Volume Title: The Economics of Artificial Intelligence: Health Care Challenges

Volume Authors/Editors: Ajay Agrawal, Joshua Gans, Avi Goldfarb, and Catherine Tucker, editors

Volume Publisher: University of Chicago Press

Volume ISBNs: 978-0-226-83311-8 (cloth); 978-0-226-83312-5 (electronic)

Volume URL: https://www.nber.org/books-and-chapters/economics-artificial-intelligence-health-care-challenges

Conference Date: September 22–23, 2022

Publication Date: March 2024

Chapter Title: Artificial Intelligence, the Evolution of the Health Care Value Chain, and the Future of the Physician

Chapter Author(s): David Dranove, Craig Garthwaite

Chapter URL: https://www.nber.org/books-and-chapters/economics-artificial-intelligence-health-care-challenges/artificial-intelligence-evolution-health-care-value-chain-and-future-physician

Chapter pages in book: p. 9 – 45

1

Artificial Intelligence, the Evolution of the Health Care Value Chain, and the Future of the Physician

David Dranove and Craig Garthwaite

1.1 Introduction

It is often said that the most expensive medical "technology" is the physician's pen.[1] While this is an obviously apocryphal statement, it is rooted in the fundamental centrality of physicians to the health care economy. In his foundational book, Fuchs (1974) characterized the physician as the "captain of the team," i.e., the economic actor that directs the application of medical technology and as a result serves as the primary determinant of medical spending. Not much has changed in the fifty years since Fuchs advanced this argument—physicians still dominate medical decision making—except the team has gotten larger and much more expensive.

As the "captain of the team," physicians diagnose illnesses, recommend, and perform treatments. As described by Arrow (1963), patients trust their physicians to make the correct choices about their treatments and physicians continue to earn high scores in trust, especially when compared with other occupations (Gallup 2022). Yet physicians are fallible, often misdiagnosing

David Dranove is the Walter McNerney Distinguished Professor of Health Industry Management at Northwestern University's Kellogg School of Management, where he is also professor of strategy.

Craig Garthwaite is the Herman R. Smith Research Professor in Hospital and Health Services, a professor of strategy, and the director of the Program on Healthcare at Kellogg (HCAK) at Kellogg School of Management, Northwestern University, and a research associate of the National Bureau of Economic Research.

We thank Bingxiao Wu for her assistance in assembling tables 1.1 and 1.2. For acknowledgments, sources of research support, and disclosure of the authors' material financial relationships, if any, please see https://www.nber.org/books-and-chapters/economics-artificial-intelligence-health-care-challenges/artificial-intelligence-evolution-healthcare-value-chain-and-future-physician

1. See https://www.hcinnovationgroup.com/home/blog/13018116/whats-the-most-expensive-technology-the-doctors-pen.

cases and making the wrong treatment recommendations. Recent research has demonstrated that this can involve both undertreating those who are quite ill and overtreating those who are largely healthy (Mullainathan and Obermeyer 2021). The result is the undesirable combination of higher costs and increased rates of preventable death, injury, and illness.

Given the inherent fallibility of physicians, over the past several decades the medical community, payers, and regulators have experimented with incentive and provided physicians with information about best practices in an attempt to influence and improve medical decision making. A large research literature suggests that these efforts have had, at best, mixed results.[2]

Advances in data collection and analytic methods enabling the development of artificial intelligence (AI) offer new and unprecedented opportunities to improve medical decision making. Across a variety of cases, AI has shown the potential to reduce false positive and false negative rates of diagnosis. AI can also provide more appropriate treatment recommendations, often tailoring them to highly specific sets of symptoms and patient characteristics that could be difficult for every human medical provider to accurately diagnose. Finally, AI has the potential to overcome some inherent biases of various actors in the system, although this may be a matter of replacing the biases of physicians with the biases of data analysts and those who direct their work (Obermeyer et al. 2019).

To better understand the potential implications of AI in health care, we rely on the economic intuition established by Autor, Levy, and Murnane (2003), which examines the effect of greater automation on the distribution of wages and tasks across workers of different skill types. While Autor, Levy, and Murnane primarily consider the impact of robots doing relatively routine work, technological progress in AI means automation can increasingly accomplish some of the nonroutine tasks that were thought to be the solely the domain of human workers. They may even be able to accomplish some of these nonroutine tasks *better* than humans can.

But it is also apparent that other tasks routinely performed by physicians and other medical personnel remain well beyond the reach of even the most optimistic proponents of AI. As suggested by Autor (2022), the impact of a greater use of AI in medicine will therefore depend on the degree to which the routine and nonroutine tasks AI takes on complement or substitute for tasks that must still be performed by medical providers. As we discuss below, this impact will likely differ by specialty, skill set, and the patient's medical condition—with potentially wide-ranging ramifications for the medical profession.

Broadly speaking, physicians are responsible for both gathering information from patients and using that information to diagnose conditions

2. For example, see Wickizer and Lessler (2002) and Eijkenaar et al. (2013), which review the literature on utilization review and pay for performance, respectively.

and develop treatment protocols. To the extent the information gathered is purely physical (e.g., a blood sample, image, or tissue sample) and the diagnosis can be automated, physicians could conceivably lose their centrality in the role of "captain" and be replaced by medical providers serving as technicians carrying out the decisions made by third-party AI algorithms. This could result in physicians ceding much of their practices to lower-cost allied medical professionals, such as nurse practitioners working in retail clinics. In that case, value may be captured by the owners and implementors of the AI systems or the clinics or patients.

To the extent the information must still be gathered by interactions between human medical providers and patients, the ability to more accurately use that information to make a diagnosis and develop a treatment plan is a complement to a provider's effort. Providers who are better able to gather data from patients or utilize the additional information from AI may capture much of the valued created. However, that provider may not be the same type of doctor that currently completes those tasks—allied medical personnel may be equally (or more) capable of incorporating AI information into medical decisions. Therefore, AI's impact will be dictated by the set of tasks that currently comprise a physician's role in the system—which we demonstrate below varies meaningfully across specialties.

In the economic literature, the creation and adoption of AI is often either explicitly or implicitly modeled as an exogenous event emerging as a result of broader technological progress. In health care, however, there are a number of barriers to the success of AI that are specific to the sector and will likely influence the eventual existence and nature of automation. At a minimum, there are both institutional and legal barriers to assembling large data sets containing information about patients. Perhaps more importantly, the success of AI may depend on buy-in from the very individuals whose success it threatens—physicians. Accordingly, we discuss below how the predictions of ALM on the impact of AI on physicians are likely to also shape the types of AI that emerge and are adopted by the medical community.

In addition, we consider the role of government and the competitive environment in determining which types of technologies emerge. Absent some sort of standard-setting body, it is unclear how technologies will be adopted. This is particularly true if small differences emerge in the accuracy of these technologies and if the legal environment for liability is unclear.

We begin by describing important features of the health care market that are central to understanding the economic implications of AI. These include both the medical decision-making process and the history of third-party interventions. We next examine the implications of the labor economics literature on the distribution of economic surplus in the health care value chain and demonstrate the important role of different tasks in understanding this prediction. We conjecture about how AI will shape the future of physicians and allied personnel. We then take the perspective of AI firms—will they capture the value they create, and will they be forced to adapt to the potential backlash from physicians? We close with a discussion of additional areas of economic research that would help to understand the potential implications of AI in this market.

1.2 The Value Chain and Medical Decision Making

We begin by describing a highly stylized value chain in medical care, which highlights the central role of the physician. While the total value chain is quite complex, and includes medical innovation as well as health insurance, we focus on key steps from onset of illness to delivery of treatment:

- A patient visits a medical provider, usually a physician, either complaining of symptoms or for a routine check-up.
- The provider and the patient discuss the patient's health, and the provider performs additional diagnostic tests and procedures based on the information gleaned from the patient.
- The provider diagnoses any existing medical problems.
- The provider recommends a course of treatment, which may include watchful waiting, medication, additional diagnostic tests, and/or a surgical procedure.
- If the patient agrees, additional tests and treatment are rendered.

In a seminal paper, Arrow (1963) describes how and why physicians have played a central role in this health care value chain. Acting as learned agents for their patients, physicians help patients determine what medical services they require and who should provide them. Patients trust their physicians to be competent and compassionate. Physicians earn this trust through professional training as well as years of experience. Indoctrination during medical school as well as professional peer pressure further encourage physicians to serve as perfect, or near perfect agents (Dranove 1988). As Arrow put it, "The social obligation for best practice is part of the commodity that the physician sells."

Playing a central role in the value chain, physicians capture a sizable portion of the value they create. Physicians are among the highest-compensated individuals in the United States. For example, in 2017 the average physician earned nearly $350,000 per year, and half of all physicians were in the top two percent of all US earners (Gottlieb et al. 2020). These averages mask meaningful heterogeneity, with primary care physicians having an average income of approximately $250,000 and the average surgeon earning nearly $500,000. Allied medical providers, while still earning salaries that are well above average, do not approach these levels.

Other actors in the value chain, such as hospitals, drug and device makers, and even insurers, also contribute to value creation and capture an economically significant portion of that value as wages and profits. Patients capture the residual value—the difference between the health benefits created by the value chain and their payments in the form of both out-of-pocket spending and insurance premiums.

Almost as quickly as Arrow had described the physician–patient relationship, researchers began identifying ways in which the physician was far from a perfect agent. We discuss this research in the next section. Concerns about biases and errors have led both providers and third-party payers to use research evidence and practical experience to improve medical decision making. In the next section, we describe the history of these interventions with an underlying conceptual framework that AI is the latest and potentially most powerful example of these long-running efforts.

1.3 Third-Party Intervention in the Value Chain

Even before Arrow (1963) described the trustworthy physician-agent and Fuchs (1974) named physicians the "captain of the team," researchers were concerned about medical decision making. One line of research focused on the pernicious effects of fee-for-service reimbursements, which provided incentives to physicians to overtreat their patients.[3] The past fifty years have seen numerous efforts to remedy these potentially negative incentives.[4]

A second line of research on physician agency identified widespread variation in medical practice from doctor to doctor and across regions, such that seemingly identical patients often receive different treatments (Wennberg and Gittelsohn 1973; Cutler et al. 2019). To some extent, this could reflect differences in patient preferences or physician skills and might not indicate inefficiency (Chandra and Staiger 2007; Finkelstein, Gentzkow, and Williams 2016). There is a broad consensus, however, that at least some portion of practice variations reflects poor medical decision making, whether due to poor training, limited experience, or personal biases (Cutler et al. 2019). While most of the literature on inducement and practice variations focuses on their impact on health spending, a prominent 2000 report from the Institutes of Medicine provides alarming evidence of problems with quality that also reflected poor medical decision making (Kohn, Corrigan, and Donaldson 2000). Substandard quality may lead to over 100,000 unnecessary deaths annually in the United States.

If physicians are making poor decisions, it stands to reason that some oversight may be warranted.[5] There is a long history of medical providers

3. For example, see Shain and Roemer (1959), Roemer (1961), Evans (1974), Fuchs (1978), and Luft (1978).

4. These include the introduction of fixed payments per hospital admission (the DRG system), as well as a variety of payment innovations for physicians, often referred to as "payment reform." The latter may include bonuses based on following treatment guidelines or achieving quality metrics.

5. Much of this historical perspective is from Gray and Field (1989).

reviewing each other's decisions. For example, hospital quality assurance committees review medical records to assess the decisions of their medical staff. The first examples of third parties intervening in medical decision making date to the 1950s, when organizations including labor unions and some health maintenance organizations instituted second surgical opinion programs. As the name suggests, payers would not authorize reimbursement for a surgery without approval from an independent surgeon. Organized medicine resisted, and second opinion programs did not rapidly spread.

The Social Security Amendments of 1972 (PL 92-603) catalyzed third-party review by creating professional standards review organizations (PSROs). PRSOs were panels of local physicians that used their own expertise to develop "objective" standards of care for physicians practicing in their communities. By the late 1970s, a congressional subcommittee claimed that there were over two million unnecessary surgeries each year (American College of Surgeons, 1982). Congress authorized Medicare to augment PSROs with second surgical opinion programs, also developed by panels of independent physicians.

These programs worked via two distinct mechanisms that are salient to any consideration of modern AI. First, the review panels published their standards, which practicing physicians could use to inform their medical decisions. In this way, the panels would have complemented physician decision making. Second, the panels could review claims data and punish providers who failed to conform to the standards. In effect, physicians would have to follow the panel's recommendation or face punishment. In this way, the judgments of the panels would have substituted for physician decision making. As it turned out, the panels' guidelines were not well publicized, and the panels lacked meaningful punishment powers. As a result, the programs neither complemented nor substituted for physician decision making. In 1982, Congress replaced PSROs with peer review organizations (PROs), which had slightly stronger enforcement powers. Punishment remained relatively rare, however, and PROs were nearly as ineffective as PSROs.

While government oversight of medical decision making floundered, the private sector took notice of the potential benefits of these efforts. With employers grousing about rising health care benefits costs, commercial insurers introduced utilization review (UR) programs, essentially PROs with more meaningful teeth to match their bark. Insurers typically outsourced UR to independent companies. Interqual, one of the largest of these UR service providers, offers a good example of how UR worked.[6] Interqual would not authorize payment unless a case met two criteria. Intensity of service criteria included "diagnostic and therapeutic services generally requiring hospitalization," whereas severity of illness criteria included "objective,

6. This example drawn from Dranove (1993).

clinical parameters reflecting the need for hospitalization" (Interqual 1989). Interqual's medical advisors developed these criteria from literature reviews and their own experiences. Interqual developed computer algorithms to implement them. An Interqual employee (typically a nurse) could enter relevant clinical data and the algorithm would determine if the patient met the criteria. At the physician end of the interaction, a doctor would typically assign staff (again, typically a nurse) to provide the required data to Interqual. If the patient did not meet one or both of the criteria, the physician might get personally involved, providing further justification for the treatment decision.

Commercial UR programs intervened far more often than did PRSOs and PROs, and this led to a backlash from patients and physicians. Sixty percent of respondents to a 1998 Commonwealth Fund physician survey reported that they had serious problems with external reviews and limitations on their clinical decisions.[7] Physicians expressed concern about the impact on their patients' health. One common complaint was about the opacity of the UR algorithm. Another was that the physician possessed information about the patient that was not incorporated into the UR algorithm. How did the patient sound when they described their condition? How did the patient respond to prior treatments? Do they adhere to prescriptions? Do they have a supportive home environment?

Simply put, UR algorithms might produce the optimal treatment for a patient presenting with a limited, identifiable set of demographic and clinical characteristics. That is, UR can generate "norms" of care. But physicians have additional information about their patients' "idiosyncrasies" that the algorithm omits—often because there is not a plausible way for a physician to communicate the wealth of information that they have about each patient. This creates a tension: Is it better to force potentially biased physicians to conform to norms, or allow them to make their own decisions, factoring in idiosyncrasies? As we discuss, AI does not eliminate this tension, but it may tilt the calculus. This is particularly true if it becomes easier for AI systems to take in larger amounts of complex data.

What often grated on physicians about UR, and may apply equally well to AI, was that reviews were time consuming and cut into their incomes (at least as a measure of dollars per hour worked). A more subtle but potentially far more important factor was that UR threatened the physician's status as "captain of the team." If a computer algorithm could supplant the physician's judgment, this would totally subvert the value chain. Patients would no longer have to place their trust solely in the judgement of their

7. See "The Commonwealth Fund Survey of Physician Experiences with Managed Care," March 1997, https://www.commonwealthfund.org/publications/fund-reports/1997/mar/commonwealth-fund-survey-physician-experiences-managed-care.

physicians—they could, after all, get superior advice from a computer. This, in turn, could transform physicians from professionals whose judgments saved lives to technicians who merely followed directions, and could put the future earnings potential of physicians in jeopardy.

Politicians took notice of the backlash against UR. US House Minority Leader Steve Ellmann (R-MO) stated that "doctors and consumers . . . all have a horror story to tell you about the insurance company that wouldn't pay on the claim" (Hilzenrath 1997). Amid lobbying from organized medicine, many states enacted laws exposing insurers to malpractice regulations; we discuss below how these laws may impact AI. The US House of Representatives passed legislation that would prohibit insurers from overruling physician decisions, and President Clinton proposed a "Patient Bill of Rights" that would have provided recourse for patients when UR agencies denied coverage. Organized medicine widely praised these efforts.[8] Under intense political pressure, and with research studies failing to find consistent cost savings from UR, insurers changed course (Wickizer, Wheeler, and Feldstein 1989; Wickizer 1990). By the early 2000s, they no longer threatened to withhold payments from providers who failed to follow guidelines. Instead, UR would be purely informative.

For the next decade or longer, both the government and private insurers drew on an ever-growing volume of published research, as well as in-house data, to refine treatment guidelines. These remain almost exclusively informative rather than punitive. In the 2000s, the federal Agency for Healthcare Research and Quality sponsored nearly two dozen patient outcome research teams (PORTS), which developed treatment standards for a range of medical conditions, from lumbar spine stenosis and osteoarthritis to prostate cancer and heart attacks.[9] The PORTS developed standards by conducting meta-analyses of the relevant research literatures. Their recommendations came to be described as "evidence-based medicine" or "treatment protocols." When implemented by third parties such as health insurers or, increasingly, electronic health records (EHR) suppliers, these protocols are often referred to as clinical decision support (CDS).

As research evidence has grown, analysts have access to more granular EHR data. This means that the algorithms can incorporate information about what used to be "idiosyncrasies," more finely tailoring recommendations to specific patient needs. This tilts the calculus in favor of third-party oversight, and the evidence-based medicine movement has accelerated. Hospitals and/or payers often provide physicians with highly detailed

8. Reacting favorably to Clinton's proposal, American Medical Association president Thomas Reardon stated, "Restoring public confidence begins by allowing physicians to be advocates for their patients." Cited in "Clinton Proposes Patients' Bill of Rights," *British Medical Journal*, 1997, 315: 1397. For a discussion of the political jousting between organized medicine and health insurers, see Toner (2001).

9. For a detailed history of PORTS, see Freund et al (1999).

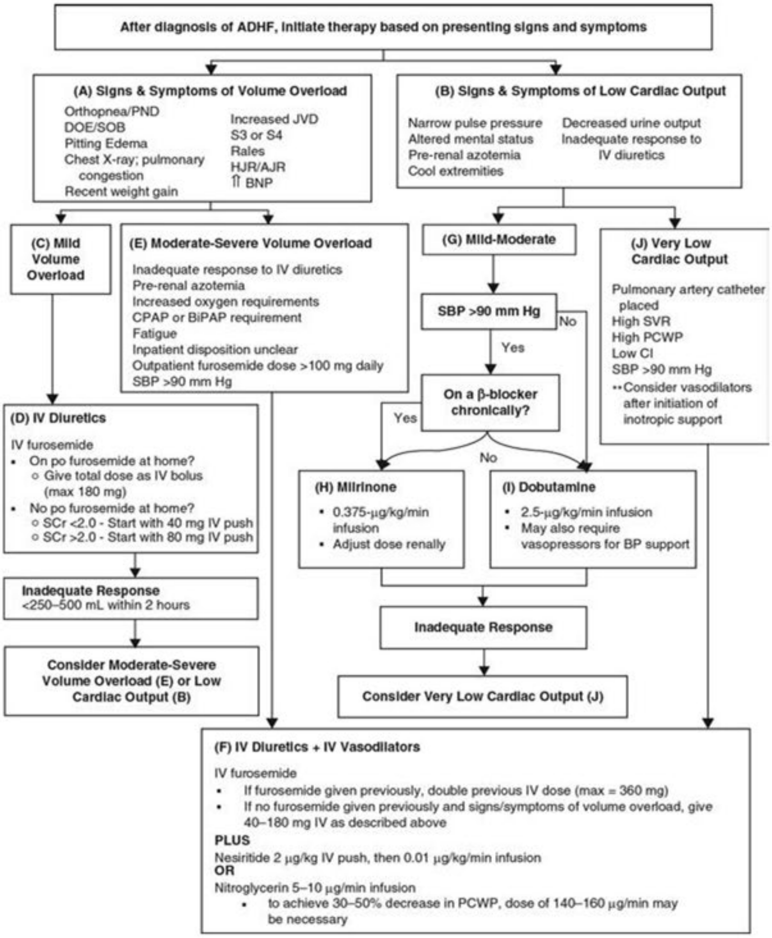

Fig. 1.1 Treatment protocol for acute decompensated heart failure

Source https://www.researchgate.net/figure/Acute-decompensated-heart-failure-ADHF-treatment-algorithm-AJR-abdominal-jugular_fig1_6415930. DiDomenico, Robert J., Hayley Y. Park, Mary Ross Southworth, Heather M. Eyrich, Richard K. Lewis, Jamie M. Finley, and Glen T. Schumock. 2004. "Guidelines for Acute Decompensated Heart Failure Treatment, *Annals of Pharmacotherapy* 38 (4): 649–60, copyright © 2004 by SAGE Publications. Reprinted by Permission of SAGE Publications. Exclusive rights are owned by SAGE and permission for any further reuse must be obtained from SAGE.

treatment protocols, indicating what tests to order, what diagnoses to render based on test results, and what treatments to deliver based on diagnoses and other pertinent patient information. Figure 1.1 depicts a treatment protocol for acute decompensated heart failure. Protocols are usually advisory, and most physicians believe the positive aspects outweigh the negatives, though a sizeable minority believe they limit their ability to make clinical decisions.[10] For the most part, then, evidence-based medicine has served to complement existing physician practice.

AI represents the next step in the development of treatment protocols. By applying advanced data analytics to large data sets, computers incorporate ever more granular data and develop more sophisticated and fine-tuned protocols, including some that target very specific clinical indications. AI also offers new opportunities for oversight, not just of treatment recommendations, but of the diagnostic process. Up to now, a radiologist's reading of an MRI image or a pathologist's analysis of a tissue sample have been inputs into third-party algorithms. AI affords the opportunity to have machines read the MRIs and analyze the tissue samples. The potential for AI to either complement or substitute for physician practice is therefore spreading well beyond anything presented by prior third-party interventions. Understanding the implications for this widespread adoption of automation requires a clearer conceptual model of how such systems can impact the distribution of economic rents in the value chain.

1.4 Automation and the Distribution of Economic Value

Over the past several decades, technological progress has allowed for an increasing set of tasks to be completed by machines rather than humans. This began as primarily substituting for "blue collar" physical labor (e.g., steam shovels replacing physical shovels, tractors replacing horse-drawn plows) resulting in decreased employment among workers in those sectors (Rasmussen 1982; Olmstead and Rhode 2001). Eventually, advancements in computers allowed automation to move into more "white collar" professions resulting in declining employment for particular types of workers in those sectors (Autor 2014).

The ability of automation to undertake such tasks has caused a combination of consternation in the popular press and academic curiosity among researchers. These concerns focus on how the implementation of automation affects employment levels, wages, and inequality. Nonacademics have primarily focused on the ability of automation to substitute for workers and potentially decrease wages, often with doomsday predictions for the future of workers.

Economic theory, however, generates a far more ambiguous and heterogenous set of predictions about the impact of increased automation. Understanding the potential effects of automation requires starting from the idea that inputs to the value chain are generally rewarded based on their

10. Deloitte 2016 Survey of US Physicians, https://www2.deloitte.com/content/dam/Deloitte/us/Documents/life-sciences-health-care/us-lshc-physician-survey-hit-factsheet.pdf. Sixty-one percent of responding physicians agreed or strongly agreed with the statement "Overall, the positive aspects of having protocols outweigh the negatives." Forty-four percent agreed or strongly agreed that protocols "Limit physicians' ability to make clinical decisions."

productivity (Autor 2022). This productivity is itself a function of the input's capabilities (i.e., the economic value they can create) and its relative scarcity (i.e., the economic value it can potentially capture). Given the variety of ways in which value can be created and the changing nature of competition in the market, the productivity of inputs varies over time.

Predicting this variation in productivity requires considering that economic production is actually the result of a bundles of tasks, some of which are accomplished by labor and some by capital. The mix of these inputs varies meaningfully by occupation and over time, as the degree to which production can rely on labor and capital is a function of technological progress. While each of these tasks for production are necessary, changes in the relative cost of each type of input will vary the optimal mix of tasks and the optimal use of labor and capital. It will also be a function of the degree to which newly developed technologies create a *displacement effect* by simply replacing tasks done by labor or a *productivity effect* by increasing the value of other types of labor inputs. To the extent that these new capital inputs raise the value of labor, they will increase rather than decrease demand for these types of labor. Thus, labor may increase from automation, but the effect will vary across the distribution of workers (see Acemoglu and Restrepo 2020 for a discussion of these effects and a broader discussion of the labor economic research into the effects of automation).

Perhaps the most canonical study in this task-based approach of considering automation is Autor, Levy, and Murnane (2003). The authors focus on a task-based approach to the impacts of automation on wages and inequality and posit that to the degree economic value is created by a combination of tasks, the role of increased automation is a function of how it affects the relative contribution of these tasks to create economic value.

Autor, Levy, and Murnane break tasks up into two broad categories—routine and nonroutine. At the time the paper was written, technological limitations meant automation was primarily relegated to completing "routine" tasks, those that follow a well-defined set of rules and an order of operations that can be clearly documented and communicated to computers in the form of a program. These categories resulted from the limits of computer programming and technology at that time. Tacit human knowledge was difficult to communicate to computers, and this served as a fundamental boundary between the types of tasks that could be automated and those that could not. This, in turn, provided some clear bounds of the amount of substitution that could occur.

Autor, Levy, and Murnane illustrate how the impact of automation depends on the degree to which new technologies serve as a substitute or a complement for the work currently done by humans. They find that the rise of automation in routine tasks resulted in a reduction in labor inputs for those tasks. They also found that as the costs of automating routine tasks fell, there was an increase in demand for labor performing nonroutine tasks that were complements to automation. Over time, technology progressed so that even more routine tasks could be automated, with subsequent declines in clerical and administrative occupations (Autor 2014). This trend has continued with the rise of industrial robots, i.e., autonomous machines that can complete well defined tasks without human oversight. Acemoglu and Restrepo (2020) find that the increased use of industrial robots for tasks such as welding, painting, and manufacturing is responsible for a decline in employment for these sectors.

Some described the advent of AI, and its ability to accomplish tasks that cannot be specifically programmed, as overturning that paradigm discussed in Autor, Levy, and Murnane 2003 (e.g., Susskind 2020). In reality, advances in AI have simply shifted the frontier of jobs that could be automated from purely routine tasks to the nonroutine tasks that were reserved for humans identified by Autor, Levy, and Murnane. We posit that while this has clear implications for which types of tasks could ultimately be automated, the fundamental economic points remain largely unchanged and will continue to dictate the distribution of rents across the value chain. Ultimately, the impact of AI-based automation will be a function of the degree to which it displaces labor inputs or increases the productivity of other types of labor inputs—noting that even advancements that increase productivity could result in a net decline of economic rents collected by labor. We also note that the impact of AI could vary greatly across the distribution of workers as certain types of labor may find their tasks are less replaceable than others.

1.5 Modeling AI in Health Care

We draw on this labor economics literature to consider the variety of ways that an increased use of AI could impact the distribution of value in health care. As we consider the relative impact of AI on various actors in the value chain, there are questions both about the degree of complementarity and the relative scarcity of various types of employees. For example, consider a situation where the widespread adoption of AI for diagnostic testing allowed for more medical decision making to be completed by midlevel providers such as physician assistants or nurse practitioners rather than doctors. This would increase the value that could be created by these midlevel providers. However, there are relatively fewer restrictions on entry for this profession, and as a result new workers could be attracted into this sector. As a result, while the value created by the shift to midlevel providers could be quite high, it is not clear whether those providers would capture much of it.

As a starting point, we must consider the appropriate definition of "productivity" in this context. As discussed by Autor, Levy, and Murnane (2003), productivity is the result of the amount of value created and the scarcity of an input in creating the value. Inputs are rewarded based on their productivity. In our context of medical decision making, productivity is related to the inputs used to reach a medical decision and the quality of that medical decision. For example, imagine that the true diagnosis sits along a line or around a circle. There is some reported diagnosis, based on labor and AI inputs, that sits on the same line or circle. The smaller the distance between the reported and true diagnosis, the better the health outcome for the patient. (One can easily include costs into the calculus.) In this way, productivity can be equated to the proximity of the true and reported diagnosis. Likewise for the productivity of the treatment decision. Thus, there is a natural correspondence between productivity as defined by Autor, Levy, and Murnane and productivity in medicine.

Historically, productivity in medicine was primarily the result of a physician's effort with little reliance on technology or third parties. We have discussed how third parties use evidence-based medicine to improve productivity in treatment recommendations. While the jury is out regarding the magnitude of these improvements, there is little doubt that technological change has led to substantial improvements in diagnostic productivity—contrast MRIs with X-rays to diagnose breast cancer, for example. Some new diagnostics require very little judgement or insight. For example, a cholesterol test produces a specific number measuring heart health, and a blood test for measuring glomerular filtration rate provides a clear estimate of kidney function. In these settings, there is little expertise required to perform or interpret the test. Instead, physicians are primarily responsible for knowing which tests to order and what to recommend given a particular set of results. While that frontier of recommendations is obviously moving over time, it is not particularly cumbersome for physicians to follow the frontier in their specialty.

Other innovations, including imaging and genetic tests, require more physician input into reading and interpreting test results. For example, radiologists have historically been critical to reading scans to detect various cancers or other abnormalities. Similarly, orthopedic surgeons read images such as MRIs and X-rays to determine whether patients are candidates for surgeries as opposed to other more conservative interventions. Developing treatment plans from testing that requires more judgement currently requires patients and third-party payers to rely even more heavily on the recommendation of a medical provider.

Let us return to the notion that productivity reflects the difference between true and reported diagnosis/treatment recommendation. It is important to recognize that even if the physician knows what is best for the patient, with no uncertainty, the physician might not truthfully report what is best. Unlike other settings of increased use of automation, where the firm that employs AI will choose the most productive use, it may not be financially advantageous for the physician to use AI. More importantly, the physician might not be constrained by market forces to use it. Consider the case of patients with back pain. Some of these patients may require surgery to address their underlying condition—a treatment plan that may generate significant value for the provider. Of course, there are other patients with back pain that results from less severe underlying medical conditions that would benefit from a more conservative path such as physical therapy and weight management. In a world where the physician is independently responsible for developing and reporting a treatment plan, that physician may recommend surgery when it is unwarranted but financially beneficial to the physician. This may reflect both demand inducement and practice variations that we discussed earlier.

If AI serves the same function as PORTS or treatment protocols—informational but not dictatorial—then demand inducement and practice variations may still lead to suboptimal decisions. Even so, the introduction of AI may lead to far superior outcomes than existing treatment protocols, for several reasons. First, to the extent that AI provides more accurate diagnoses and treatment decisions than even the best current protocols, physicians will increasingly accept its recommendations. This may have a secondary benefit. In an effort to reduce medical spending, insurers have attempted to force physicians to take financial responsibility when their patients' costs exceed various benchmarks. Physicians often resist, arguing that medical costs are too unpredictable. AI can add predictability to both diagnoses and treatment costs, encouraging more physicians to accept payment reforms. Second, accurate AI would give insurers more confidence to challenge physician decisions. In effect, the insurer may prefer an unbiased decision based solely on AI input over a potentially biased physician-determined weighting of AI and physician input. Consider further that while most physicians have not embraced payment reform, many hospitals have, in the form of accountable care organizations and other new payment structures. AI may give hospitals the tools to accurately evaluate the productivity of their own medical staff. Physicians may prefer following treatment recommendations offered by their hospital employers more than from commercial insurers.

Another economically important feature of AI is the ability to use a different combination of inputs in the medical decision-making process. Suppose, for example, that there is some nontrivial fixed cost to physicians reading radiology scans or tissue samples. If AI reading of scans and tissue samples is sufficiently inexpensive and sufficiently accurate, it would be efficient to bypass the physician altogether. As some of our examples show, AI diagnostic accuracy can far exceed what physicians have accomplished, even when physicians incorporate AI into their diagnoses. We suspect that radiologists and pathologists have much to be concerned about as AI use expands into more areas of diagnosis. If industrial robots can replace welders, painters, and others in the broader economy, then can these medical specialists be far behind?

Can the same be said for broader areas of medical decision making? Can robots replace doctors? Rendering diagnoses and making treatment recommendations often require information from patients about their underlying health. Traditionally, physicians obtain this information during office visits and incorporate it into their "personal algorithms." Even if gathering of this information is a crucial step in the value chain, there may be no reason why the physician needs to be involved. This information could be gathered by a midlevel provider such as a physician assistant or a nurse practitioner. To the extent that the information and resulting decisions are colored by various nuances, such as the patient's affect when responding to questions, it might not be sufficient for a midlevel provider to feed the answers to rote questions into a computer. We lack the expertise to state which types of conditions have such subjectivity in reporting, but this seems likely to be an important determinant of when midlevel providers will effectively substitute for physicians. We also note that it is not immediately clear whether physicians are the optimal labor input even when information requires some subjectivity.

Questions about this type of substitution are particularly important because different medical providers take part in related but distinct labor markets. As we discuss below, both wages and entry into these labor markets can be sticky, leading to long run inefficiencies in the labor market response to AI.

1.5.1 Examples of a Task-Based Approach to Examining the Economic Effects of AI on the Health Care Value Chain

This discussion makes it obvious that the distribution of a physician's tasks and the availability of substitute inputs is a key factor in determining value capture. This set of tasks differs by specialty (and likely within specialty across geography and setting). Better understanding which specialties will be most impacted by AI requires examining this variation in the nature of tasks performed and how it intersects with existing and potential AI technology.

To provide a simple illustration of this variation, tables 1.1 and 1.2 list the most common procedures and services delivered to Medicare beneficiaries by two specialists—general internal medicine and radiologists.[11] Table 1.1 ranks the top ten procedures and services by frequency; table 1.2 ranks them by payments.

While nonmedical experts often refer to the profession of a "physician" fairly generically, these lists of job tasks demonstrate the fundamental heterogeneity across different specialties of physicians. Of particular interest is the extent to which the specialties differ in the extent of interaction with patients. As the tables and exhibit show, not surprisingly, radiologists primarily bill payers for their engagements with technology. Radiologists largely read X-ray, CT, MRI, and other diagnostic images; the actual imaging (and engagement with patients) is usually performed by allied medi-

11. We are grateful to Bingxiao Wu for assembling these data.

Table 1.1 Top 10 procedures and services in 2020 (ranked by total number of Medicare services)

	Radiology	
Rank	Service	Number of Medicare services (in millions)
1	X-ray scan	32
2	CT scan	18.9
3	Ultrasound examination	7.2
4	Mammography	6.1
5	MRI scan	5.1
6	Digital tomography	4.8
7	Bone density measurement	1.6
8	Nuclear medicine study	0.7
9	Imaging for evaluation of swallowing function	0.2
10	Biopsy	0.2
	Internal Medicine	
Rank	Service	Number of Medicare services (in millions)
1	Injection of drug	30.8
2	Established patient office or other outpatient service	27.9
3	Subsequent hospital inpatient care	18.9
4	Blood test	8.1
5	Insertion of needle into vein for collection of blood sample	5.4
6	Vaccine	5.4
7	Subsequent nursing facility visit	5.1
8	Annual wellness visit	3.7
9	Initial hospital inpatient care	3.5
10	Hospital discharge day management	3.0

Source: 2020 Medicare utilization and payment data, Physicians & Other Practitioners, at the provider-service level, https://data.cms.gov/provider-summary-by-type-of-service/medicare-physician-other-practitioners/medicare-physician-other-practitioners-by-provider-and-service.

cal personnel under the radiologist's supervision. In contrast, internists are primarily billing for their interactions with patients. The tables and exhibit show that these physicians earn most of their income directly engaging with patients, in office and hospital visits of varying length.

It is also important to note that these particular tasks represent what physicians are able to bill for but fall short of providing a comprehensive description of the tasks necessary to complete these activities. As a result, while a typology based solely on billing codes makes it seems that radiology is devoid of human contact and internists are neo-luddites who eschew technology, the reality is far more complex.

Table 1.2 **Top 10 procedures and services in 2020 (ranked by total Medicare payment)**

	Radiology	
Rank	Service	Medicare payment (in $ millions)
1	CT scan	1,130
2	MRI scan	526
3	Mammography	405
4	Ultrasound examination	322
5	X-ray scan	284
6	Nuclear medicine study	184
7	Digital tomography	176
8	Removal of plaque in arteries	149
9	Bone density measurement	32.3
10	Biopsy	29.9
	Internal Medicine	
Rank	Service/procedure	Medicare payment (in $ millions)
1	Established patient office or other outpatient service	1,760
2	Subsequent hospital inpatient care	1,300
3	Initial hospital inpatient care	511
4	Annual wellness visit	458
5	Subsequent nursing facility visit	328
6	Hospital discharge day management	237
7	Injection of drug	226
8	Vaccine	221
9	Critical care delivery	135
10	Physician telephone patient service	107

Ultimately, the services each type of physician bills involve a combination of tasks with varying degrees of technology and human interaction. To demonstrate the complexity of tasks that underlie the billing codes in tables 1.1 and 1.2, exhibit 1.1 contains the tasks that are identified by the Occupational Information Network (O*Net) to define the occupation of a radiologist and a general internal medicine physician. Examining these tasks makes it immediately clear that a billing-based classification of activities that implies radiologists only interact with technology and internists solely work with human patients is overly simplistic.

At a minimum, radiologists must report their findings to physicians. While reports to other physicians are usually written rather than verbal, they can often contain the kind of nuance that might be difficult for AI to fully repli-

Exhibit 1.1

O*Net Tasks for Radiologists

- Obtain patients' histories from electronic records, patient interviews, dictated reports, or by communicating with referring clinicians.
- Prepare comprehensive interpretive reports of findings.
- Perform or interpret the outcomes of diagnostic imaging procedures including magnetic resonance imaging (MRI), computer tomography (CT), positron emission tomography (PET), nuclear cardiology treadmill studies, mammography, or ultrasound.
- Review or transmit images and information using picture archiving or communications systems.
- Communicate examination results or diagnostic information to referring physicians, patients, or families.
- Provide counseling to radiologic patients to explain the processes, risks, benefits, or alternative treatments.
- Instruct radiologic staff in desired techniques, positions, or projections.
- Confer with medical professionals regarding image-based diagnoses.
- Coordinate radiological services with other medical activities.
- Document the performance, interpretation, or outcomes of all procedures performed.
- Establish or enforce standards for protection of patients or personnel.
- Develop or monitor procedures to ensure adequate quality control of images.
- Recognize or treat complications during and after procedures, including blood pressure problems, pain, oversedation, or bleeding.
- Participate in continuing education activities to maintain and develop expertise.
- Participate in quality improvement activities including discussions of areas where risk of error is high.
- Perform interventional procedures such as image-guided biopsy, percutaneous transluminal angioplasty, transhepatic biliary drainage, or nephrostomy catheter placement.
- Develop treatment plans for radiology patients.
- Administer radioisotopes to clinical patients or research subjects.

Exhibit 1.1 (cont.)

- Advise other physicians of the clinical indications, limitations, assessments, or risks of diagnostic and therapeutic applications of radioactive materials.
- Calculate, measure, or prepare radioisotope dosages.
- Check and approve the quality of diagnostic images before patients are discharged.
- Compare nuclear medicine procedures with other types of procedures, such as computed tomography, ultrasonography, nuclear magnetic resonance imaging, and angiography.
- Direct nuclear medicine technologists or technicians regarding desired dosages, techniques, positions, and projections.
- Establish and enforce radiation protection standards for patients and staff.
- Formulate plans and procedures for nuclear medicine departments.
- Monitor handling of radioactive materials to ensure that established procedures are followed.
- Prescribe radionuclides and dosages to be administered to individual patients.
- Review procedure requests and patients' medical histories to determine applicability of procedures and radioisotopes to be used.
- Teach nuclear medicine, diagnostic radiology, or other specialties at graduate educational level.
- Test dosage evaluation instruments and survey meters to ensure they are operating properly.

O*Net Tasks for General Internal Medicine Physicians

- Analyze records, reports, test results, or examination information to diagnose medical condition of patient.
- Treat internal disorders, such as hypertension, heart disease, diabetes, or problems of the lung, brain, kidney, or gastrointestinal tract.
- Prescribe or administer medication, therapy, and other specialized medical care to treat or prevent illness, disease, or injury.
- Manage and treat common health problems, such as infections, influenza, or pneumonia, as well as serious, chronic, and complex illnesses, in adolescents, adults, and the elderly.

(*continued*)

Exhibit 1.1 (cont.)

- Provide and manage long-term, comprehensive medical care, including diagnosis and nonsurgical treatment of diseases, for adult patients in an office or hospital.
- Explain procedures and discuss test results or prescribed treatments with patients.
- Advise patients and community members concerning diet, activity, hygiene, and disease prevention.
- Make diagnoses when different illnesses occur together or in situations where the diagnosis may be obscure.
- Refer patient to medical specialist or other practitioner when necessary.
- Monitor patients' conditions and progress and reevaluate treatments as necessary.
- Collect, record, and maintain patient information, such as medical history, reports, or examination results.
- Provide consulting services to other doctors caring for patients with special or difficult problems.
- Advise surgeon of a patient's risk status and recommend appropriate intervention to minimize risk.
- Immunize patients to protect them from preventable diseases.
- Direct and coordinate activities of nurses, students, assistants, specialists, therapists, and other medical staff.
- Prepare government or organizational reports on birth, death, and disease statistics, workforce evaluations, or the medical status of individuals.
- Conduct research to develop or test medications, treatments, or procedures to prevent or control disease or injury.
- Operate on patients to remove, repair, or improve functioning of diseased or injured body parts and systems.
- Plan, implement, or administer health programs in hospitals, businesses, or communities for prevention and treatment of injuries or illnesses.

cate. To the extent that radiology reports are formulaic (e.g., they characterize the size and nature of an observed lesion or cyst and state a probability that the lesion or cyst is malignant), AI might be able to produce the same type of report, with greater accuracy. However, there could still be tasks that are necessary for a radiologist to complete in partnership with these results. For example, radiologists also must often work with patients to help them to understand their testing procedures and results—a task that would be hard to imagine being supplanted by AI in the near future.

Nor do the findings from claims data mean that all tasks involved with being an internist require the types of patient interactions that cannot be overtaken by advances in AI. As exhibit 1.1 shows, internists need to be able to accurately diagnose medical conditions from a variety of data, order appropriate tests, and make treatment recommendations for patients. As the task list demonstrates, this is often based on information about a variety of symptoms and ailments, and advanced AI could do better at both diagnosing and identifying treatments in those cases. Even in that case, a medical provider is critical to gather information for the AI system. However, there is no definitive reason that task needs to be completed by a physician. Allied medical personnel can and often do engage in these kinds of patient interactions. To the extent those personnel can serve as a complement to advances in AI, the centrality of even physicians who currently interact a lot with patients could be threatened.

1.6 AI in Medical Practice

Fully understanding the potential scope for automation to serve as either a substitute or a complement for physician productivity requires more knowledge about the types of AI that have currently been developed or could conceivably emerge over a reasonable time frame.

Taking a step back, we note that while technological progress has allowed AI to take part in nearly all aspects of medical practice, at a broad level, these technologies fall into two primary categories:[12]

- *Scanning of test samples to perform diagnoses*. Radiologists visually assess medical images to detect and characterize disease (Hosny et al. 2018). AI algorithms are particularly effective in recognizing and interpreting complex images and therefore may produce faster and more accurate diagnoses than human physicians (Alkhaldi 2021). For example, Kim et al. (2020) partnered with five hospitals to collect mammography scans and patient outcome data. They found that AI improved the detection of breast cancer, with false negative rates falling from 25 percent to 15 percent. The greatest improvement was for early-stage cancers, which are hardest to detect. Mawatari et al. (2020) found that when radiologists relied on deep convolutional neural network (DCNN) software that was trained using data from one hospital to detect hip fractures, false negative rates fell from 17 percent to 9 percent.[13] AI can also aid in the screening of blood and tissue samples. For example, Hollon et al.

12. The lone exception, surgery, faces disruption from another new technology: robots.

13. When radiologists incorporated DCNN results but also considered their own independent reading of images, false negative rates increased to 12 percent.

(2020) studied the time required to interpret histologic images during cancer surgery performed on over 400 patients at one hospital. The surgeon must wait while the samples are read, so every minute counts. They found that DCNN reduces the time required for the pathologist to analyze samples from 40 minutes to 3 minutes, with no difference in accuracy.

- *Mining of clinical data*: Data mining—identifying reproduceable patterns in big data—has several potential uses in health care, from extracting relevant information from EHRs to forecasting diseases before they happen, to recommending treatments tailored to highly detailed clinical information. Mining EHRs can turn up health indicators that predict the onset of disease. For example, the US Department of Veterans Affairs, partnering with DeepMind Health, developed a model to predict acute kidney injury during hospital stays. The model incorporates new health data as it is entered into the EHR and predicts 90 percent of kidney injuries that would require dialysis up to 48 hours before the injury. Such early prediction allows doctors to take steps to reduce the progression of the injury and potentially prevent the need for dialysis. The model also indicates the relevant clinical factors that led to the prediction and the relevant blood tests for monitoring the patient.

Broadly these tasks fall under the category of CDS, which represents a potentially far-reaching use of data mining. An important task for physicians is to translate diagnostic information into treatment recommendations, from which drugs to prescribe to whether the patient requires major surgery. These decisions can be highly complex, involving dozens of clinical indicators (Croskerry 2018). AI can digest information in published research as well as mine thousands of clinical records to identify the best treatments to recommend, at a granularity that is limited only by the size of the data set and the range and precision of variables included in the data. In an early application, the University of North Carolina Cancer Center used IBM Watson's Genomic project to personalize treatments for patients with specific genetic defects (Patel et al. 2018). Admittedly, the use of Watson also stands as evidence of existing limitations with AI. However, it is unclear how binding this limitation will be over time.

These two categories of AI both contribute toward the ultimate goal of medical practice: obtaining an appropriate diagnosis and treatment plan that increases a patient's health. Of course, the economic implications of these two types of AI developments could be vastly different.

We also note that while these are the two broad areas that AI *could* fall into, there has been far more development of diagnostic tools that substitute for potential physician effort. The development of sophisticated CDS that truly guides physicians or other medical providers has not fully emerged into the market. This could, of course, be a function of simple technological progress. However, as we discuss in the next section, there are a variety of economic and market-based barriers that could limit the development of particular types of AI.

1.7 Barriers to AI Development and Implementation

To the degree that automation decreases the value captured by traditional medical providers, it will create opportunities for other parts of the value chain to capture value. As such, we also discuss below how the effect of AI on the value chain affects the market for developing new AI in this sector. The existing labor economics literature on the role of automation often thinks of these technological developments as exogenous—and often in the case of manufacturing or more generic routine tasks, this is a reasonable assumption. However, the development and adoption of AI for medical decision making will require the active participation of physicians and other medical decision makers before its adopted.

While AI has the potential to serve the two broad features of performing diagnoses and supporting clinical decision making, to date we have primarily seen it adopted in a role of diagnosis. Even in the role where AI is being developed to serve a diagnostic role, its widespread adoption has been more limited than some would have expected if one only considers the pace of technological progress.

Some of this lack of adoption both within and across categories could be the result of different applicability of technologies. However, we would argue that in addition to any pure technological features, there are a number of economic factors that have limited the adoption of these services.

1.7.1 Access to Data

Regardless of its application, AI requires data, from scans, blood samples, clinical and outcome data from medical records, and so on. These data largely reside in EHRs and, in principle, are already available for data scientists to explore (though we also are aware that some valuable data currently resides in the minds of physicians). In some countries, such as England, EHRs are universal, uniform, and consolidated—all data is centralized and uniformly reported. In the United States, however, data are fragmented across many EHR systems with limited interoperability. That is, data exchange across platforms is limited. Moreover, even when providers use the same platform, there is substantial customization, which again limits data exchange.

There are also regulatory barriers to assembling and using medical data. Data are protected by the Health Insurance Portability and Accountability Act of 1996 (HIPAA), which can limit the sharing of personally identifiable health information across medical providers. This makes it extremely difficult for third parties to access claims data and to pool data across providers or over time; Tschider (2019) calls this "the healthcare privacy–artificial intelligence impasse." This barrier is particularly problematic for the many technology-focused firms that are attempting to develop health care for AI but exist outside of the traditional medical system. Even if HIPAA were relaxed to allow for more data sharing, one could argue that the sensitive nature of the data increases fears of litigation or other negative events when working with such data from outside of your own firm.

The upshot is that in the United States, analytics are often confined to "in-house" data, often from a single hospital or health system. Indeed, the published literature on AI is replete with studies derived from surprisingly small samples drawn from individual hospitals and systems. An important exception are studies involving the US Veterans Administration, which bills itself as the largest integrated health care system in the country. With EHRs covering 9.1 million patients, the VA has proven to be fertile ground for AI development.[14] The VA has even established a National Artificial Intelligence Institute and has published numerous studies of AI in the VA system.[15] Kaiser Permanente, the largest private integrated health system, is also active in AI development. Kaiser and other large systems may find AI becoming a new source of value creation, as their privately developed diagnostic tools and CDS systems give them an edge over smaller providers lacking access to similar data.

As large health systems continue to facilitate AI development, it is unclear how the resulting decision tools will filter out into general use. Systems may want to protect their intellectual property so as to maintain competitive advantage. Even if systems feel charitable, sharing their algorithms may require compromises to accommodate variations in the kinds of data available in different EHRs. It is also concerning whether any relative homogeneity of the patients or the practice of medicine in these organizations could lead to biased AI technologies. This is particularly true when, as we describe below, part of the process of AI involves developing ways in which automation can occur through processes that are not immediately obvious or knowable by humans. Quite simply, as the machine learns, we may not be fully aware what it is learning and whether there is truly external validity to these processes.

Many health insurers have significant skills in data analytics. So too do many companies outside of health care. In order to access and use sensitive health data without running afoul of HIPAA, these companies may need to own the provider practices that generate the data. This may partially explain the integration strategy of Optum, which is the nation's leading employer of physicians. In the absence of widespread data sharing, the potential for both

14. See https://www.research.va.gov/naii/.

15. For example, see Piette et al. (2016), Lee et al. (2021), Rodriguez-Diaz et al. (2021), and Jing et al., (2022)

value creation and value capture by large vertically integrated organizations using in-house data to develop AI systems may be palpable.

1.7.2 Sticky Prices and Entry Barriers

AI will reduce the productivity gap between physicians and allied medical professionals for at least some medical services. In a well-functioning market, we would expect physician fees for the affected services to fall. In equilibrium, physicians might continue to provide these services, albeit at lower wages that reflect the reduced scarcity of their ability in these new production functions. In the long run, declining physician compensation would likely reduce entry by new physicians and drive up the fees for those services not affected by AI or those that are complements to AI and can only be performed by physicians.

The idiosyncrasies of physician fee schedules, however, suggest that this dynamic may not play out exactly as predicted by economic models from outside of health care. Medicare and most private insurers use the resource-based relative value scale (RBRVS) as the basis for setting a fee schedule. Even when private insurers pay a multiple of the Medicare rate, the *relative* value of these services is dictated by this schedule unless the payer engages in effort to separately negotiate the fee (Clemens, Gottlieb, and Molnár 2017).

The RBRVS assigns each of several thousand different physician services a relative value unit (RVU). The fee for any given service is the corresponding RVU for that service, multiplied by a dollar multiplier. The Centers for Medicare and Medicaid Services sets the multiplier for services delivered to Medicare beneficiaries. Private insurers either set or negotiate a separate multiplier for their enrollees. The important point is that the relative fees for all services are dictated by the RVU, and that RVUs are based on the resource inputs required to produce the service—essentially physician and office staff time and overhead. In other words, the relative fees for different services are effectively based on a labor theory of value rather than on the productivity of the input. Market forces only enter to the extent that they influence the overall multiplier and not the relative value of various tasks.[16]

The implication is that while AI may reduce the relative productivity of certain physician services, it is not likely to lead to a reduction in the relative fees for those services. To the extent that AI affects fees, it will depend on whether AI increases or decreases the amount of time it takes for physicians to render a final diagnosis/recommendation. It is not clear which direction this will go.

Sticky fees will accelerate the shift away from using physicians for ser-

16. Some payers set or negotiate separate fees for a small number of "carved-out" services, such as joint replacement surgery and deliveries of newborns. These fees may be based on market forces. The lion's share of reimbursements are based on RVUs.

vices where their productivity advantage has declined. After all, if a physician has become far less productive relative to affiliated medical providers but wages have not adjusted to reflect this decline, payers may be more inclined to use the affiliated provider. What is interesting, and is an area for more research, is the situation where the productivity advantage has declined but the best potential medical outcome is still the result of the combination of a physician with the newly developed AI technologies. In these settings, there could be a conflict between what is the most effective and what is the most cost-effective treatment—particularly if the wages of physicians are unable to adjust. It is a broader political economic question as to how such conflict would be resolved, but given the history of physician reimbursement and the role of policymakers on limiting the ability of payers to dictate care, it is not obvious that we would reach the most economically efficient outcome.

As the prices for various services evolve in a market, we would normally expect the entry and exit of affected economic actors. Various frictions in the form of entry and exit barriers would, however, limit this movement. This is particularly evident in the labor market for medical providers where a variety of credentialing organizations limit the free flow of individuals into the market. However, these barriers are not the same across different actors. For example, there are far more limits on individuals becoming physicians than there are for other allied medical professionals. This extends beyond simply the amount of time to complete the training. The number of training slots (both seats in domestic medical schools and residency slots at hospitals) are broadly fixed and limit expansions of the supply in response to changing economics. Similarly, physicians are highly trained individuals who may have far worse outside options in the labor market than practicing medicine. This could limit their willingness to move out of the labor force. In contrast, it is relatively less arduous to train for other medical occupations, and similar limitations do not exist constraining supply.

These different entry and exit barriers are important when considering the implications for understanding the potential impact of AI on the distribution of economic rents in the value chain. This is particularly true given the relatively fixed reimbursement of physicians over time that we discuss below.

1.7.3 Medical Malpractice Concerns[17]

A physician who makes an incorrect diagnosis or makes the wrong treatment recommendation, either of which resulting in harm to the patient, may be liable in court for damages and risk professional discipline.[18] This is true

17. Many of the concepts in this section are taken from Sullivan and Schweikart (2019)

18. The key cases are *Sarchett v. Blue Shield of California* 43 Cal. 3d 1, 233 Cal. Rptr. 76, 729 P. 2d 267, (1987) and *Wickline v. California* 192 Cal. App. 3d 1630, 239 Cal. Rptr. 810 (1986). Our discussion here draws on Gray and Field (1989).

even if the physician is following the recommendations of an informed third party, including government entities, such as Medicare-sponsored UR agencies, and private insurers. Physicians may also be liable if they implement suggestions developed through AI. This applies both when they perform services that proved to be medically unnecessary and when they failed to perform medically necessary procedures.

Case law suggests that physicians are ultimately liable for treatment decisions, even when third parties are involved. In *Sarchett v. Blue Shield of California*, the court affirmed the rights of third parties to disagree with a physician's diagnosis and determination of medical necessity. The court added, however, that any doubts about coverage should be construed in favor of the patient. In other words, if the physician insists that a procedure is medically necessary, the insurer should generally be required to cover it. While this seems to protect physicians, the subsequent *Wickline v. California* case limited that protection. The upshot of *Wickline* is that in situations where the physician deems a procedure is medically necessary but the payer denies coverage, both parties may be liable if failure to perform the procedure results in harm. In particular, the burden is on the physician to appeal the insurer's decision. At the same time, if the physician accepts the insurer's recommendation and something goes wrong, the physician is again liable. It seems that a physician who blindly accepts third-party oversight is inviting litigation.

It is not clear how these legal doctrines, which focus on human conduct, will apply to AI (Bathae 2018). As noted by Chinen (2016), "The more autonomy machines achieve, the more tenuous becomes the strategy of attributing and distributing legal responsibility for their behavior to human beings." Even if AI is held responsible (whatever that means), it could prove difficult to find a responsible party, as many individuals and companies contribute to the creation of AI systems. This could leave the physician as the only easily identifiable target in liability suits.[19] On the other hand, if the AI algorithm is developed in partnership with a health system, as is often the case, then plaintiffs will have a clearly identifiable and deep pocketed target.

Malpractice concerns do not entirely weigh against AI adoption. To the extent that AI improves the quality of third-party recommendations, it will reduce the malpractice risks inherent in the current system. Moreover, as much as physicians are at risk for following third-party recommendations that prove incorrect, they are also at risk if they fail to deliver medically necessary treatment or deliver what proves to be the objectively wrong treatment. As AI improves diagnostic accuracy and the appropriateness of treat-

19. It is worth adding that both the Employee Retirement Income Security Act of 1974 and various state doctrines effectively state that corporations cannot practice medicine and therefore cannot be liable for malpractice. (Trueman 2002).

ment recommendations, physicians can reduce their exposure to malpractice risk by following AI recommendations.

Concerns about medical malpractice could be exacerbated in a world where it is not entirely obvious how AI is making particular medical decisions without full knowledge of the process. Autor (2022) describes how we have moved from a world of Polanyi's paradox ("We do not know what we know) to Polanyi's revenge ("We do not know what the computer knows"). While there are a variety of tasks where the productivity gains are sufficient such that we may not care about this lack of knowledge, it is not clear that medical diagnoses and treatment falls in that category—particularly if physicians are worried that such a lack of knowledge could contribute to their liability in the event of a negative health outcome.

1.7.4 Resistance from Organized Medicine

As we discuss above, there is a long history of third-party intervention in medical decision making. When that intervention threatens physician discretion, as in 1990s-style UR, physicians and patients have openly resisted. Physicians are more receptive to advice from third parties, as with PORTS and evidence-based treatment protocols. When third parties partially base reimbursements on whether physicians follow that advice, the reaction has been mixed. This suggests physicians are likely to tolerate AI, provided it complements medical practice. AI that substitutes for physicians will be met with stubborn resistance. If the adoption of AI comes down to a battle between physicians and insurers, we suspect their role as a "trusted agent" will allow physicians to ultimately win. As in the past, insurers may limit using AI to dictate medical practice, and legislators may remove any malpractice exemptions for AI developers.

While resistance by organized medicine is often thought of as an impediment to the *adoption* of existing AI technologies, it is important to consider that in equilibrium the expectation of such resistance by the developers of technology would likely shape the very frontier of what comes to market. Without some amount of deliberate decision making by individuals both within and outside of the health care sector, the ability of value-creating technologies to enter the market in the face of such resistance and as a result society may fail to realize the full potential of AI for health care. This is also true to the degree that the optimal AI systems require meaningful interaction with medical practice to reach their full potential. In one way this is related to the development of such technologies. This, however, could be accomplished by a relatively small set of medical professionals who could have sufficient capital invested in the firms developing AI to overcome any financial resistance. It also could be that only through the adoption and iteration of technology across physicians without a financial stake in the process can we enjoy the most productive AI in health care. In those settings it may be hard to ever have AI reach its full potential.

1.7.5 Resistance from Patients

Perhaps the biggest reason for the managed care backlash of the 1990s was that consumers trusted their physicians more than their insurers. In particular, if payers dictated that particular tests were not medically or economically justifiable but physicians and patients desired such tests, there was little faith put into the "evidence-based medicine" recommended by the payers. There are certainly a number of reasons for this to occur. Part of this is the inherent trust in the physician that described by Arrow (1963). This trust has resulted in a fundamental belief that physicians are primarily interested in the health of their patients—an assertion that we do not contend with but that leaves an economically meaningful amount of "wiggle room" at the margin for medical procedures and tests that are financially advantageous to the physician without being overly injurious to the patient.

Another reason for the inherent distrust of insurers is that as the residual claimants on premiums not spent on medical services, they themselves have inherent economic biases to *undertreat* patients. In current settings, it is rare to see conflict between a physician and a payer be centered on the physician wanting a more conservative treatment path and the payer recommended more expensive and intensive treatments. However, in a world of expanded AI for medical decision making, such paths become more likely. This is particularly true in areas where physicians may have a higher rather of false negative diagnoses than a more automated system. It is unclear whether the emergence of such treatments would shift support away from physicians and toward payers.

One factor that may overcome provider and patient resistance to AI is the rapid evolution of payment modalities. Payers increasingly offer bonuses or in other ways tie compensation to cost savings and/or better outcomes. Payment reforms for hospitals are especially common, with many hospitals participating in accountable care organizations that allow the hospitals to share in any cost reductions, a far cry from fee-for-service and cost-based reimbursement methods common in the times of Arrow and Fuchs. Hospitals may find that following the dictates of AI allows them to enjoy large financial windfalls, and they may push the use of AI onto their doctors and allied medical staff. In this way, payers may indirectly impose the dictates of AI without necessarily feeling the same backlash.

1.8 AI and the Future of Physicians

We have argued that AI can either complement or substitute for labor. The labor economics literature contains a number of predictions about how the degree of complementarity versus substitution will impact the distribution of economic rents in the system. It is an open question beyond the scope of this paper or frankly our expertise as to what technologies will ultimately emerge. However, our analysis suggests that whether new technologies will be substitutes or complements depend on three factors:

1) The nature of the service—diagnosis versus clinical decision making
2) The extent to which physicians have access to information that is not available to or decipherable by a computer
3) The magnitude of biases in physician decision making

We lack the requisite medical knowledge to make definitive statements, but we can make some high-level observations. Regarding the nature of the service, the majority of published AI studies to date appear to target diagnostic accuracy. Studies suggest that AI produces sharply lower false negative and false positive rates, and at least one study shows that AI on its own outperforms physicians who incorporate AI into their final diagnosis. Given these facts, radiologists and pathologists—two relatively highly paid specialties—likely have a lot to fear from AI.

With regard to clinical decision making, there is likely to be a well-identified set of clinical conditions for which treatment recommendations can be standardized and for which physician expertise contributes little extra to value creation. Once patients with these conditions are identified, nurses or other allied medical personnel could issue treatment recommendations, as dictated by the AI system. The question is how to perform the necessary triage. In other words, someone has to obtain and enter the required information into the computer. It remains unclear whether physicians or other allied medical personnel will be better at soliciting such information from patients.

This brings us to the second consideration. When it comes to a computer issuing treatment recommendations, the old expression "Garbage in, garbage out" applies. We can imagine that there are some sets of symptoms and diagnostic test results that leave little margin for error. At the risk of proving our lack of medical knowledge, we suspect that conditions such as conjunctivitis (pink eye) or an ear infection are good examples. For patients presenting with the symptoms of these conditions, physician expertise is not required for the appropriate treatment recommendation. In fact, given concerns about the overuse of antibiotics, it is possible that having an unbiased and automated system may actually be superior in some of these situations.

At the other end of the spectrum are the array of rare diseases diagnosed by television character Dr. Gregory House, who frequently combined clues obtained from personal interactions with the patient and family with years of experience diagnosing rare conditions to make life-saving treatment recommendations. While this fictional character makes for an obvious extreme case, it is clear that value maximization by real world providers will continue to require careful judgments at the time patients present with symptoms and test results. After all, Polanyi's famous quote about the inability to explain what we know applies to patients as much as it does to the creators of automation. One important task of medical providers is the ability to elicit large amounts of information from patients and then determine what is important for the purpose of a medical diagnosis—some of which may be plainly obvious to a patient and some of which may only be apparent to a trained medical professional. For those specific symptoms and tests, it remains an open question as to whether these provider-patient interactions would be improved by the adoption of AI-based clinical decision making. In addition, given malpractice concerns, is it even possible that AI remains anything other than a complement to human medical decision making for all but the simplest clinical conditions?

We finally turn to third-party intervention into medical decision making. This has long been premised on the belief that physicians were biased in favor of performing too many unnecessary services. Even if physicians have information not available to the third party, such bias can justify limiting physician discretion. Utilization review and its descendants are just one way to address bias. In the past decade, payers have introduced a wide range of payment reforms that both limit incentives for overtreatment and reward providers for achieving quality metrics. AI offers an obvious alternative means of implementing more sophisticated means of UR. Again, considering the economic motivations of various actors here will be important. It is clear that a greater automation of UR will ease the burden on medical providers—an existing hassle cost that is a common lament of medical providers. What is unclear is whether that will be viewed as a positive for third-party payers. Given concerns that physicians may ultimately figure out how to "teach to the test" and provide an AI system with the information necessary to always receive approval for treatments, it may be that the cost of an arduous UR system is a feature and not a bug for the payer. That is to say, a higher cost for the physician may discipline how often the provider wants to seek additional treatments and ideally will cause the physician to sort these interactions based on the value created for the patient. While we admit that the current system may not be optimal, it is unclear that a new system relying on AI will be optimal given the economic incentives of the various parties in the value chain.

1.9 Value Capture by AI Developers

AI seems likely to affect value creation and value capture by physicians. In this section we explore the potential for AI developers to capture some of the value they create. We start with a simple observation: the AI market is highly fragmented. While IBM's Watson Health is the best-known AI vendor and IBM invested over $4 billion to build its health care capabilities,[20] it has generated no more than $1 billion in annual revenue and no profits.[21] At the same time, more than a dozen health care providers using IBM Watson

20. Reuters Staff, "IBM to Acquire Truven Health Analytics for $2.6 Billion," February 18, 2016, https://www.reuters.com/article/us-truven-m-a-ibm-idUSKCN0VR1SS.

21. Laura Cooper and Cara Lombardo, "IBM Explores Sale of IBM Watson Health," *Wall Street Journal*, February 18, 2021.

halted or reduced their oncology-related products, and there is little research evidence to show improvements in patient outcomes.[22] IBM sold the Watson Health unit to private equity firm Francisco Partners in 2022. Google Health was created in 2018 to consolidate that company's data-driven health care initiatives, which ranged from Google Brain (its AI initiative) to Fitbit. Google Health shut down after three years, with Google Brain moving into Google Research. Thus far, Google Brain has had little to show in terms of usable AI products in health care (or other sectors of the economy, for that matter.) The rest of AI development in health care is a hodgepodge of provider organizations, start-up tech companies, or joint ventures between the two.[23] By one count, health care accounts for a fifth of all venture funding in AI, and a recent publication highlighted forty start-ups from what is undoubtedly a much larger number.[24] To the extent that different companies are focusing on different areas of diagnosis and treatment, competition for AI products may be limited. That said, the market is likely to be fragmented for the foreseeable future.

At first blush, such fragmentation may seem surprising, given the obvious scale economies associated with data analytics. The history of the EHR market suggests otherwise. Hospitals began adopting advanced EHR systems, which include CDS, in the 2000s. The market was highly fragmented at first, but scale economies and network effects favored consolidation. The market has consolidated, yet remains only "moderately concentrated" (using merger guidelines), with leader Epic holding a 33 percent share of the hospital EHR market and the top four vendors (Epic, Cerner, Meditech, and CPSI) together holding 83 percent of the market.[25] Data on sales to physicians are harder to come by, but it appears that there are ten or more EHR companies for physicians.[26] While the reasons why the market has not further consolidated remains unclear, it does suggest that consolidation in the AI market might also be slow.

Why does fragmentation matter? Consider that most AI applications to date are developed through partnerships between AI developers and health care systems. Bearing in mind that competition among health care providers is local, a successful partnership should allow a local system to create more value. Back up the value chain to the beginning—where AI developers compete to partner with the local system—and we see that fragmentation

22. Daniela Hernandez and Ted Greenwald, "IBM Has a Watson Dilemma," *Wall Street Journal*, August 11, 2018.

23. For further discussion of AI start-ups, see Bertalan Mesko, "Top Artificial Intelligence Companies in Healthcare to Keep and Eye On," *The Medical Futurist*, January 19, 2023, https://medicalfuturist.com/top-artificial-intelligence-companies-in-healthcare/#.

24. See https://builtin.com/artificial-intelligence/artificial-intelligence-healthcare. Accessed August 10, 2022.

25. See https://www.beckershospitalreview.com/ehrs/ehr-vendors-ranked-by-percentage-of-hospital-market-share.html. Accessed August 3, 2022.

26. See https://www.praxisemr.com/top-ehr-vendors.html. Accessed August 3, 2022.

among developers would force them to compete away their rents, leaving them to local providers. To the extent that the local provider market is also fragmented, the health systems will themselves compete away their rents, leaving patients to enjoy the lion's share of benefits.

Consolidation of the AI market would change the calculus of value capture. This calculus also changes to the extent that AI developers and health care providers form deep partnerships with substantial specific investments. Learning distinct data systems and earning the trust of physicians can take time. A developer that embeds itself in a health system stands to capture a sizable portion of the value it creates.

1.10 AI in Less Developed Economic Settings

Much of the discussion of AI in this paper (and in the existing literature) has focused on its adoption in developed country markets and its interaction with the economic incentives of medical provider and third-party payers in those markets. We would be remiss, however, not to also discuss the vastly different implications of a widespread use of AI in less economically developed settings—particularly those without meaningful access to trained medical providers. After all, it is one thing to debate whether AI is superior to a physician alone, an affiliated medical provider working with AI, or some other combination of trained inputs. It is quite another when the counterfactual is no treatment or diagnosis at all—which sadly remains the case in many developing countries. These considerations can also influence discussions about the optimal organization of medical markets in rural settings of developed countries such as the United States—which also often lack ready access to specialists of all types.

In cases where access to medical professionals is constrained and it is not immediately obvious how to relax such constraints, there could be very different welfare implications of even relatively poorly performing automation. After all, in such settings it is not obvious automation should be evaluated against a hypothetical ideal medical diagnosis but instead against a realistic counterfactual of the available standard of care.

That said, as we consider the incentives of the developers of AI, it is possible that the very economic institutions that constrain the availability of medical providers may decrease the economic value of AI to firms developing such technologies. Consider the case of the biopharmaceutical industry, which develops products using a market-based, for-profit model. Under such a model a host of medical conditions endemic to developing countries, such as malaria and other neglected tropical diseases, remain underinvestigated. This is not because of a lack of social value—after all over 400,000 individuals die each year from malaria. Instead, this lack of investment stems directly from the inability of firms to capture a sufficient amount of the social value that they create.

Could AI for developing countries suffer the same fate? It is possible that there are a host of automated technologies that could develop meaningful value in rural or developing country settings but remain overlooked because of the lack of a reasonable expectation of reimbursement by innovators.

Solutions to this possibility are not immediately obvious. While there is a role for government or nongovernmental organizations to step into this area, it is not clear for political economy reasons that we will see such actions. It is one thing for a philanthropy to propose funding a cure for currently incurable condition; it is another to offer funding for an AI system that would not be implemented in a developed market but offers superior efficacy to the existing conditions in a less developed market.

1.11 Conclusion

As the technological frontier advances, the possibility for AI to generate meaningful economic value increases. While this is true for the entire economy, we highlight a series of unique features in the health care sector that would change some implications and predictions for technology in this sector.

While it is well beyond our expertise to predict the future of what technologies can emerge, economics offers important insights into the impact of certain types of technology on market actors. Understanding how the potential emergence of AI can alter the existing distribution of economic surplus in the value chain is important for both predicting and managing the impact of this technology. This is true for allocators of capital and policymakers alike.

It is clear there is great potential for AI to create welfare across a variety of health care settings in developed and developing countries. However, this impact will be a function of exactly which technologies are both developed and adopted. A particularly important point is for actors from outside of health care to understand how the incentives of existing medical providers can influence the future of AI. This could highlight areas where a greater degree of intervention from outside of the sector may be warranted.

References

Acemoglu, Daron, and Pascual Restrepo. 2020. "Robots and Jobs: Evidence from US Labor Markets." *Journal of Political Economy* 128 (6): 2188–244.

Alkhaldi, Nadejda. 2021. "Artificial Intelligence in Radiology—Use Cases and Trends." Ritrex Strategy. June 9, 2021. https://itrexgroup.com/blog/artificial-intelligence-in-radiology-use-cases-predictions/#header.

American College of Surgeons. 1982. *Second Surgical Opinion Programs: A Review and Progress Report*. Chicago: American College of Surgeons.

Arrow, Kenneth J. 1963. "Uncertainty and the Welfare Economics of Medical Care." *American Economic Review* 53 (5): 941–73.

Autor, David H. 2014. "Skills, Education, and the Rise of Earnings Inequality among the 'Other 99 Percent.'" *Science* 344 (6186): 843–51.

Autor, David. 2022. *The Labor Market Impacts of Technological Change: From Unbridled Enthusiasm to Qualified Optimism to Vast Uncertainty*. NBER Working Paper No. 30074. Cambridge, MA: National Bureau of Economic Research.

Autor, David H., Frank Levy, and Richard J. Murnane. 2003. "The Skill Content of Recent Technological Change: An Empirical Exploration." *Quarterly Journal of Economics* 118 (4): 1279–333.

Bathae, Yavar. 2018. "The Artificial Intelligence Black Box and the Failure of Intent and Causation." *Harvard Journal of Law & Technology* 31 (2): 889–938.

Chandra, Amitabh, and Douglas O. Staiger. 2007. "Productivity Spillovers in Health Care: Evidence from the Treatment of Heart Attacks." *Journal of Political Economy* 115 (1): 103–40.

Chinen, Mark A. 2016. "The Co-evolution of Autonomous Machines and Legal Responsibility." *Virginia Journal of Law & Technology* 20 (2): 338–93.

Clemens, Jeffrey, Joshua D. Gottlieb, and Tímea Laura Molnár. 2017. "Do Health Insurers Innovate? Evidence from the Anatomy of Physician Payments." *Journal of Health Economics* 55: 153–67.

Croskerry, Pat. 2018. "Adaptive Expertise in Medical Decision Making." *Medical Teacher* 40 (8): 803–8.

Cutler, David, Jonathan S. Skinner, Ariel Dora Stern, and David Wennberg. 2019. "Physician Beliefs and Patient Preferences: A New Look at Regional Variation in Health Care Spending." *American Economic Journal: Economic Policy* 11 (1): 192–221.

Dranove, David. 1988. "Demand Inducement and the Physician/Patient Relationship." *Economic Inquiry* 26 (2): 281–98.

Dranove, David. 1993. "The Five Ws of Utilization Review." In *American Health Policy: Critical Issues for Reform*, edited by Robert B. Helms. Washington, DC: American Enterprise Institute.

Eijkenaar, Frank, Martin Emmert, Manfred Scheppach, and Oliver Schöffski. 2013. "Effects of Pay for Performance in Health Care: A Systematic Review of Systematic Reviews." *Health Policy* 110 (2–3): 115–30.

Evans, Robert G. 1974. "Supplier-Induced Demand: Some Empirical Evidence and Implications." In *The Economics of Health and Medical Care*, ed. Mark Perlman, 162–73. London: Palgrave Macmillan.

Finkelstein, Amy, Matthew Gentzkow, and Heidi Williams. 2016. "Sources of Geographic Variation in Health Care: Evidence from Patient Migration." *Quarterly Journal of Economics* 131 (4): 1681–726.

Freund, Deborah, Judith Lave, Carolyn Clancy, Gillian Hawker, Victor Hasselblad, Robert Keller, Ellen Schneiter, and James Wright.. 1999. "Patient Outcomes Research Teams: Contribution to Outcomes and Effectiveness Research." *Annual Reviews of Public Health* 20: 337–59.

Fuchs, Victor R. 1974. *Who Shall Live? Health, Economics and Social Choice*. Singapore: World Scientific.

Fuchs, Victor R. 1978. "The Supply of Surgeons and the Demand for Operations." NBER Working Paper No. 236. Cambridge, MA: National Bureau of Economic Research.

Gallup Survey. 2022. "Honesty/Ethics in Professions." Accessed August 5, 2022. https://news.gallup.com/poll/1654/Honesty-Ethics-Professions.aspx

Gottlieb, Joshua D., Maria Polyakova, Kevin Rinz, Hugh Shiplett, and Victoria Udalova. 2020. "Who Values Human Capitalists' Human Capital?: Healthcare

Spending and Physician Earnings." US Census Bureau, Center for Economic Studies, Working Paper No. CES-20-23.

Gray, Bradford H., and Marilyn J. Field, eds. 1989. *Controlling Costs and Changing Patient Care? The Role of Utilization Management*. Washington, DC: National Academies Press.

Hilzenrath, David S. 1997. "Backlash Builds Over Managed Care." *Washington Post*, June 30, A01.

Hollon, Todd C., Balaji Pandian, Arjun R. Adapa, Esteban Urias, Akshay V. Save, Siri Sahib S. Khalsa, Daniel G. Eichberg, et al. 2020. "Near Real-Time Intraoperative Brain Tumor Diagnosis Using Stimulated Raman Histology and Deep Neural Networks." *Nature Medicine* 26 (1): 52–58.

Hosny, Ahmed, Chintan Parmar, John Quackenbush, Lawrence H. Schwartz, and Hugo J. W. L. Aerts. 2018. "Artificial Intelligence in Radiology." *Nature Reviews Cancer* 18 (8): 500–510.

Interqual. 1989. *The ISD-A Review System with Adult Criteria: Consultants, Educators and Publishers to Healthcare Providers*. North Hampton, NH: Interqual.

Jing, Bocheng, W. John Boscardin, W. James Deardorff, Sun Young Jeon, Alexandra K. Lee, Anne L. Donovan, and Sei J. Lee. 2022. "Comparing Machine Learning to Regression Methods for Mortality Prediction Using Veterans Affairs Electronic Health Record Clinical Data." *Medical Care* 60 (6): 470–79.

Kim, Hyo-Eun, Hak Hee Kim, Boo-Kyung Han, Ki Hwan Kim, Kyunghwa Han, Hyeonseob Nam, Eun Hye Lee, and Eun-Kyung Kim.. 2020. "Changes in Cancer Detection and False-Positive Recall in Mammography Using Artificial Intelligence: A Retrospective, Multireader Study." *Lancet Digital Health* 2 (3): e138–e148.

Kohn, Linda T., Janet M. Corrigan, and Moola S. Donaldson, eds. 2000. *To Err Is Human: Building a Safer Health System*. Washington, DC: National Academy Press

Lee, Aaron Y., Ryan T. Yanagihara, Cecilia S. Lee, Marian Blazes, Hoon C. Jung, Yewlin E. Chee, Michael D. Gencarella, Harry Gee, April Y. Maa, Glenn C. Cockerham, Mary Lynch, and Edward J. Boyko. 2021. "Multicenter, Head-to-Head, Real-World Validation Study of Seven Automated Artificial Intelligence Diabetic Retinopathy Screening Systems." *Diabetes Care* 44 (5): 1168–175.

Luft, Harold S. 1978. "How Do Health-Maintenance Organizations Achieve Their 'Savings'?" *New England Journal of Medicine* 298 (24): 1336–43.

Mawatari, Tsubasa, Yoshiko Hayashida, Shigehiko Katsuragawa, Yuta Yoshimatsu, Toshihiko Hamamura, Kenta Anai, Midori Ueno, Satoru Yamaga, Issei Ueda, Takashi Terasawa, Akitaka Fujisaki, Chihiro Chihara, Tomoyuki Miyagi, Takatoshi Aoki, and Yukunori Korogi. 2020. "The Effect of Deep Convolutional Neural Networks on Radiologists' Performance in the Detection of Hip Fractures on Digital Pelvic Radiographs." *European Journal of Radiology* 130: 109188.

Mullainathan, Sendhil, and Ziad Obermeyer. 2021. "Diagnosing Physician Error: A Machine Learning Approach to Low-Value Health Care." *Quarterly Journal of Economics* 137 (2): 679–727.

Obermeyer, Ziad, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. 2019. "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations." *Science* 366 (6464): 446–53.

Olmstead, Alan L., and Paul W. Rhode. 2001. "Reshaping the Landscape: The Impact and Diffusion of the Tractor in American Agriculture, 1910–1960." *Journal of Economic History* 61 (3): 663–98.

Patel, Nirali M., Vanessa V. Michelini, Jeff M. Snell, Saianand Balu, Alan P. Hoyle, Joel S. Parker, Michele C. Hayward, et al. 2018. "Enhancing Next-Generation

Sequencing-Guided Cancer Care through Cognitive Computing." *Oncologist* 23 (2): 179–85.

Piette, John D., Sarah L Krein, Dana Striplin, Nicolle Marinec, Robert D Kerns, Karen B Farris, Satinder Singh, Lawrence An, and Alicia A Heapy. 2016. "Patient-Centered Pain Care Using Artificial Intelligence and Mobile Health Tools: Protocol for a Randomized Study Funded by the US Department of Veterans Affairs Health Services Research and Development Program." *JMIR Research Protocols* 5 (2): e53.

Rasmussen, Wayne D. 1982. "The Mechanization of Agriculture." *Scientific American* 247 (3): 76–89.

Rodriguez-Diaz, Eladio, György Baffy, Wai-Kit Lo, Hiroshi Mashimo, Gitanjali Vidyarthi, Shyam S. Mohapatra, and Satish K. Singh.. 2021. "Real-Time Artificial Intelligence–Based Histologic Classification of Colorectal Polyps with Augmented Visualization." *Gastrointestinal Endoscopy* 93 (3): 662–70.

Roemer, Milton I. 1961. "Bed Supply and Hospital Utilization: A Natural Experiment." *Hospitals* 35: 36–42.

Shain, Max, and Milton I. Roemer. 1959. "Hospital Costs Relate to the Supply of Beds." *Journal of Occupational and Environmental Medicine* 1 (9): 518.

Sullivan, Hannah R., and Scott J. Schweikart. 2019. "Are Current Tort Liability Doctrines Adequate for Addressing Injury Caused by AI?" *AMA Journal of Ethics* 21 (2): 160–66.

Susskind, Daniel. 2020. *A World without Work: Technology, Automation and How We Should Respond*. London: Penguin.

Tschider, C. 2019. "The Healthcare Privacy-Artificial Intelligence Impasse." *Santa Clara High Technology Law Journal* 36 (4): 439.

Toner, R. 2001. "Debate on Patients' Rights Sends Lobbyists into Battle." *New York Times*, June 20, Section A, Page 1.

Trueman, D. 2002. "The Liability of Medical Directors for Utilization Review Decisions." *Journal of Health Law* 35 (1): 105–43.

Wennberg, John, and Alan Gittelsohn. 1973. "Small Area Variations in Health Care Delivery: A Population-Based Health Information System Can Guide Planning and Regulatory Decision-Making." *Science* 182 (4117): 1102–8.

Wickizer, Thomas M. 1990. "The Effect of Utilization Review on Hospital Use and Expenditures: A Review of the Literature and an Update on Recent Findings." *Medical Care Review* 47 (3): 327–63.

Wickizer, Thomas M., John R. C. Wheeler, and Paul J. Feldstein 1989. "Does Utilization Review Reduce Unnecessary Care and Contain Costs?" *Medical Care* 27 (6): 632–46.

Wickizer, Thomas M., and Daniel Lessler. 2002. "Utilization Management: Issues, Effects, and Future Prospects." *Annual Review of Public Health* 23: 233–54.