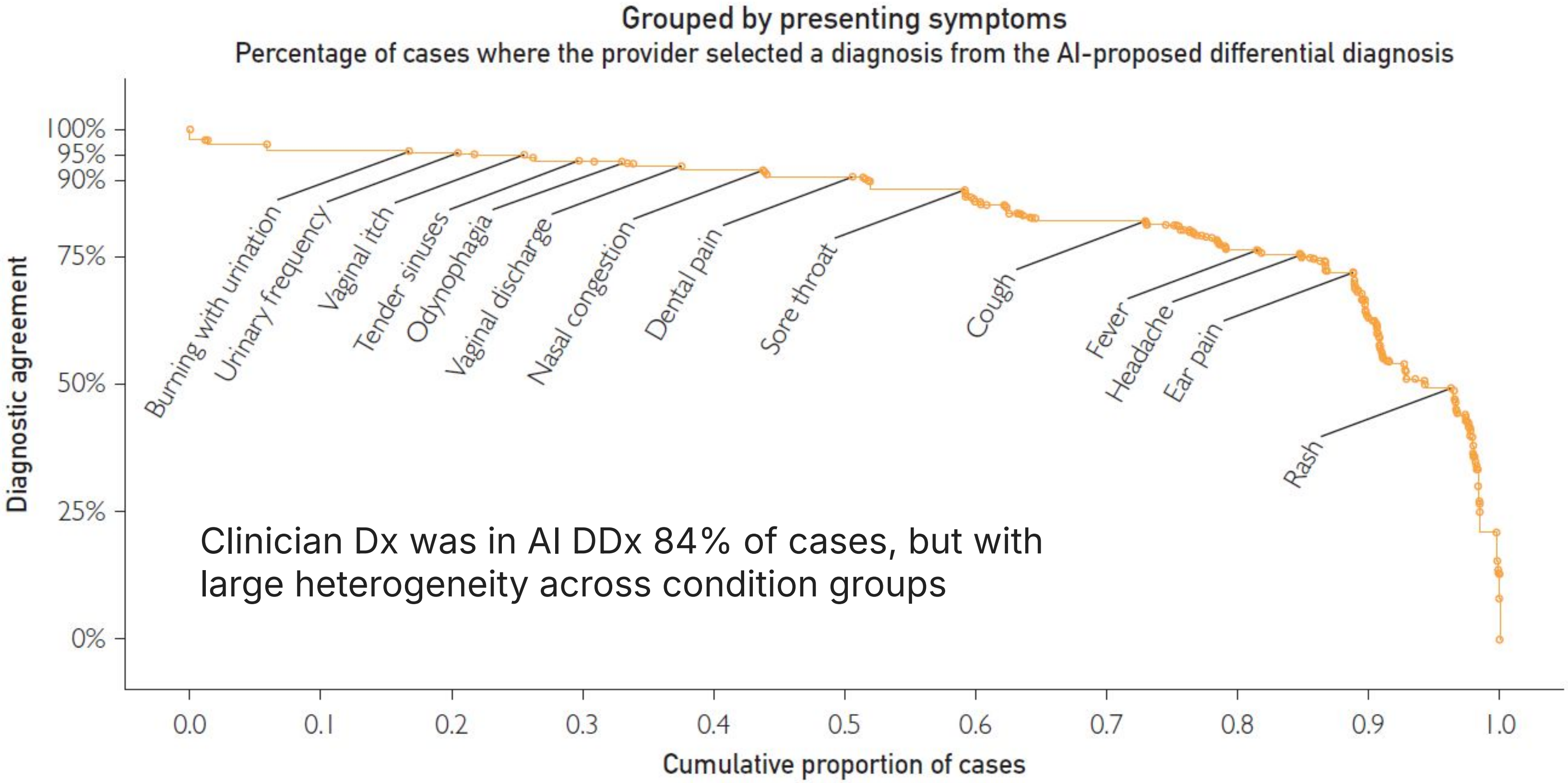
AI model: intake and diagnosis (ensemble of ML models trained on EMR data)

- Chat collects patient demographics, history, symptoms
- The model then generates intake summary + differential diagnosis (≤5 dx; median 2)

Clinician (Physician or Nurse): conducts text/video consult and makes all final decisions.

Main outcome: Share of cases where the provider diagnosis was within the model differential.

Diagnostic Accuracy in Virtual Primary Care



Clinician Dx was in AI DDx 84% of cases, but with large heterogeneity across condition groups

Annals of Internal Medicine ORIGINAL RESEARCH

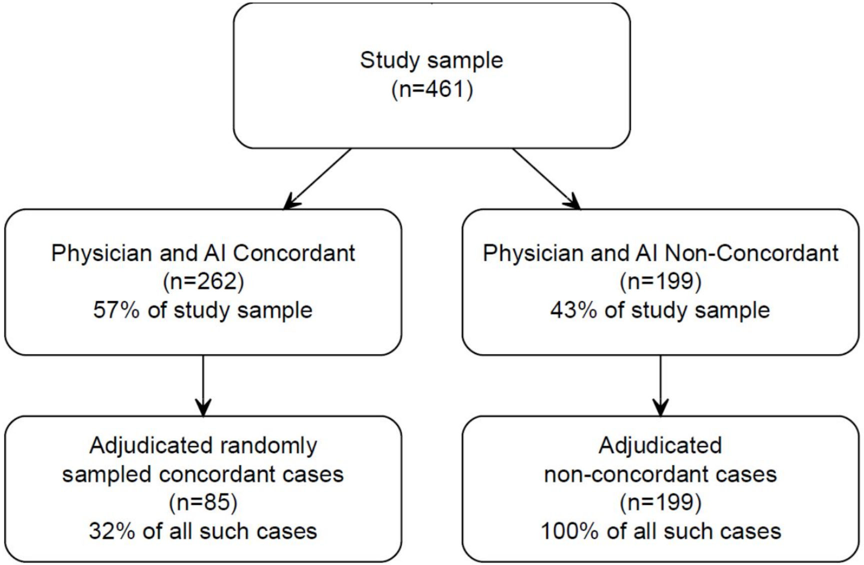
Comparison of Initial Artificial Intelligence (AI) and Final Physician Recommendations in AI-Assisted Virtual Urgent Care Visits

Dan Zeltzer, PhD; Zehavi Kugler, MD; Lior Hayat, MD; Tamar Brufman, MD; Ran Ilan Ber, PhD; Keren Leibovich, PhD; Tom Beer, MSc; Ilan Frank, MSc; Ran Shaul, BASc; Caroline Goldzweig, MD, MSHS; and Joshua Pevnick, MD, MSHS

- **Setting:** Virtual urgent care clinic based in Los Angeles (Cedars Sinai)
- **Sample:** 461 visits with for adults with respiratory, urinary, vaginal, eye, or dental symptoms from June-July 2024. Seen by 18 physicians (avg. visit 18 minutes)
- **AI input:**
  - AI conducts preliminary intake + extracts information from patient records.
  - Recommends diagnosis and care management (prescriptions, labs, referrals).
  - Recommendations suspended if AI confidence is low (20% of cases)
- **Physicians** conduct a video visit and make decisions

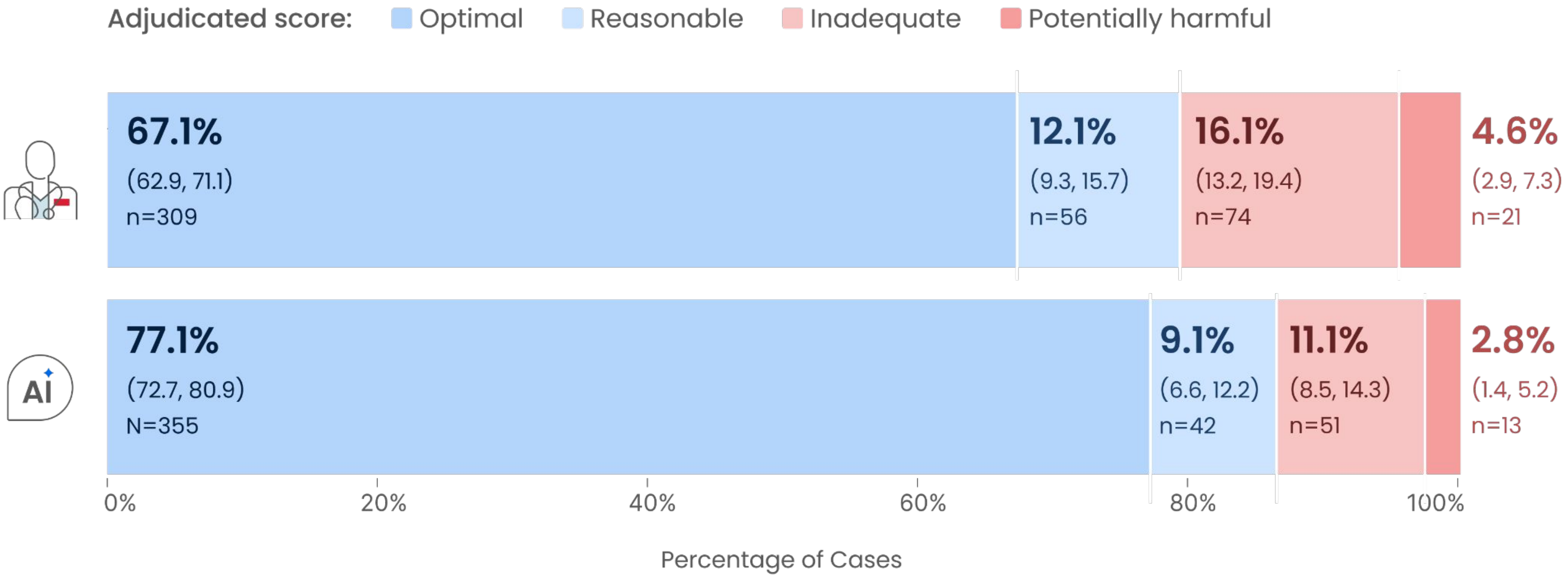
Empirical Approach: Expert Adjudication

- **Two physicians reviewed each visit** (intake, AI output, transcript, clinician decision) and rated **AI** and **physician** on a **4-point quality scale:** optimal to potentially harmful.
- **If expert reviewer disagreed** a third physician joined to discuss; consensus was encouraged, but not required.
- **Key limitations: single-center, retrospective**; reviewers **not blinded** to source; unclear whether clinicians **saw AI recommendations**.



Summary of Physician and AI Scores

Reweighted sample of n=461 cases. 95% CI in parentheses.



Summary AI vs. Physician Scores

Equal 68%, AI better 21%, Physician better 11%

Cross-Classification of Adjudicated Physician Versus AI Scores:
Weighted Sample of ñ = 461 Cases

A

AI (95% CI), %	Physician (95% CI), %: Optimal	Reasonable	Inadequate	Potentially Harmful
Optimal	58.3 (54.3 to 62.3) ñ = 269	5.4 (3.8 to 7.6) ñ = 25	10.8 (8.6 to 13.6) ñ = 50	2.5 (1.5 to 4.2) ñ = 12
Reasonable	2.7 (1.6 to 4.4) ñ = 12	4.5 (2.7 to 7.3) ñ = 21	1.7 (0.9 to 3.1) ñ = 8	0.2 (0.0 to 0.7) ñ = 1
Inadequate	5.6 (3.9 to 7.8) ñ = 26	2.1 (1.2 to 3.6) ñ = 10	3.2 (1.8 to 5.8) ñ = 15	0.1 (0.0 to 1.0) ñ = 1
Potentially Harmful	0.5 (0.2 to 1.7) ñ = 3	0.1 (0.0 to 0.8) ñ = 1	0.3 (0.1 to 1.4) ñ = 1	1.8 (0.8 to 4.4) ñ = 8

Relative Score

AI better (ñ = 96)
20.8% (17.8% to 24.0%)

Equal score (ñ = 313)
67.9% (64.8% to 70.9%)

Physician better (ñ = 52)
11.3% (9.0% to 14.2%)

Adjudicator Reasons for Difference in Scores

Sample: 111 non-concordant cases with AI and physician rated differently, and at least one rated inadequate or potentially harmful

AI Rated Better (64%):

Guideline adherence:

- **Better lab use (22.8%):** Recommended necessary tests, avoided empiric treatments.
- **More guideline adherent (16.3%)**: Avoided unnecessary antibiotics/corticosteroids.

Risk identification:

- **Effective referrals (15.2%)**: Identified need for urgent evaluations in worsening cases.
- **Correctly identified red flags (4.4%)**: Highlighted symptoms suggesting elevated risk that providers missed.

Physicians Rated Better (36%):

Adaptability to shifting information and complexity:

- **Adapted to new information** (6.2%)
- **Handled complex cases** (4.4%)
- **Accounted for history** (3.1%)

Referrals and decision-making:

- **Avoided unnecessary ER referrals (8.0%)**: Did not refer mild cases (e.g., healthy COVID-19).
- **Made appropriate referrals (5.9%)**

Summary: AI vs physicians in virtual urgent care

In high AI-confidence visits (80%; AI withholds recommendations in 20%):

- **AI and physicians were fully concordant 57% of the time.**
- **When they disagreed, adjudicators rated AI equal or better in ~89% of cases** (68% equal; 21% AI better).
    - **AI strengths:** guideline-adherent meds/tests/referrals + stronger detection of red flags.
    - **Physician strengths:** adapting to new/inconsistent info during the visit and other nuanced judgment (e.g., avoiding unnecessary ED referrals).

Findings suggest **AI may support physicians in decision-making, especially for common cases** (which account for most of the clinic volume).

Potential to help reduce unwarranted variation in care.

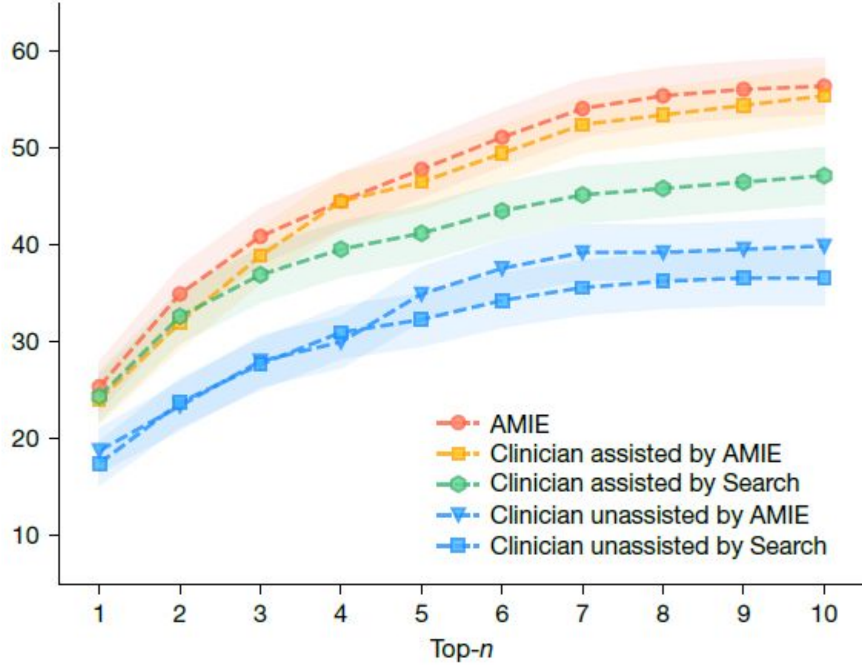
Integrating Humans and Algorithms

AI Diagnostic Assistance

Tao et al., Nature (2025)

- Clinician provided differential diagnosis using patient histories from 300 (hard) NEJM cases.
- Compared medical-LLM assistance (Google AMIE) with internet search
- Specialists rated the differential on quality and completeness.
- Result: LLM assistance improved physician differential accuracy.

Percentage of DDx including the actual diagnosis



Setting expectations for AI's near-term role

Work in progress

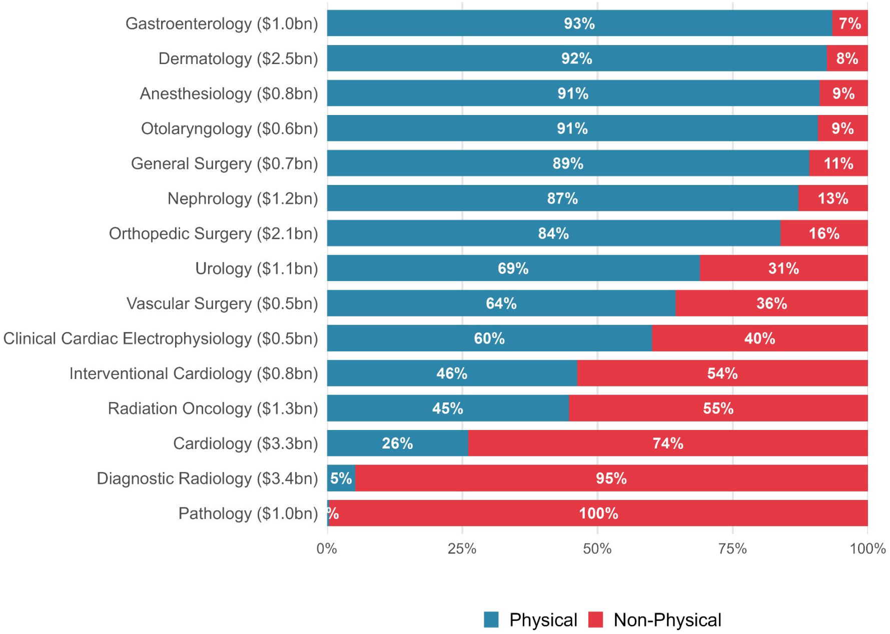
Question: What share of physician services could AI affect in the near term?

Approach: We classified physician services by physical interaction requirement as an upper bound on AI automation.

Result: ~56% of billed services (by $) require physical contact overall, but ranges from nearly 0% (pathology, radiology) to 90%+ (surgery, anesthesiology).

Share of payments for classified services by physical interaction

Top 15 specialties by dollar volume (excluding high-unclassified specialties)



How is AI Used in Healthcare Today?

- **Documentation** (e.g., Abridge, Ambience AI scribes): Massive deployed (150+ systems, 50M conversations in 2025, reduces after-hours documentation); creates novel data layer beyond EHR/claims
- **Evidence Retrieval** (e.g., OpenEvidence): Over half of US providers reported using AI for real-time access to medical literature.
- **Risk Prediction & Stratification:** hospitals use integrated predictive AI and AI for imaging/radiology to support personalized care, targeting high-risk patients.
- **Chronic Care Management:** RPM with AI analytics
- **Clinical Copilots** for preliminary triage, routing, and diagnostic assistance

Sources: JAMIA 2025 (Poon et al.); Abridge/Menlo Ventures 2025; HealthIT.gov; NEJM AI 2025; Talencio 2024

AI in Healthcare: Opportunities and Challenges

Challenges and open questions

- **Implementation and safety:** Understanding when AI helps vs. harms; preventing over-reliance
- **Equity:** Bridging access gaps and addressing subgroup biases
- **Incentives and governance:** payment models; liability allocation; regulatory standards for adaptive systems
- **Workforce and workflow:** Where does human judgment add most value? What new roles are needed?

Opportunities

- **AI Triage & Async Care**: Patient-facing AI providing asynchronous care with escalation to clinicians only when needed
- **Continuous chronic care:** RPM, wearables, and AI alerts for proactive chronic care management.
- **Population health at scale:** risk stratification and targeted follow-ups
- **Healthcare operations:** Administrative task automation (e.g., coding, preauth)

Advantage for healthcare systems with unified digital infrastructure!

Thank you!

Bibliography

Telehealth Studies

Zeltzer, D., Einav, L., Rashba, J., & Balicer, R. D. (2024). The Impact of Increased Access to Telemedicine. *Journal of the European Economic Association*, 22(2), 712-750. https://doi.org/10.1093/jeea/jvad035

Zeltzer, D., Einav, L., Rashba, J., Waisman, Y., Haimi, M., & Balicer, R. D. (2023). Adoption and utilization of device-assisted telemedicine. *Journal of Health Economics*, 90, 102780. https://doi.org/10.1016/j.jhealeco.2023.102780

Nakamoto, C. H., Cutler, D. M., Beaulieu, N. D., Uscher-Pines, L., & Mehrotra, A. (2024). The Impact of Telemedicine on Medicare Utilization, Spending, and Quality, 2019-22. *Health Affairs*, 43(5). https://doi.org/10.1377/hlthaff.2023.01142

Physician Variability and Diagnostic Decision-Making

Chodick, G., Goldstein, Y., Shurtz, I., & Zeltzer, D. (2025). Challenging encounters and within-physician practice variability. *Review of Economics and Statistics*, 1-9.

Wennberg, J. E., & Gittelsohn, A. (1973). Small area variations in health care delivery. *Science*, 182(4117), 1102-1108.

Chandra, A., & Staiger, D. O. (2007). Productivity spillovers in healthcare: Evidence from the treatment of heart attacks. *Journal of Political Economy*, 115(1), 103-140.

Cutler, D., Skinner, J. S., Stern, A. D., & Wennberg, D. (2013). Physician beliefs and patient preferences: A new look at regional variation in health care spending. *American Economic Review*, 103(3), 100-104.

Molitor, D. (2018). The evolution of physician practice styles: Evidence from cardiologist migration. *American Economic Journal: Applied*