McKinsey & Company

Healthcare Practice

The coming evolution of healthcare AI toward a modular architecture

The surge in the use of AI healthcare tools is likely to prompt a shift away from point solutions. Early investors will set the foundation for an integrated architecture and clinical-data foundries.

by Aneesh Krishna, David Friend, Neeraj Gohad, and Prashanth Reddy

November 2025

AI has captured the zeitgeist, and healthcare organizations now have an overwhelming number of AI solutions to choose from. Confronted with deep structural challenges, such as labor shortages, rising care costs, and persistent margin compression, it's no surprise that widespread adoption of AI-enabled point solutions (applications that improve a single activity or task) from healthcare services and technology (HST) companies has taken off.

Indeed, it's an embarrassment of riches. AI solutions are proliferating faster than most organizations can absorb them, pressuring healthcare organizations to evolve how they work. The most effective organizations will act local but think enterprise-wide by placing focused bets on deeply tactical AI applications while reimagining business domains. The next horizon of innovation will emerge from the convergence of agentic AI, streamlined data connectivity, and cross-functional coordination for enterprise transformation.

We believe this wave of AI innovation is poised to spark two critical shifts that many stakeholders may not yet be considering but should begin preparing for now. First, leaders will lay the groundwork for a modular, connected AI architecture that brings together point solutions, data infrastructure, and intelligent agents into a cohesive whole. Second, this will create opportunities for healthcare organizations to convert their vast stores of patient records into key enablers by creating clinical-data foundries, which can spur innovation, improve clinical-care delivery and patient outcomes, and unlock new sources of value. Data governance will be at the core of both, and organizations that take establish strong governance frameworks with the appropriate risk guardrails early on to ensure privacy and confidentiality will be able to deliver sustained outcomes.

Realizing this potential will require more than adopting new tools and technologies. The rush toward point solutions has created a fragmented AI environment and new operational friction. Unless leaders' course correct now, they will merely automate today's inefficiencies. Healthcare organizations that invest now in redesigning core business domains and processes for AI-native workflows will be best positioned to capture enterprise value and deliver effective clinical outcomes. Stakeholders in healthcare organizations are at a strategic crossroads: Will they remain passive sellers, buyers, and users of third-party tools, or will they become active shapers of the flexible, integrated healthcare AI architecture that's now a tantalizing possibility?

This article explores how emerging innovations, evolving profit pools, and early signals of change will potentially shape the healthcare landscape. Healthcare AI is shifting from tactical, workflow-specific tools to federated, modular architecture and clinical-data foundries. Consequently, the real battleground will be who controls the data and orchestration layers.

The next horizon is a modular, connected AI architecture

Point solutions are winning because they address real workflow friction with specificity and deliver tangible, near-term value within the domains they're applied to (see sidebar "The success of point solutions"). But they're also creating a new fragmentation problem. Therefore, the long-term opportunity lies in assembling these capabilities into a modular, connected AI architecture that can coordinate point solutions across workflows and domains.

The success of point solutions

Over the past few years, healthcare organizations have been deploying AI in areas with clear functional impact and scalability. Here are some examples of successful point solutions and how AI enablement has spurred authorizations by the US Food and Drug Administration (FDA) and investments.

Proven solutions

Point solutions have been working because they're built for healthcare's deep vertical workflows and help solve operational constraints. AI-enabled ambient-listening tools, for example, integrate seamlessly into administrative and clinical workflows and are being adopted by hundreds of care organizations. Ambient AI helped a large, California-based health system save nearly 16,000 hours in documentation time over a 15-month period.[1]

Additionally, the use of AI in clinical decision-making has gained traction in areas like cardiology and radiology. Tools like those from ChatMD and OpenEvidence have enabled care organizations to rapidly access clinical guidelines and peer-reviewed literature in real time to help them answer complex questions during care delivery to enhance diagnostic confidence. While human-in-the-loop oversight remains essential, AI is increasingly valuable within clinical operations.

AI has also been applied to functions of operations and revenue cycle teams, from scheduling and intake to claims processing and document workflows. Furthermore, AI-powered staffing platforms have helped care organizations fill shifts and reduce reliance on agency labor, contributing to greater workforce stability and cost savings.[2]

Health insurers have implemented chatbots across a variety of duties, such as benefit and claims explanation, triage of member requests, and matching of members with in-network providers. Payers are also working with vendors of electronic health records to enhance provider collaborations to improve care coordination, reduce coverage gaps, and streamline claims processing with AI tools. However, the use of AI tools for prior authorization has led to provider and member concerns, with the US Centers for Medicare & Medicaid Services issuing guidance on prior-authorization automation.[3]

Life science organizations have also deployed AI across the value chain, from discovery to commercialization. Examples include AI Structural Biology Network and OpenFold Consortium's OpenFold3, which is a federated data-sharing platform developed to use proprietary data sets from pharma companies to advance drug discovery; F. Hoffmann-La Roche's digital pathology platform that hosts 20-plus AI algorithms for cancer research and detection, giving pathologists access to third-party innovation; Novartis's use of AI for clinical-trial feasibility and site selection; Amgen's use of deep machine learning in manufacturing to improve quality control; and pharma companies filing regulatory submissions three times faster by using AI tools.[4]

Backed by the market

The surge in AI innovation in US healthcare is reflected in the FDA authorization of more than 1,000 AI-enabled medical devices between 2015 and 2024 (exhibit). Almost 80 percent of these are in medical imaging, particularly radiology, ultrasound, and computed tomography applications.[5] Over the same period, approximately 2,400 AI-enabled healthcare-related companies launched. Roughly 1,750 of those were backed by venture capital, and activity before venture funding also picked up, particularly between 2023 (85 companies) and 2024 (155 companies), which points to continued momentum and early-stage interest.[6] Investment is keeping pace: Between 2023 and 2024, healthcare-AI-venture funding rose by nearly 20 percent, and early 2025 data suggest continued growth. While deal volumes remain modest, average check sizes are climbing sharply, with a growing share of capital concentrated in large-scale bets.

AI holds the promise of becoming the engine of healthcare productivity, and the wave of AI megarounds (funding of at least $100 million) in 2025 reflects a strategic pivot from stand-alone tools to workflow platforms. The largest financing so far this year was $210 million in Series B funding for OpenEvidence, which has an AI-powered medical-information search tool and has been positioning it as a system-level interface for evidence-based

1 Giles Bruce, "16K hours saved: Ambient AI scribes at Kaiser Permanente," Becker's Health IT, June 13, 2025.
2 For more, see "Using AI to create work schedules significantly reduces physician burnout, study shows," American Society of Anesthesiologists press release, January 28, 2022, and Nadine Bienefeld, Emanuela Keller, and Gudela Grote, "AI interventions to alleviate healthcare shortages and enhance work conditions in critical care: Qualitative analysis," *Journal of Medical Internet Research*, January 2025, Volume 27.
3 For more, see "CMS Interoperability and Prior Authorization Final Rule (CMS-0057-F)," US Centers for Medicare & Medicaid Services, January 17, 2024; Jennifer Lubell, "How AI is leading to more prior authorization denials," American Medical Association, March 10, 2025; and Susanna Vogel, "Lawmakers scrutinize AI's role in prior authorization, mental healthcare in House hearing," Healthcare Dive, September 4, 2025.
4 For more, see "Amgen's deep learning approach to vial inspection," Pharmaceutical Online, March 31, 2025; Marie Roehm, "AISB network expands Federated OpenFold3 Initiative with three new pharma contributors," Apheris, October 1, 2025; Moe Alsumidaie, "How Novartis is using AI for clinical trial feasibility and site selection," Clinical Trial Vanguard, December 30, 2024; "Rewiring pharma's regulatory submissions with AI and zero-based design," McKinsey, August 1, 2025; "Roche advances AI-driven cancer diagnostics by expanding its digital pathology open environment," Roche press release, September 9, 2024.
5 "Artificial intelligence-enabled medical devices," US Food and Drug Administration, July 10, 2025.
6 PitchBook, accessed October 2025.

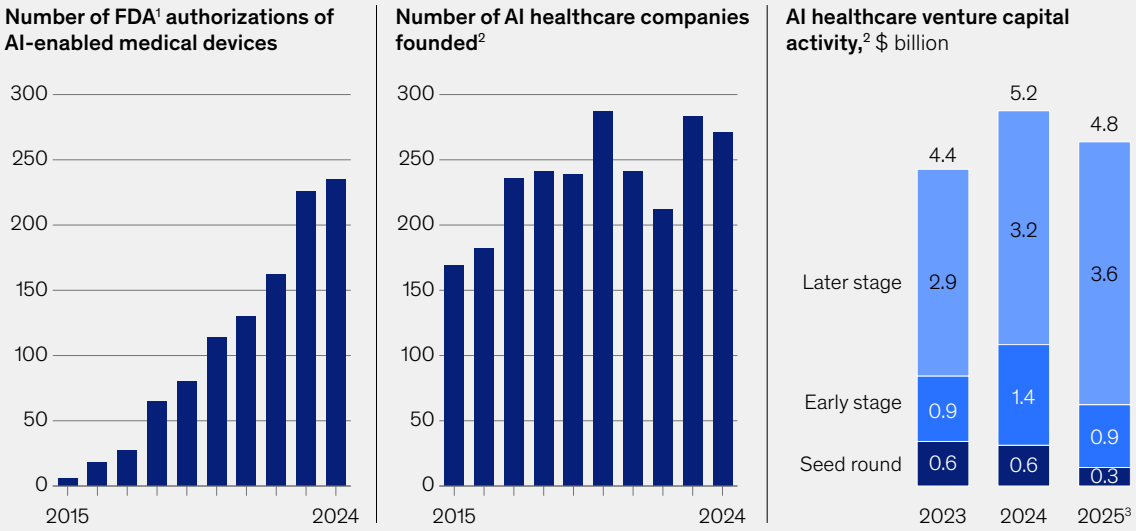
The success of point solutions (continued)

clinical decision-making.[7] The company has since raised another $200 million in Series C funding.[8] Another example is Aidoc, which raised $150 million as it scales its FDA-cleared solutions in cardiovascular care, oncology, and rib fracture triage.[9] The backing from major health systems and Nvidia's venture arm signals growing confidence in AI's role in clinical-decision support.

Exhibit

A rise in innovation in US healthcare AI is evident in the number of related authorizations and start-ups and the increase in later-stage venture capital.

AI in the US healthcare sector



[1]US Food and Drug Administration.
[2]Excludes pharmaceutical and biotechnology companies.
[3]Through June 2025.
Source: FDA; PitchBook

McKinsey & Company

[7] "OpenEvidence, the fastest-growing application for physicians in history, announces $210 million round at $3.5 billion valuation," OpenEvidence press release, July 15, 2025.
[8] Michael J. de la Merced, "OpenEvidence raises $200 million for a ChatGPT for medicine," *New York Times*, October 20, 2025.
[9] "Aidoc secures landmark FDA clearance for first foundation model-powered clinical AI solution of its kind," Aidoc press release, February 18, 2025; Heather Landi, "Clinical AI company Aidoc lands $150M backed by General Catalyst, Nvidia's venture arm," Fierce Healthcare, July 23, 2025.

A modular architecture would combine several key layers: domain-specific AI models that excel in particular functions, intelligent agents acting as connectors that coordinate interactions among these models, and protocols—such as the Model Context Protocol (MCP)—that enable secure, real-time access to data where ever it resides. Together, these components create an architecture that links workflows both within and across domains. Organizations that build an interoperable architecture will cut the time from implementation to impact, lower administrative burden, and reduce revenue leakage.

Investors' current position

Given the proliferation of AI tools, investors are rethinking how to differentiate opportunities in an increasingly crowded field and are becoming more selective about where and how they place capital. Exit timelines remain uncertain, prompting many funds to prioritize companies with embedded distribution, integration readiness, and the potential to become plug-in assets. Some are also pursuing nontraditional plays, such as models aligned with payers and providers, infrastructure-adjacent tools (for example, claims APIs and orchestration layers), and "roll ups" led by private equity that create synergistic ecosystems. In this context, diligence extends beyond product strength to include ecosystem positioning, integration velocity, and long-term defensibility.

Competitive dynamics

Recent product announcements from major electronic-health-record (EHR) companies have intensified the pressure on AI point solution providers.[1] AI capabilities native to EHR companies range from patient engagement tools and scribes to revenue cycle management, analytics that connect and contextualize patient data, and solutions for clinical-trial management and data sharing. These are poised to accelerate natural selection and consolidation, reshaping the competitive landscape. AI may even become the catalyst that redefines not only EHR market dynamics but also the ecosystem of point solution vendors themselves, as vendors are likely to gravitate toward EHR platforms that promote multivendor interoperability and enable provider choice.

Despite the stickiness of EHR workflows, the magnitude of change management required suggests it's still uncertain whether AI will reinforce entrenched proprietary, EHR-native, potentially vendor-locked systems or will tip the balance toward a marketplace model for point solution vendors. However, while the field for HST companies will inevitably consolidate, it's unlikely that any single platform can deliver across all dimensions, and solutions with true differentiation are well positioned to thrive. AI solutions with proven traction, frictionless workflows, agile development cycles, and clear differentiation from EHR-native tools have the potential to withstand the pressure.

The decisions providers face

Care organizations may prefer EHR-native solutions for their simplified procurement, integrated workflows, and minimal deployment burden, but they remain cautious about overdependence on a single platform. Furthermore, concerns about innovation pace, the transition from pilots to full deployment, and access to cutting-edge technologies—often pioneered by AI-native HST companies—will sustain demand for strong point solutions. This has been illustrated by OpenEvidence, which recently launched a no-cost offering to verified providers that combines AI-powered ambient listening with insights from the latest clinical literature. Ultimately, the healthcare AI landscape will be shaped by the juxtaposition of EHR incumbents' method of proceeding with AI integration, HST vendors' continued ingenuity and innovation agility, and the enduring realities of technology adoption and change management in healthcare.

Payer–provider 'bot wars'

An intriguing dynamic is also emerging between health insurers and care providers. While providers deploy AI to streamline clinical documentation and automate both prior authorization and claims processing, payers are adopting similar tools to verify claims and manage denials.

[1] For more, see Ashley Capoot, "Epic touts new AI tools for patients and doctors at company's annual meeting," CNBC, August 20, 2025, and "Oracle to bring new AI capabilities to its patient portal, making it easier for people to understand their medical records," Oracle press release, September 10, 2025.

This evolving interplay has been described as "bot wars," with payer and provider algorithms at times conflicting in automated workflows. Although payers adopted AI solutions early on for utilization management and payments, the pace of adoption among providers suggests that payers will need to rethink and redesign processes and workflows to match the level of efficacy and precision that AI promises to bring to the claims process.

The hope is that the bot wars will result in payer–provider systems that arrive at the right answers more quickly and more often so that manual intervention is generally limited to only a few complex situations that are quickly surfaced to the human in the loop to resolve. This should reduce friction and enable faster, more accurate outcomes. And when ambient listening and AI-enabled claims processing become omnipresent, the quality of solutions, the underlying models they were trained on, and the data that enable this will be the true differentiators. In this environment, success will depend on the strength of underlying enablers—namely, high-quality data, robust governance, traceability, transparency, and system-level design.

The role of hyperscalers in the evolving landscape

As the importance of high-quality data and the need for a scalable AI infrastructure increase, tech giants are rapidly expanding their presence in US healthcare. No longer just infrastructure providers, they're positioning themselves as the operating layer for healthcare AI by developing healthcare-specific cloud platforms, clinical-documentation tools, and clinical-workflow solutions.[2] Nvidia's collaboration with the Mayo Foundation for Medical Education and Research (Mayo Clinic) to build the Mayo Clinic Digital Pathology platform and Google DeepMind's work on MedGemma, a dedicated open model for healthcare-related AI development, are two prominent examples.[3] The recent product rollouts by EHR companies could further strengthen their position. For example, one major EHR incumbent's recently released AI scribe is built in collaboration with a hyperscaler.[4] Others are implementing large language models (LLMs) from a major technology company or AI-powered search summaries in their patient portals.[5]

Look underneath the hood, and hyperscalers are also defining orchestration protocols and interoperability standards that will govern how intelligence flows across systems. Open architectures, such as MCP, are allowing new AI agents to directly access functional data across care organizations. This reduces the need for legacy data lakes, enables context-aware workflows, streamlines the use of dedicated health AI models, and allows LLMs to be incorporated into various workflows.

In parallel, the US Centers for Medicare & Medicaid Services (CMS) interoperability mandates, along with the formation of the CMS Interoperability Framework and CMS-Aligned Networks, are reinforcing this shift toward open architecture and standardized data flows across the healthcare ecosystem.[6] This has the potential to bring about a broader shift from fragmented, task-specific solutions to enterprise-wide AI integration that supports faster innovation, greater scale, and more targeted care delivery. Rather than having one centralized system, the future of healthcare AI is likely to be modular, AI-native architecture that integrates agentic AI mesh and enables interoperability across siloed systems.

---

2 Naomi Diaz, "Tech giants push further into healthcare AI," Becker's Health IT, April 3, 2025.

3 For more, see "Nvidia partners with industry leaders to advance genomics, drug discovery and healthcare," Nvidia press release, January 13, 2025.

4 For more, see Heather Landi, "Epic unveils AI solutions for clinicians, patients and RCM, gives sneak peek at Cosmos AI for risk prediction," Fierce Healthcare, August 19, 2025.

5 For more, see Laura Dyrda, "Oracle, OpenAI dig deep for patient empowerment," Becker's Health IT, September 11, 2025.

6 US Centers for Medicare & Medicaid Services' interoperability mandates include, but are not limited to, data exchange requirements under the Trusted Exchange Framework and Common Agreement and the Fast Healthcare Interoperability Resources standard. For more, see "CMS Interoperability and Prior Authorization Final Rule CMS-0057-F," Centers for Medicare & Medicaid Services, January 17, 2024.

The emerging opportunity of clinical-data foundries

AI healthcare models will be built on curated, high-quality clinical data. Access to longitudinal clinical data, therefore, is likely to emerge as a core competitive advantage. Hyperscalers are starting to collaborate with healthcare organizations to unlock insights from deep clinical data sets—without compromising patient privacy—underscoring the pivotal role health systems play in realizing this opportunity.

The top ten health systems in the United States (by number of hospitals) operate more than 1,200 hospitals with nearly 175,000 beds, giving them substantial control over longitudinal, high-quality clinical data.[7] While making sense of structured and unstructured data has traditionally been tough, gen AI and agentic AI can transform inputs into usable, structured formats. This capability opens entirely new ways for health systems to transform their data assets into a source of strategic strength by creating a new profit pool.

Health systems or AI-enabled data aggregators could license de-identified data sets, with patient consent and underpinned by robust data governance—what we call "data foundries"—for model training to HST companies. Or they could enter revenue-sharing partnerships with pharma and medtech companies. This could spur the evolution of a new archetype of HST company with a business model focused on developing proprietary AI tools in partnership with health systems and rooted in local populations.

The road ahead won't be straightforward, however. A major EHR company's recent announcement about building foundation models on its de-identified database may further reinforce this dynamic. Despite this database's substantial scale, less than 40 percent of health systems that use the EHR, by some estimates, contribute their data.[8] This leaves sufficient space for health systems and an emerging set of HST companies to play an active role in the development and monetization of clinical-data foundries, where health systems can shape the terms of agreement for using their data.

While many AI-enabled HST companies have built point solutions that solve complex care and administrative challenges, the archetype of a full-fledged clinical-data foundry capable of combining device outputs, diagnostic results, genomic data, pharmacy and claims data, physician notes, and trial registries has yet to fully emerge. Execution depends on not just regulatory compliance but also building robust data governance, fostering patient and clinician trust, and aligning incentives across stakeholders to ensure long-term value realization.

Governance, in particular, looms as a critical differentiator. Multidisciplinary AI teams are needed for their oversight and authority over risk assessment, clinical validation, and vendors.[9] Without these guardrails, there's a risk of regulatory exposure, resistance from care organizations, and safety issues.

7 Ethan Popowitz, "Top 10 largest health systems in the US," Definitive Healthcare, April 4, 2025.
8 Hannah Nelson, "How Epic's Cosmos supported clinical research with de-identified data," Xtelligent Healthcare Marketing, February 2, 2022.
9 Agustina D. Saenz et al., "Establishing responsible use of AI guidelines: A comprehensive case study for healthcare institutions," *npj Digital Medicine*, November 2024, Volume 7, Number 1.

Care organizations and data aggregators with native healthcare AI models have the opportunity to shift clinical data from a passive record to an active asset—monetized, integrated, and deployed throughout domains to support decision-making across the healthcare value chain. The emergence of clinical-data foundries may define the next major profit pool in healthcare, and those who curate and activate these assets will shape the economics of the healthcare AI ecosystem.

Recommendations for stakeholders

The rapid evolution of AI in healthcare is reshaping the industry, creating distinct opportunities and imperatives for all parties. Each group brings distinct capabilities and priorities, yet their choices will collectively determine how quickly innovation scales and integrates into clinical practice. Here we provide recommendations for investors, HST companies, hyperscalers, and strategic companies (see sidebar "Recommendations for healthcare CEOs").

Investors and healthcare services and technology companies

Investors should prioritize near-term opportunities with HST companies whose AI-enabled point solutions demonstrate clear traction and scalable use cases and with proven success in navigating the complex procurement environment in healthcare. HST companies should, over time, expand into platform solutions, attracting capital and facilitating strategic roll ups and partnerships with large technology companies or health systems that are well positioned to shape the emerging healthcare AI architecture and clinical-data-foundry opportunity.

In the long run, the most successful exits are likely to come from companies that are building ecosystems rather than stand-alone tools. As healthcare organizations treat AI enabled point solutions as on ramps to the modular AI architecture and clinical-data-foundry strategy, organizations should pick priority domains and redesign them end to end for AI integration as opposed to fragmented pilots.

Organizations that sequence their strategy from high ROI point tools to a connected architecture across domains will be best positioned to lead in the development of clinical-data foundries and create and deliver value. In parallel, these organizations should build a robust multidisciplinary data governance framework that catalogs structured and unstructured assets, defines data

Recommendations for healthcare CEOs

For CEOs across the healthcare value chain, we suggest the following:

- *Prove it, then scale it.* Start where ROI is immediate. Deploy agentic AI to own full workflows (such as documentation, patient communications, and scheduling) and expand only after verification of burden reduction, throughput gains, and quality improvement.
- *Build for the end state, not the pilots.* Treat point solutions as on-ramps. Pick priority domains and redesign and rewire end to end toward a modular AI platform and participation in a clinical-data foundry.
- *Signal scale readiness.* Establish rigorous data governance (including for catalogs, data rights and sharing, model inventory with owners, traceability, and human oversight) and forge partnerships that enable multisite rollouts and clear data foundry agreements.

rights and sharing policies, and maintains a transparent registry of AI models with clear accountability, owners, traceability, and human oversight.

These choices will generate tangible proof points for investors—such as pilots expanding into multisite deployments, measurable improvements across domains, active data foundry partnerships with defined rights, and clear governance practices that demonstrate readiness to scale. Such markers indicate the potential for more widespread, durable adoption and where value will accrue.

Hyperscalers and strategics

Given the scale and scope of the healthcare value chain, hyperscalers are likely to arrive at a crossroads: Should they approach healthcare as one of the verticals in their broader enterprise go-to-market strategy, or should they approach healthcare with a singular focus? With US health spending reaching $4.9 trillion in 2023, healthcare represents not only an economic opportunity but also a sector primed for transformation.[10] Unlocking its full potential will require connecting fragmented data and reimagining domains, systems, and processes. This is an at-scale challenge well suited to the capabilities of hyperscalers.

To lead in this evolving landscape, large tech companies and strategics should consider accelerating partnerships across the healthcare value chain, with an eye on M&A opportunities for AI-enabled point solutions that have deep traction and wide reach, then integrating these solutions into unified platforms. In parallel, companies should look to expand access to longitudinal clinical data through data-sharing agreements. As investment in both AI infrastructure and compute continues to fuel the development of increasingly powerful models (and perhaps ultimately even artificial general intelligence), focus will inevitably turn to one of the most intransigent of sectors—healthcare, which accounts for nearly 17.6 percent of US GDP.[11] Platforms positioned to integrate these advancements will define the next generation of healthcare AI infrastructure.

A defining hallmark of the AI era is monthly and even daily upheaval that forces industries to make sense of today's turmoil and stay ahead of upcoming trends so they can shape a brighter tomorrow. In the long run, organizations that can scale horizontally—across the various clinical and administrative verticals—will be at the forefront of building a modular, connected healthcare AI architecture. This architecture will enable data-driven, system-level value creation across functions and transform how the healthcare value chain operates, ultimately enabling the rise of clinical data foundries.

[10] "Trends in health care spending," American Medical Association, April 17, 2025.
[11] "National health expenditure data: Historical," Centers for Medicare & Medicaid Services, updated December 18, 2024.

Aneesh Krishna is a senior partner in McKinsey's Bay Area office; **David Friend, MD,** is a senior adviser in the Boston office; **Neeraj Gohad** is an associate partner in the Southern California office; and **Prashanth Reddy** is a senior partner in the New Jersey office.

The authors wish to thank Julia Perry for her contributions to this article.

This article was edited by Querida Anderson, a senior editor in the New York office.

Designed by McKinsey Global Publishing