

CLINICAL AI & PATIENT CARE

The Preclinical Signal in Routine Abdominal CT

How Mayo's REDMOD and the Pre-Diagnostic Pancreas Force a Rethink of Cancer Screening Math, Workflow Economics, and the Multimodal Future of Inference

MAY 02, 2026 • PAID



Video Preview



Podcast, Part I (free)



Abstract

Pancreatic ductal adenocarcinoma kills roughly 87% of the people who get it, mostly because by the time a tumor is visible on imaging, the underlying biology has been progressing for years and curative resection is off the table. On April 28, 2026, Goenka and Mukherjee and the Mayo plus MD Anderson team published a model called REDMOD in Gut that quietly reframes the problem. Headline numbers: 71% sensitivity and 88% specificity for detecting PDAC signal on abdominal CTs that human radiologists already read as normal, with median lead time around 16 months and outliers stretching past two years. Specialist baseline on the same scans was 10%. Test-retest agreement on repeat scans came in at 90 to 92%, which matters far more than people realize. The viral framing on social calls this AI detecting cancer 3 years early. That is not really what is happening. What is happening is much weirder and considerably more interesting.

Things this essay covers, in shorthand:

- The biology lag (why the pancreas is special and why screening keeps failing effectiveness)
- What REDMOD is and is not (radiomics pipeline, not a 3D CNN tumor detector)
- The pre-diagnostic CT as a new and underexploited asset class
- Why retrospective AUCs flatter everyone and what the AI-PACED prospective trial will probably show
- Bayes and why an 88% specificity model can still bankrupt a screening program
- Workflow economics across the ~30 to 40M annual US abdominal CTs
- Multimodal convergence: CA19-9, ctDNA, methylation MCED, EHR signals the new-onset diabetes wedge
- Commercialization, reimbursement, and the ownership fight nobody has had

Quick anchor numbers: ~67,530 expected US PDAC diagnoses in 2026 per ACS, five-year survival, ~85% of cases caught after locoregional or metastatic spread, ~71% sensitivity and 88% specificity for detecting PDAC signal on abdominal CTs in REDMOD validation, USPSTF still grades general-population PDAC screening as not recommended.

a D, and the wedge cohort (new-onset diabetes after 50) has roughly a 1% three-y PDAC conversion rate per the Sharma cohort work.

Table of Contents

1. The biology lag that ate pancreatic cancer for forty years
2. What REDMOD is actually doing under the hood
3. The pre-diagnostic CT as a new and underexploited asset class
4. Why retrospective AUCs flatter you and what AI-PACED will probably show
5. Bate's hates you and the screening math problem nobody wants to discuss
6. Workflow economics and what 40 million abdominal CTs a year actually imply
7. The multimodal arms race that nobody can credibly avoid
8. Commercialization, reimbursement, and the ownership fight nobody has had

The biology lag that ate pancreatic cancer for forty years

PDAC has been the wall in solid tumor oncology for a generation. The pancreas retroperitoneal, deep, surrounded by bowel gas, draped behind the stomach, and shielded by the duodenal C-loop. By the time a discrete mass is visible on contrast the malignant transformation has often been underway for something like ten to fifteen years. The Yachida and Iacobuzio-Donahue work in Nature back in 2010 established the metastatic timing of pancreatic cancer using genomic dating and basically established that the founder mutation predates clinically detectable disease by c to a decade, with another five years between locally advanced and metastatic disease. So the screening problem in PDAC is not that the field lacks tools. It is that the disease is functionally invisible during its actually curable phase.

The numbers are brutal. American Cancer Society projections for 2026 put expected US PDAC diagnoses around 67,530 with deaths around 51,750, which is more than half. Five-year overall survival sits around 13% across all stages, and that already

number masks the dispersion: localized disease with margin-negative resection and modern adjuvant chemo (FOLFIRINOX or gem-abraxane) gets you to 30 to 40% 1-year survival. Stage IV disease is closer to 3%. Tumors caught under 1 cm with negative nodes can run above 50% five-year survival. The entire game is stage shift. Catching the disease earlier is the only modifiable variable that materially changes the survival curve, and yet the field has spent thirty years failing to identify a screening modality with acceptable PPV in average-risk populations.

USPSTF gave general-population PDAC screening a D grade in 2019 and reaffirmed it in 2024, citing harms exceeding benefits in average-risk adults. The reasoning is Bayesian. Annual incidence in the over-50 US adult population is roughly 20 to 40 cases per 100,000 person-years. Even a great test produces a flood of false positives at this prevalence. Existing high-risk surveillance protocols (CAPS for first-degree relatives of PDAC patients, BRCA1/2 and Lynch carriers, Peutz-Jeghers, hereditary pancreatitis, IPMN follow-up) capture maybe 10 to 15% of incident cancers and have marginal PPV outside familial syndromes. Most patients still present with painless jaundice, weight loss, or non-specific abdominal symptoms after the disease has spread. The clinical default for forty years has been to wait for the disease to declare itself, which means waiting until palliation rather than cure is the realistic objective. This is the wall REDMOD is poking at.

What REDMOD is actually doing under the hood

This is not a transformer or a 3D CNN doing tumor segmentation. It is a radiomics pipeline, which is a meaningfully different beast. The model takes a contrast-enhanced abdominal CT, runs automated pancreas segmentation (almost certainly a UNet or nnUNet derivative pretrained on something like the Pancreas-CT-82 cohort or NIH TCIA datasets), then extracts hundreds of quantitative imaging features according to the IBSI standard (the Image Biomarker Standardization Initiative formalized this in 2020). The features themselves are old-school stuff: gray-level co-occurrence matrices for texture, gray-level run length and size zone matrices for spatial arrangement of intensities, first-order intensity statistics, shape descriptors, and wavelet

decompositions across multiple scales. On top of those features sits a classifier, and certainly tree-based given the longitudinal stability profile (gradient-boosted models like XGBoost or LightGBM tend to produce the kind of 90 to 92% test-retest agreement REDMOD reported, which is unusual for deep models that ride on subtle parenchymal differences).

The 73% sensitivity at 88% specificity in the low-prevalence normal-read cohort is doing real work. The model is not seeing tumors because there are no tumors yet or no tumors visible to a competent abdominal radiologist. What it is picking up is upstream tissue dynamics. Subtle parenchymal heterogeneity, progressive ductal caliber drift, fat fraction shifts in the pancreas itself, focal volume loss in segments of the gland, vague atrophy patterns, and probably some peri-pancreatic stranding signatures that are too low-amplitude for a human reader to consistently flag. None of these are diagnostic of cancer. They are imaging correlates of pre-neoplastic biology: PanIN progression, peritumoral fibrotic remodeling, ductal stricturing from microscopic obstruction, lobular atrophy from upstream cellular loss. The pancreas is leaving a fingerprint that something is wrong long before that something is a malignancy.

The piece of the methodology that matters most for people thinking about this commercially is the multi-institutional validation. The team validated across CT scans from multiple institutions, multiple imaging vendors, and multiple acquisition protocols. A model that only works on a single GE scanner at a single center is not a product, it is an academic paper. The fact that REDMOD performed consistently across scanner manufacturers and protocols is what makes it interesting as deployable infrastructure rather than as a curiosity. The other quietly significant part is that the pipeline runs automatically without time-intensive manual preparation. That is the difference between something that lives on a research workstation and something that can wire into a PACS gateway. Mayo and MD Anderson clearly built this for deployment, not just publication.

The pre-diagnostic CT as a new and underexploited asset class

Here is the thing the AI hype crowd keeps missing. The actual moat in REDMOI is not the architecture or the radiomics math. The radiomics literature has been around since Lambin and Aerts started publishing in 2012, and texture analysis goes back further than that. The moat is the dataset. Linking incidental abdominal CTs to longitudinal patient outcomes, where outcomes means cancer registry confirmation, death index linkage, and downstream diagnosis codes with reasonable timestamp accuracy, is genuinely hard. Most academic centers can do this internally on a small scale by linking their PACS to their tumor registry. Building it across institutions with consistent labeling, demographic diversity, scanner heterogeneity, and clean case-control matching, is a regulatory and integration nightmare.

Pre-diagnostic imaging is essentially a new asset class that did not really exist as a structured resource ten years ago. Paired imaging plus future labels with median 12-month lead time is unbelievably valuable. Whoever owns the longest follow-up windows, the most demographically diverse cohorts, the cleanest case-control pairs, and the most reliable outcomes labeling wins the next decade of imaging AI in oncology. This is exactly why the team here is Mayo plus MD Anderson rather than some startup. Mayo has a multi-decade enterprise imaging archive linked to claims, EMR, and tumor registry data, and MD Anderson is the largest oncology-focused academic system in the country. The combination yields exactly the kind of paired, prospective-but-retrospectively-labeled imaging that REDMOD needed.

Several venture-backed companies have been quietly trying to build this longitudinal imaging plus outcomes graph at scale, with mixed success. The privacy and HIPAA piece pushes most of these toward federated learning architectures where models train across institutions without data leaving any single site. That solves the regulatory problem but introduces new problems around model drift across institutions, distribution shift in the underlying patient mix, and consistency of feature extraction across heterogeneous PACS environments. The imaging exchange and ROI infrastructure layer is also relevant here, since the real bottleneck for these datasets is often not the AI side but the data acquisition and de-identification side. Most retrospective studies that move quickly are running on a single institution because moving images across institutions remains slow, expensive, and operationally

fraught. The companies that build clean longitudinal imaging plus outcomes graphs across institutions are sitting on something that compounds in value the longer you run, because the labels (future cancer diagnoses) only get more accurate with time.

Why retrospective AUCs flatter you and what AI-PACED will probably show

Time for some grown-up statistics talk. The 73 over 88 numbers in the Gut paper are retrospective. The cohort selection was thoughtful but cannot fully escape the structural biases that retrospective imaging studies bake in. All of the scans in the validation cohort were originally interpreted as normal by a clinical radiologist, which means they passed through a human screener who already filtered out the obvious findings. So REDMOD is essentially being benchmarked against the residual signal that human readers missed. That is fair, that is exactly the use case, and that is the right comparison to make. But the 39% specialist baseline is comparing the model to humans operating in routine workflow conditions, not humans given unlimited time and explicit instructions to look for pre-neoplastic signal. With prompting and unlimited time, a subspecialty pancreas radiologist might do considerably better than 39%. That gap matters when thinking about whether REDMOD-class AI is replacing radiologists or augmenting them.

The bigger structural caveat is that the cases in the cohort were selected because they later got diagnosed. So ground truth on the positives is essentially perfect. The matched controls, however, are almost certainly cleaner than a random walk-in cohort would be. Real-world deployment will pull in patients with all kinds of confounders: chronic pancreatitis, cystic lesions, post-surgical anatomy, severe steatosis, prior radiation. Each of those is a potential false-positive driver. The retrospective specificity of 88% probably overstates real-world specificity by at least a few points and possibly more.

This is exactly why AI-PACED is the actual proof. Artificial Intelligence for Pancreatic Cancer Early Detection is the prospective trial that Mayo announced alongside the paper. It is enrolling patients at elevated risk (especially the new-onset

diabetes after 50 cohort and high-risk surveillance candidates), running REDMC existing CTs prospectively, and tracking forward for cancer diagnosis, false positive time to workup, and clinical outcomes. That trial will produce real-world PPV numbers, false positive volumes, downstream MRI and EUS demand, and ultimately the cost-effectiveness data that payers and the FDA actually care about for any screening claim. Until then, the 73% headline is best read as a ceiling under near-retrospective conditions. The number to watch is how much it falls in the prospective setting and how much enrichment of the screened population is needed to maintain a workable PPV.

Bayes hates you and the screening math problem nobody wants to discuss

This is the section where the Twitter pundits get destroyed. PDAC incidence in average-risk over-50 US adults runs roughly 20 to 40 cases per 100,000 person-years depending on the age slice. Call it 0.03% annual prevalence in a typical screened cohort. Plug REDMOD's reported 73% sensitivity and 88% specificity into Bayes and you get something brutal. In a hypothetical screening cohort of one million average-risk over-50 adults, you would expect somewhere around 300 actual cancers in the next year. Of those, REDMOD would correctly flag about 219 (73% of 300). The 12% false positive rate applied to the 999,700 cancer-free patients yields roughly 119,964 false positives. Total positives flagged: about 120,183. PPV: roughly 0.18%. That means for every 555 patients flagged positive, about one actually has cancer.

Every flagged patient gets a workup cascade. Contrast-enhanced MRI with secretin-stimulated MRCP, possibly endoscopic ultrasound, possibly EUS-guided fine needle aspiration if a lesion is found. EUS-FNA carries a non-trivial procedural complication rate, with post-procedure pancreatitis running 1 to 2% in most series and rare but serious adverse events around 0.5%. Workup costs for a single false positive can run \$3,000 to \$8,000 between MRI, EUS, professional fees, and follow-up imaging. Multiply that by 120,000 false positives across a million screened lives and you are looking at roughly \$400 million to \$1 billion in marginal workup costs to find a

hundred cancers. That is the cost-effectiveness wall that has buried every prior I screening attempt and is why USPSTF gave it a D grade.

REDMOD only works economically in enriched populations. New-onset diabetes age 50 carries roughly a 1% three-year PDAC conversion rate. First-degree relative PDAC patients run 5 to 10x baseline risk, and BRCA2 carriers and Peutz-Jegher carriers run substantially higher. In a CAPS-eligible cohort with prevalence around 1%, the Bayes math flips. Out of 10,000 screened, you would expect 100 cancers, REDMOD flags 73 of them and 1,188 false positives, yielding a PPV around 5.8%. That is roughly the same range as low-dose CT for lung cancer screening, which is reimbursed and covered. So the deployment story is not population screening. It is opportunistic risk inference within enriched cohorts. New-onset diabetes is the largest and most identifiable enriched cohort, because the diabetes itself is often paraneoplastic and shows up in claims data within weeks of the metabolic shift. Diabetes is the wedge.

Podcast Part II (paid) and concluding sections below paywall...



Continue reading this post for free, courtesy of Thoughts on Healthcare.

[Claim my free post](#)

[Or purchase a paid subscription.](#)

[← Previous](#)

