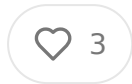


THE MAYO CLINIC VENTURE PLAYBOOK WHY THEIR DATA ACCESS MIGHT BE THE MOST UNDERVALUED EDGE IN HEALTHCARE AI

NOV 11, 2025



DISCLAIMER: These thoughts are my own and do not reflect those of my employer.

ABSTRACT

Mayo Clinic's venture arm offers portfolio companies something most healthcare startups would kill for: structured access to one of the world's most comprehensive clinical datasets spanning over a century of medical records from 1.3 million patients annually. This essay examines the strategic implications of Mayo Clinic Ventures' access arrangements, comparing their model to traditional health system venture arms, analyzing the specific technical advantages of Mayo's unified electronic health record system built on Epic's platform, and evaluating how data partnerships translate into defensible competitive moats for AI-enabled healthcare companies. Key findings include:

- Mayo Clinic Ventures provides portfolio companies with tiered data access through formal collaboration agreements rather than raw data dumps, enabling compliant training and validation of clinical AI models
- The Rochester campus operates as a unified system with consistent documentation standards across specialties, offering higher quality training data than multi-hospital aggregations

- Portfolio companies gain access to Mayo's Clinical Data Analytics Platform which includes structured clinical notes, imaging archives, genomic data, and longitudinal outcomes across complex patient populations
- Mayo's venture model emphasizes clinical validation infrastructure including IT support, physician champion networks, and deployment pathways that compress typical healthcare sales cycles from 18-24 months to 6-9 months
- The data access advantage appears most valuable for companies building clinical decision support tools, rare disease diagnostics, and multimodal AI models that require integrated clinical and imaging data

TABLE OF CONTENTS

The Real Value Proposition Beyond Capital

How Mayo's Data Architecture Actually Works

What Portfolio Companies Get Access To

The Clinical Validation Shortcut That Matters More Than Data

Why Mayo's Model Beats Other Health System Venture Arms

The Catch: What Founders Give Up

Where This Creates Actual Moats

The Companies Getting It Right

The Real Value Proposition Beyond Capital

Most health system venture arms are basically checking a box for innovation. They write small checks, get board observer seats, host demo days, and occasionally facilitate a pilot that goes nowhere. Mayo Clinic Ventures plays a different game.

entirely and the difference comes down to something boringly concrete: they act give portfolio companies structured access to clinical data that can be used to train and validate AI models in ways that matter for FDA clearance and health system procurement.

The numbers behind Mayo's dataset are legitimately impressive in ways that matter for machine learning applications. They see about 1.3 million unique patients per year across their three main campuses in Rochester, Phoenix, and Jacksonville. But the real story is Rochester where they have over a century of continuously maintained medical records through their Unified Data Platform. This is not some hodgepodge of data acquired from different practices with different EMR systems and documentation standards. It's a single integrated academic medical center that has been maniacally organized to collect and maintain clinical documentation since the Mayo brothers started keeping meticulous surgical records in the 1880s.

For AI companies, this matters tremendously because the data quality and consistency issues that plague most healthcare datasets are dramatically reduced. When you're training a model on records from a dozen different community hospitals that got bought by some health system over the past twenty years, you are dealing with different coding practices, different levels of documentation completeness, different imaging protocols, and different patient populations. Mayo Rochester is much closer to a controlled environment where a patient with complex pancreatic cancer gets worked up with similar thoroughness as another patient with the same diagnosis, documented using similar templates and terminology, imaged on similar equipment with similar protocols.

The venture arm launched officially in 2016 although Mayo had been doing strategic investments for years before that through various corporate development vehicles. They typically write initial checks between two and five million dollars and focus on Series A and B stage companies. The portfolio is not huge, maybe forty to fifty active investments, but the curation is pretty deliberate. They are looking for companies where Mayo's clinical expertise and data assets create meaningful strategic value beyond just financial returns.

How Mayo's Data Architecture Actually Works

Understanding what portfolio companies actually get access to requires understanding how Mayo built their data infrastructure because this is not just a matter of querying an EMR system. Mayo has invested hundreds of millions of dollars over the past fifteen years building what they call the Clinical Data Analytics Platform, which is essentially a research-grade data warehouse that sits on top of their Epic production environment and aggregates clinical, imaging, genomic, and operational data into formats that can actually be used for machine learning.

The platform includes about 10 million patient records with varying levels of completeness going back to the mid-1990s when they started electronic documentation in earnest, plus digitized records and coded data from paper charts going back much further for certain patient populations. The structured data includes everything you would expect from a comprehensive Epic implementation: problem lists, medication orders, laboratory results, vital signs, procedures, billing codes. Mayo has gone further than most health systems in structuring clinical notes through natural language processing and creating standardized data models for research.

The imaging archives are particularly valuable because Mayo runs a high-volume academic radiology department that has been methodical about DICOM standards and image quality. They have north of 15 million imaging studies stored in PACS going back about twenty years including CT, MRI, PET, ultrasound, pathology slides, and specialized imaging modalities. For companies building computer vision models for diagnostic imaging, having access to this volume of professionally interpreted images from a tertiary referral center means you can train on rare pathologies and complex cases that you would never find in sufficient numbers at a community hospital.

The genomic data integration is where Mayo has been particularly aggressive relative to other health systems. They have been running a large-scale biobank effort for years and have genomic data on hundreds of thousands of patients that can be linked to clinical phenotypes and outcomes. This is critically important for companies who

on precision medicine applications or pharmacogenomics tools where you need to correlate genetic variants with drug responses or disease progression patterns.

What Portfolio Companies Get Access To

The access model is tiered and formalized through data use agreements that specify pretty specifically what companies can and cannot do. This is not a situation where Mayo hands over a hard drive with patient records. Everything runs through the Clinical Data Analytics Platform with appropriate anonymization, and most of the actual model development happens in a secure research environment where Mayo informaticists can monitor compliance with privacy regulations and use restrictions.

For early-stage portfolio companies that are pre-revenue and still in the algorithm development phase, Mayo typically provides access to de-identified retrospective datasets that match the clinical domain the company is targeting. So if you are building a sepsis prediction model, they will work with you to extract relevant ICD data including lab values, vital signs, medication administrations, and outcomes. If you are building a tool for detecting diabetic retinopathy in fundus photos, they provide access to thousands of retinal images with expert annotations from their ophthalmology department.

The key advantage here is not just volume but ground truth quality. Mayo physicians are generally considered among the best in the world at what they do, and their clinical documentation and diagnostic accuracy tends to be higher than what you get in community practice. When you are training a machine learning model, the quality of your labels matters as much as the quantity of your data. Getting annotations from Mayo radiologists or pathologists means your model is learning from genuine expertise rather than whoever happened to be reading studies at a random community hospital at three in the morning.

As companies mature and move toward clinical validation and FDA clearance, Mayo provides something even more valuable which is access to prospective data collected within their clinical workflows. This is where the venture relationship creates a massive shortcut relative to trying to set up a clinical trial or validation study at

academic medical center as an outside vendor. The typical process for an AI company trying to validate their model at a place like Mayo involves cold outreach to department chairs, navigating institutional politics, submitting research protocols through IRB, recruiting physician champions who are willing to use your tool, and then waiting months or years to collect enough data to demonstrate clinical utility.

Portfolio companies can compress this timeline dramatically because Mayo Ventures actively facilitates connections to relevant clinical departments and helps navigate internal bureaucracy. They have established pathways for getting portfolio company tools deployed in test environments where they can collect real-world performance data without disrupting clinical operations. This is not quite the same as having a green checkmark to deploy your half-baked algorithm in production, but it is much faster than the standard academic research collaboration process.

The Clinical Validation Shortcut That Matters More Than Data

Here is what most AI founders miss about the Mayo relationship: the data access is valuable but the clinical validation infrastructure is actually more important for getting to revenue. Healthcare AI companies die in the valley between having a promising model trained on retrospective data and actually getting that model deployed in clinical practice where it generates revenue. The reasons for this are structural and have to do with how health systems evaluate and procure new technologies.

When a hospital or health system is deciding whether to buy your AI tool, they want to see evidence that it works in real clinical settings with real patients and real workflows, not just performance metrics on a held-out test set from some dataset. They want peer-reviewed publications showing clinical utility. They want to talk to physician champions at respected institutions who have actually used the tool and can vouch for it. They want to see that you have navigated FDA clearance if applicable. All of this takes time and money and relationships that most early-stage startups do not have.

Mayo provides a shortcut to all of this for portfolio companies. They have established research protocols and IRB frameworks for evaluating AI tools across different clinical domains. They have physician champions in basically every specialty who are sophisticated about AI and willing to collaborate on validation studies. They have relationships with top-tier medical journals and can help portfolio companies get their validation work published in venues that actually matter for customer acquisition. They have regulatory experts who understand the FDA software as a medical device framework and can advise on clearance strategy.

The timeline compression here is massive. A typical AI company trying to run a clinical validation study at an academic medical center as an outside vendor might spend six months just navigating contracting and getting IRB approval, another twelve months recruiting patients and collecting data, and then another six months analyzing results and writing papers. With Mayo as an investor and partner, that entire process can happen in less than a year because the infrastructure and relationships are already in place.

This also creates a signaling effect that helps with subsequent customer acquisition. When you can tell a prospect that your tool has been validated at Mayo Clinic and published in a top journal with Mayo physicians as co-authors, you are immediately taken more seriously than if your clinical evidence comes from some community hospital nobody has heard of. Mayo's brand carries enormous weight in healthcare, and being able to associate your company with that brand through a formal partnership and published research creates credibility that would take years to build otherwise.

Why Mayo's Model Beats Other Health System Venture Arms

Most health system venture arms do not offer anything close to this level of data access and clinical validation support. UPMC Enterprises, Kaiser Permanente Ventures, Intermountain Ventures, Providence Ventures - these are all legitimate strategic investors but their value proposition is different and generally less valuable for AI companies specifically.

UPMC has good data assets and strong informatics capabilities but their EMR environment is more fragmented across their network of acquired hospitals and have not built out the same kind of research-grade data infrastructure that Mayo. Kaiser has amazing longitudinal data on their integrated patient population but closed-system model makes it harder to generalize findings to other health systems and they are notoriously slow at implementing new technologies because of their centralized structure.

Cleveland Clinic has a venture arm that is probably the closest comparable to Mayo in terms of offering meaningful data access and clinical validation support to portfolio companies. They have strong data infrastructure, similar academic credibility, and a track record of helping portfolio companies with clinical studies. But Cleveland Clinic sees fewer rare and complex cases than Mayo does because they are more regionally focused, and their data sharing has historically been more restrictive.

The other key differentiator for Mayo is that they have been more aggressive about creating formal frameworks for data partnerships rather than handling everything on a bespoke basis. They have standard data use agreements, established pricing models for different levels of data access, clear governance structures for how portfolio companies can use Mayo data and intellectual property. This might sound boring and bureaucratic but it actually makes things move faster because you are not negotiating terms from scratch for every project.

The Catch: What Founders Give Up

Nothing in life is free and Mayo's data access and clinical validation support comes with strings attached that founders need to understand before taking their money. The most significant trade-off is around intellectual property and data rights. Mayo typically negotiates terms that give them rights to use any algorithms or models that portfolio companies develop using Mayo data for Mayo's own internal purposes. Founders also usually get pretty favorable licensing terms if the company's technology ends up being something Mayo wants to deploy across their enterprise.

This is less onerous than it sounds because Mayo is not trying to own your comp IP outright, but it does mean that if you build something really valuable using th data, they get to use it themselves without paying full commercial rates. For ven backed companies targeting large markets beyond Mayo, this is usually acceptab because Mayo is just one customer. But for companies building tools that are specifically optimized for academic medical centers or niche specialties where M represents a significant percentage of your total addressable market, giving them preferential licensing terms can be more constraining.

The other consideration is commercial conflict. Mayo Ventures typically invests companies that complement rather than compete with Mayo's existing service lin but there can be gray areas. If you are building a telemedicine platform or remot monitoring service that could potentially divert patients away from Mayo faciliti you might find that they are less interested in investing or that they impose restrictions on how you can market to their patient population. This is rational f Mayo's perspective because they are ultimately a provider organization that mak money by delivering clinical care, not a pure venture fund optimizing for financi returns.

There are also practical limitations on data access that founders should understa Mayo is not going to give you unrestricted access to their production EMR environment where you could accidentally expose patient data or disrupt clinica operations. Everything happens in controlled research environments with specif datasets that have been vetted and de-identified. If your model needs iterative ac to fresh data or real-time prediction capabilities, the implementation gets more complex and may require extended partnerships beyond the basic venture arrangement.

The geographic concentration of Mayo's data is both a strength and weakness. If are building a model that needs to generalize across diverse patient populations including different socioeconomic groups, different regional disease patterns, ar different practice patterns, Mayo's data skews toward complex cases from relativ affluent patients who can afford to travel to Rochester or Phoenix for tertiary ca: This is great if you are targeting the high-acuity market that Mayo serves, but le

ideal if you want a model that works well at safety-net hospitals or rural community practices.

Where This Creates Actual Moats

The question every investor should ask is whether Mayo's data access and clinical validation support actually translates into defensible competitive advantages for portfolio companies. The answer depends heavily on what kind of company you are building and how you use the Mayo relationship.

For companies building clinical decision support tools that require training on rare diseases or complex patient presentations, Mayo's data can create a genuine moat because nobody else has comparable datasets. If you are building a tool to help diagnose rare genetic disorders and you train your model on thousands of cases from Mayo's genetics and genomics practice, good luck to any competitor trying to replicate that without similar data access. The long tail of rare diseases is truly long and you need volume from specialized referral centers to have sufficient training examples.

The validation and publication moat is real but depreciates quickly. Getting your initial clinical validation done at Mayo and published with Mayo co-authors creates a 12-18 month head start on competitors, which in healthcare is meaningful because sales cycles are so long. But once you have proven clinical utility and competitors point to your published work as evidence that the approach works, they can run their own validation studies and catch up. The Mayo association helps with credibility in early customer conversations but it does not prevent competition long-term.

Where the Mayo relationship creates the most durable advantage is in ongoing model improvement and maintaining performance as clinical practice evolves. Healthcare is not a stationary domain where you can train a model once and deploy it forever. Clinical guidelines change, new treatments get introduced, coding practices shift, and patient populations evolve. Companies that have continuous access to fresh Mayo data through an ongoing partnership can keep their models current and maintain high performance over time, while competitors are stuck with static training sets that become increasingly stale.

The deployment pathway advantage is also durable if you use it correctly. Many companies struggle to get their first few customers because health systems are risk-averse and nobody wants to be the first to try something new. If you can point to deployment at Mayo Clinic and show that their physicians are actually using you in daily practice, that de-risks the purchase decision for other health systems enormously. Mayo effectively becomes a reference customer that accelerates your cycle for every subsequent deal.

The Companies Getting It Right

Looking at Mayo's portfolio, there are a handful of companies that have leverage data access and clinical validation support particularly effectively. Tempus is probably the highest-profile example although they are well beyond early-stage at this point. They used Mayo's genomic and clinical data to train and validate their precision oncology platform, got published in high-impact journals with Mayo co-authors leveraged that credibility to build relationships with hundreds of cancer centers. Mayo continues to be both an investor and a major customer for Tempus, creating alignment that has been mutually beneficial.

Eko Health, which makes AI-powered stethoscopes for detecting heart disease, ran major validation studies at Mayo that demonstrated their algorithms could detect structural heart disease like heart failure and valvular disease from heart sounds. Mayo collaboration resulted in publications in major cardiology journals and helped them secure FDA clearance. More importantly, it gave them clinical champions at Mayo who could vouch for the technology when they were pitching to other health systems.

Image Analysis Group, which Mayo acquired in 2019 after being a portfolio company, built their medical image analysis platform using Mayo's imaging archives and radiology expertise. The acquisition shows that when the data partnership works and the technology proves valuable to Mayo's clinical operations, there is an exit through acquisition that may not be available to other investors.

On the other end of the spectrum, there are portfolio companies that have struggled to translate Mayo's data access into commercial success because they underestimated the sales and implementation challenges in healthcare. Having a great model that works on Mayo data does not mean hospitals will automatically buy your product. You need to figure out how to integrate with different EMR systems, navigate hospital procurement processes, demonstrate ROI in ways that matter to C-suite buyers, build professional services capabilities to support implementation. Mayo can open doors but they cannot close deals for you.

The other failure mode is companies that become too dependent on Mayo as a single customer and do not build a broad enough customer base. If Mayo represents 20 percent of your revenue because you gave them great pricing as part of the venture deal, and then they decide not to renew because clinical priorities shifted or they built something in-house, you have a major problem. The goal should be to use Mayo as a launching pad to build a real business, not as a crutch.

Looking at where this market is heading, I think the Mayo model of combining venture investment with structured data access is going to become more common as health systems realize they have valuable data assets that can be monetized through strategic partnerships rather than just sold to aggregators. We are already seeing Kaiser, Cleveland Clinic, and others experiment with similar models. The question is whether they can build the data infrastructure and partnership frameworks that make this actually valuable to startups rather than just innovation theater.

For founders evaluating whether to take money from Mayo Ventures, the calculation comes down to how critical high-quality clinical data is to your competitive advantage and how much value you place on fast clinical validation and Mayo's brand association. If you are building something where data quality and quantity are genuinely differentiating, not just table stakes, then Mayo's data access is worth the trade-offs around IP and pricing. If you are building something where the hard part is commercial execution and the clinical validation is straightforward, you might be better off with pure financial investors who can help more with go-to-market strategy.

The other consideration is stage. Mayo's data access is most valuable when you are still in the algorithm development and initial validation phase, roughly seed through Series A. By Series B, you should have your initial clinical validation done and be focused on scaling customer acquisition and proving unit economics. At that point Mayo's strategic value diminishes relative to growth-stage investors who have seen companies scale in healthcare and can help with the operational challenges of going from ten customers to a hundred.

The final thought is that Mayo's model only works because they have genuinely differentiated data assets and clinical expertise that matter for specific kinds of healthcare AI applications. Most health systems do not have this and founders should be skeptical of venture arms from regional health systems that claim to offer similar strategic value without the data infrastructure and academic credibility to back it up. The Mayo, Cleveland Clinic, and Kaiser tier is real. The typical integrated delivery network venture arm is mostly theater.

If you are interested in joining my generalist healthcare angel syndicate, reach out to treyrwales@gmail.com or send me a DM. We don't take a carry and defer annual for six months so investors can decide if they see value before joining officially. Accredited investors only.



3 Likes • 1 Restack

← Previous

Next

Discussion about this post

Comments

Restacks



Write a comment...