Position: AI-Driven Risk Stratification is Essential for Affordable Early Detection of Cancer

Asif Khan[1,8] **Duncan Forster**[1] **Moshir Harsh**[1,5,9] **Daniel Ritter**[1,3] **Chunlei Zheng**[2]

Rose Orenbuch[1] **Abdullah Kuzeiz**[1] **Debora S. Marks**[1,5] **Artem Gazizov**[1]

Nathanael R. Fillmore[2,4]* **Chris Sander**[1,5,6,7,8] *

[1]Harvard Medical School, Boston, MA, USA
[2]VA Boston Healthcare System, Boston, MA, USA
[3]Department of Computer Science, Cornell University, NY, USA
[4]Boston University School of Medicine, Boston, MA, USA
[5]Broad Institute of MIT and Harvard, Boston, MA, USA
[6]CEDAR Center, Knight Cancer Institute, OHSU, Portland, OR, USA
[7]DF/HCC Cancer Center, Boston, MA, USA
[8]Ludwig Center at Harvard, Boston, MA, USA
[9]Max Planck Institute of Molecular Physiology, Dortmund, Germany

Abstract

Early detection of cancer offers the best chance for cure, yet population-wide screening remains either impractical or insufficiently implemented for most malignancies due to several factors. These include heterogeneity of risk, limited technology for early detection, rarity of cases, lack of funding, and limited societal acceptance. This position paper argues that AI-driven risk stratification using longitudinal, population-scale electronic health records (EHRs) provides a practical first step toward scalable surveillance programs. We outline a three-stage framework of prediction, detection, and intervention, in which AI-based risk models nominate small high-risk cohorts to receive advanced diagnostic tests, benefit from emerging early detection technologies, and access timely therapy or preventive care. We emphasize the importance of evaluation metrics such as positive predictive value (PPV) and standardized incidence ratio (SIR), which reflect real-world feasibility, cost-effectiveness, and alignment with healthcare system capacity. As an example, we trained a pancreatic cancer risk model on the US Veterans Affairs database of 15.9 million patients, demonstrating that focusing on high-risk individuals can support a realistic surveillance program. We argue that AI-driven risk stratification, when deployed appropriately as decision support and integrated into early detection and intervention workflows, has the potential to transform cancer care through coordinated efforts across research and healthcare systems.

1 Introduction

Cancer-related mortality can be reduced by diagnosing at an early or pre-malignant stage when curative therapy is still possible. For example, the five-year survival rate for pancreatic ductal

*Co-senior authors.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: GenAI for Health Potential, Trust, and Policy Compliance.

adenocarcinoma (PDAC) detected at stage I is approximately 30% in the United States Veterans Affairs (US-VA) [1] cohort, but falls below 10% at later stages (Figure 1). Therefore, detecting cancer as early as possible creates a critical window for potentially life-saving intervention. However, highly malignant cancers such as PDAC and high-grade serous ovarian cancer (HGSOC) are typically detected at advanced stages, when effective treatment options are limited [2].

To establish screening programs for such cancers, three major constraints must be addressed:

i) *Rarity of cancer at the population level*, since the short-term baseline risk for any single cancer in average-risk adults is below 1% per year. Even highly sensitive or specific tests will yield a low positive predictive value (PPV) when applied population wide without *pretest risk* enrichment through stratification [3].

i) *Heterogeneity of risk*, which varies across individuals (age, sex, genetics, lifestyle, exposures, etc.) and across health systems (coding practices, test availability, and follow-up). These systematic differences can shift data distributions so that risk models developed on specialized cohorts become miscalibrated and underperform when applied in real-world settings [4–6].

iii) *Operational constraints* on health systems in testing and follow-up of patients can result in excessive flagging of individuals, which may overburden diagnostic services and delay essential care. Because a positive screen is followed by confirmatory tests and procedures, screening policies should align with *local capacity* constraints and should be justified by net clinical benefit and cost-effectiveness relative to competing clinical priorities [2, 7, 8].

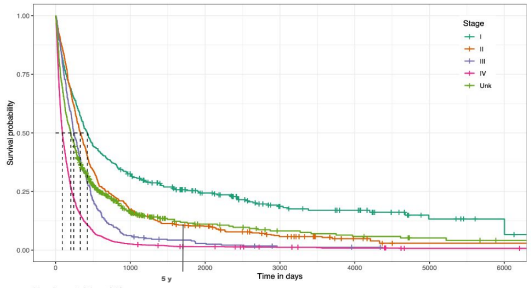

Figure 1: Survival curves for pancreatic cancer (PDAC) in a US population (US-VA) for cancer detected at different stages. Five year survival is very low compared to most other cancer types.

In this position paper, we argue that *AI-enabled risk stratification using longitudinal, population-scale EHRs* represents an important first step in a realistic three-stage *prediction–detection–intervention* surveillance program that can be adapted to the local capacity of a health system while remaining robust to systemic change.

Challenges in current risk stratification approaches for early detection. Barriers to real-world risk assessment and limits of epidemiological risk factors. Most current efforts in risk assessment and early detection are developed and validated in *specialized cohorts* that include individuals with strong inherited predispositions, patients referred for complex cases, or research volunteers who provide structured health information or undergo additional tests not available in routine care [9]. However, such selective cohorts are not representative of the general population. Risk models trained on these data may be poorly calibrated and show reduced performance when applied more broadly, since subtle but clinically relevant events in common populations are often underrepresented. Furthermore, *known epidemiological risk factors* for most cancers in the general population — where data are less structured and often lack genetic information – yield only moderate increases in risk when used individually or combined linearly (e.g., 1.5–3 for factors such as obesity or smoking [10, 11], excluding lung cancer). This makes them inadequate as the foundation for efficient population-wide surveillance.

Limits of genetic risk factors. Information on germline mutations, such as those in *BRCA1/2*, *MLH1*, or *MSH2* which substantially elevate lifetime risk for cancers including breast, ovarian, and colorectal, is rarely available in real-world populations. Moreover, these mutations account for only a small fraction of all cancer cases [12, 13]. A large Nordic twin study estimated the broad-sense heritability of overall cancer risk at 30-37%, leaving a substantial proportion of cancer variance unexplained by genetics alone [14]. Polygenic risk scores (PRS) have been developed for many cancer types, but they currently capture only 10-30% of common variant heritability and result in minimal changes in absolute risk between the lowest and highest scoring individuals [15]. Furthermore, PRS typically provide only a lifetime estimate of genetic cancer risk without modeling age-resolved risk dynamics [16]. Although clinical adoption of PRS remains limited, targeted, whole-exome, and

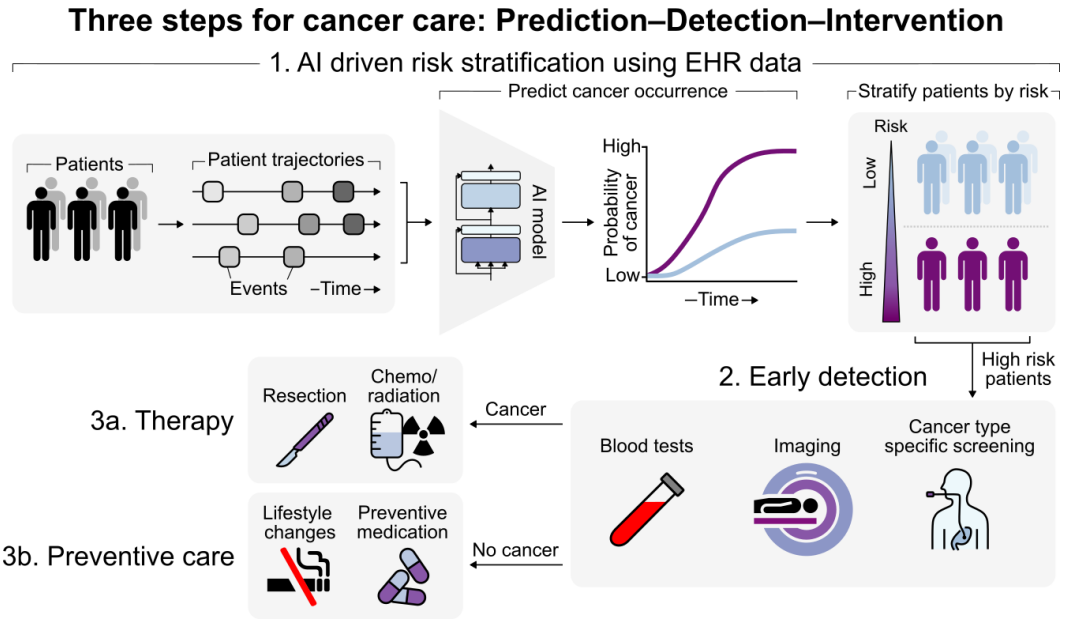

Figure 2: Risk-stratified cancer surveillance: 1.AI-driven risk stratification, 2. targeted early detection, and 3. therapy or prevention.

whole-genome sequencing are becoming increasingly common in clinical practice [17]. As more patient-level genetic data become available, future EHR-based cancer risk models should be able to capture associations between cancer occurrence and genetic, environmental, and life-history factors. At present, however, individual-level measures of genetic cancer risk remain underpowered and are not widely accessible, which limits their practical utility in population-scale risk modeling.

Limits of blood biomarkers. Common blood test biomarkers such as CA-125 for ovarian cancer have low sensitivity and specificity for early-stage disease [18], and advanced imaging modalities (e.g. MRI or PET) are costly, operator dependent, and not suitable for population-wide deployment [19–21].

Current recommendations and gaps in U.S. screening guidelines. Unlike breast cancer, for which population-wide screening with mammography is recommended and widely implemented [22], most cancers lack standardized national screening guidelines from the U.S. Preventive Services Task Force (USPSTF). As a result, potentially useful technologies, such as multi-biomarker detection from blood tests, remain underutilized in clinical practice [23, 24]. In particular, no major guideline-issuing body currently endorses population-wide screening for aggressive cancer types such as pancreatic or ovarian cancer, which are typically diagnosed at late stages, associated with high mortality, and lack low-cost screening tests [2, 25, 26]. For these cancers, a path toward population-level screening will likely require creative, incremental advances: (i) development of better risk stratification models to identify *high-risk cohorts*; (ii) improved early detection technologies (e.g., liquid biopsies based on cell-free DNA or circulating proteins) with higher sensitivity and specificity; (iii) reduced costs of clinical implementation; and (iv) continual improvement across all three areas—more accurate risk quantification, better detection accuracy, and lower cost—with the ultimate goal of expanding population coverage over the coming years.

2 Three steps to clinical decision support.

Cancers are a major cause of mortality: in the U.S., the lifetime risk of death from cancer is approximately 1 in 5 for men and 1 in 7 for women. Because prognosis improves markedly with early diagnosis, population-wide screening for early cancer detection has the potential to greatly reduce both mortality and the side effects associated with late-stage cancer therapies. However, screening tests can be expensive or involve invasive procedures such as PET scans or biopsies, making it infeasible to deploy them routinely across large populations. Being nominated for screening may also cause psychological stress and require costly medical counseling.

To maximize feasibility, a practical approach moving forward is *risk-stratified surveillance* in three stages.

(1) Train AI models for risk prediction on longitudinal, population-scale EHRs to nominate a small, high-risk cohort. Because single-cancer risk over short time horizons is low in average-risk adults, the cost of false positives is high, and specificity must therefore be very high (i.e., a low number of false positives), while maintaining high recall across the full population remains challenging. Hence, at first, the high-risk group should represent a relatively small fraction of the total population. (2) Apply early detection tests within the high-risk group, where they are most efficient, producing stronger signals compared to the general population and increasing the likelihood of successful early detection. (3) Provide early therapy when cancer is detected through the screening program.

An additional benefit of the high-risk approach, even without early detection, is that risk-guided preventive strategies, such as lifestyle modifications, may delay or prevent disease [27], thereby improving survival rates and extending healthy lifespan. Smartly designed screening programs must adhere to established principles of clinical benefit, feasibility, and proportional use of limited diagnostic capacity [2, 8]. The proposed three-step framework — (1) assess risk, (2) screen, (3) treat or prevent — requires careful trade-offs in decision-making to maximize net benefit, for example by choosing site-specific, capacity-aware thresholds and planning with realistic resources for early cancer surveillance [6, 8].

Step 1: AI driven stratification of risk-groups. The primary objective is to train a time-to-event model on large-scale longitudinal EHRs that predicts cancer risk over multiple horizons (e.g., 0-3, 0-6, 0-12, 0-24, 0-36, 0-60 months). Once trained, the model can stratify the risk of individual patients for each horizon window. In practice an *operating threshold* for nominating high risk patients is chosen to balance PPV against clinical capacity, and patients can then be ranked by predicted risk. In practice, this means nominating only the top-ranked subgroups for surveillance, with the threshold tied explicitly to health-system constraints such as the number of available screening slots, the cost of tests, and clinical preferences.

For practical usability, evaluation must extend beyond AUROC/AUPRC scores to capacity-aware measures such as the PPV and standardized incidence ratio (SIR) at different operating thresholds (e.g., top N per million). The SIR is defined as the observed incidence of cancer among individuals flagged as high-risk relative to the expected population incidence in comparable demographic strata. A higher SIR indicates the factor by which model is better compared to incidence based on age and sex alone. This provides clinicians with a measure that they can use to further review patient records or for referral to early detection tests. Selection of the operating threshold should be guided by net benefit analysis and by the expected number-needed-to-screen to identify one true case of cancer [8], ensuring surveillance is both clinically valuable and resource efficient.

Step 2: Early cancer detection methods for use in surveillance programs. The population-wide impact of a three-step program of prediction, detection, and treatment ultimately depends on advances in all three areas. Existing single-cancer screening tests, such as mammography, colonoscopy, and low-dose CT lung scans, are recommended and fairly widely used, but are limited to specific cancers ([2]). To broaden applicability, recent technological developments, such as liquid biopsies, show promise for the detection of early cancers or precancer states while reducing costs [24]. In particular, blood-based multi-cancer early detection tests (MCEDs) such as profiling of circulating tumor cells, exosomes, DNA mutations, DNA methylation and fragmentation patterns [28] or protein profiles using protein mass spectrometry [29–32], antibody pairs [33–35] or RNA aptamers [36].

These detection methods are currently undergoing retrospective and prospective evaluations. For e.g., methylation-based MCED assays are being studied in observational and interventional cohorts such as "Pathfinder" [37] and in health-system scale programs such as NHS-Galleri and "Symplify" trials in the UK [38]. A US-NCI-supported clinical trial (Vanguard Study on Multi-Cancer Detection Tests) is currently enrolling 24,000 participants using tests provided by ClearNote Health and Guardant Health [39]. Early findings suggest feasibility and enrichment for cancers not typically screened with a lower cost per cancer diagnosis. The clinical and economic value of MCEDs is maximized when applied to a small high-risk cohort nominated by Step 1, improving PPV, reducing unnecessary procedures, and supporting efficient use of diagnostic capacity.

Step 3: Intervention: therapy or preventive care. When a screening test is positive, patients are referred for appropriate diagnostic follow-up and clinical management. For these patients,

care must proceed through guideline-based diagnostic confirmation, determine how far cancer has progressed, and, when necessary, begin timely treatment such as surgery, drug therapy, or radiation therapy [40, 41]. Importantly, not all individuals flagged by prediction models will have a positive outcome of the test in Step 2. In these cases, the high risk flag is useful to analyze patient history and guide preventive strategies. Patients may be counseled on changing lifestyle-related risk factors such as smoking, obesity, diabetes, alcohol use, and diet that are known to influence cancer incidence and outcomes [42–44].

3 From retrospective modeling to prospective decision support

Prototyping to enable risk screening. AI-based risk models should first be piloted in "silent mode" to measure calibration, PPV@N (at threshold) and projected workload without influencing care. Next, limited deployment in selected clinics with approved budgets can test integration into workflows. Finally, the models can be deployed in decision support mode as HIPAA-secure web calculators using HL7 standards, such as Fast Healthcare Interoperability Resources (FHIR) [45], and integrated with EHR workflows using single sign-on (SSO) mechanisms [46]. In parallel, code can be executed directly population-wide at the health system level to identify patients at high risk who may benefit from screening. In our view, this incremental implementation has the lowest barrier to initial adoption for healthcare systems. Reports prepared from such pilots can be shared with clinicians to obtain feedback on model outputs, interpretability of factors which most strongly influenced risk prediction, and overall suitability for clinical use. Such retrospective tests create a foundation for prospective evaluation: starting with "silent background deployment", progressing to a multi-site clinical trial of risk-informed screening versus standard of care, where patients could be recruited either by proactive outreach, or at their next point-of-care encounter.

Promoting acceptance by feature interpretation. Identification of top risk factors can help clinicians contextualize AI-derived risk within their experiential and domain knowledge and is critical for designing of clinical trials. Attribution methods such as integrated gradients (IG) [47], to quantify the contribution of individual events in a patient's history towards the risk of cancer. At the *patient level*, IG scores highlight specific clinical events influencing an individual's risk, enabling personalized assessment. At the *population level*, IG scores can be aggregated to identify events that consistently appear to influence predictions, providing broader insight into the model's decision-making. This dual-level interpretability improves transparency, and trust for individualized patient assessment.

Mitigating data drift and ensuring fairness. EHR data changes over time as clinical practice evolves, with new tests being introduced, coding standards updated, and pattern of healthcare access differing across patient groups. These changes, along with the fact that who gets seen and tested (and when) is not random, and can bias both the inputs and the outcomes used for prediction model. For the reliability of AI, it is essential to monitor data drift, recalibrate when practice patterns shift, and plan external re-validation across sites and time [4–6]. To avoid perpetuating disparities, model performance should be evaluated at the same operating threshold for all demographic subgroups (e.g., re-weighting or groupwise calibration), with explicit reporting of any observed biases and the steps to correct them.

Achieving economic sustainability. Screening programs add follow-up testing, clinical workload, and thus financial costs. To keep surveillance programs sustainable, threshold N should match local capacity, using fixed alert limits or flexible screening intervals when needed. Economic evaluation should include the cost per cancer detected, the cost per early-stage case (when treatment is more effective), stage specific treatment costs (substantially higher at late stages) [48], estimated gain in quality-adjusted life years (QALY) [49], and patient out-of-pocket burden [50], comparing risk-stratified surveillance and intervention with current standard care.

3.1 Illustration: Risk stratified surveillance of PDAC

Here, we develop a PDAC risk model using the US-VA database, which has EHR data of 15.9M patients, including 19,426 PDAC cases. Each patient's longitudinal trajectory is represented by time-stamped ICD diagnosis codes and prescription medications. The combined history is modeled with a transformer that learns inter-dependencies across events to assess risk over five time horizons (0-3, 0-6, 0-12, 0-36, and 0-60 months after risk assessment). Model performance is evaluated using AUROC, PPV and SIR metrics. PPV and SIR curves are for the top N high-risk patients per million (Figure 3), where the optimal value of N can be chosen according to the capacity and

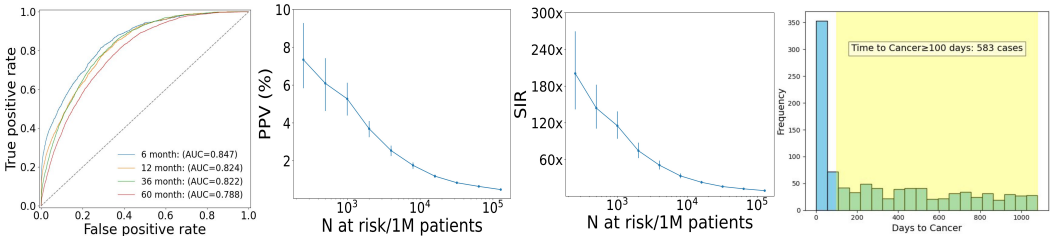

Figure 3: Left: AUROC, PPV, and SIR for a PDAC risk prediction model trained on EHR trajectories of medication and diagnosis codes. Right: for patients correctly predicted to be at high risk (36 month horizon) the actual time of cancer diagnosis.

budget of a surveillance program. For clinical interpretation, we apply IG [47] to identify the ten most predictive diagnoses and medications (Supplementary C). These attributions provide insights into the risk factors contributing most to the model predictions (Figure S1). We further examine the distribution of time-to-cancer among the 1000 high risk patients with a correct cancer prediction (Figure 3). For a prediction horizon of 36 months, 37% of these patients developed cancer more than 100 days after being flagged, showing that elevated risk of condition by several months. Such evaluations can help guide the selection of appropriate tests and the scheduling of repeat follow-up visits in the design of surveillance programs. Additional PDAC risk models have been developed on large-scale EHR data using deep learning approaches [51–53].

4 Conclusion

This position paper argues that realistic surveillance programs for early detection of cancer can be achieved by implementing AI-driven risk prediction in healthcare systems. Such programs would enable health systems to identify and monitor high-risk cohorts within large, real-world populations, supporting timely interventions that shift cancer diagnosis toward earlier, more treatable stages. To make this possible, coordinated collaboration is needed among AI researchers, technology developers of early detection methods, and clinicians across diverse healthcare systems, ultimately worldwide [44]. These developments should be accompanied by advice from patient advocate groups. Such collective efforts can lead to more equitable and effective cancer surveillance and ensure that advancements in early detection technologies and their implementation in early detection clinics benefit a broad spectrum of the population and reduce the overall burden on healthcare systems.

References

[1] Lauren E Price, Kimberly Shea, and Sheila Gephart. The veterans affairs's corporate data warehouse: uses and implications for nursing research and practice. *Nursing administration quarterly*, 39(4):311–318, 2015.

[2] Douglas K Owens, Karina W Davidson, Alex H Krist, Michael J Barry, Michael Cabana, Aaron B Caughey, Susan J Curry, Chyke A Doubeni, John W Epling, Martha Kubik, et al. Screening for pancreatic cancer: Us preventive services task force reaffirmation recommendation statement. *Jama*, 322(5):438–444, 2019.

[3] Douglas G Altman and J Martin Bland. Statistics notes: Diagnostic tests 2: predictive values. *Bmj*, 309(6947):102, 1994.

[4] Samuel G. Finlayson, Adarsh Subbaswamy, Karandeep Singh, John Bowers, Annabel Kupke, Jonathan Zittrain, Isaac S. Kohane, and Suchi Saria. The clinician and dataset shift in artificial intelligence. *New England Journal of Medicine*, 385(3):283–286, 2021. doi: 10.1056/NEJMc2104626.

[5] Geoffrey McGee, Nathan Kallus, Mingyu Yin, et al. On the nature of informative presence bias in analyses of electronic health records. *Journal of the American Medical Informatics Association*, 29(12):2072–2080, 2022. doi: 10.1093/jamia/ocac197. URL `https://pmc.ncbi.nlm.nih.gov/articles/PMC8633193/`.

[6] Gary S. Collins, Karel G. M. Moons, Paula Dhiman, Richard D. Riley, Andrew L. Beam, Ben Van Calster, Marzyeh Ghassemi, Ying Liu, Johannes B. Reitsma, Maarten van Smeden, et al. Tripod+ai statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*, 385:e078378, 2024. doi: 10.1136/bmj-2023-078378.

[7] Mitchell R Zavon. Principles and practice of screening for disease. *Archives of internal medicine*, 123(3):349–349, 1969.

[8] Andrew J Vickers and Elena B Elkin. Decision curve analysis: a novel method for evaluating prediction models. *Medical Decision Making*, 26(6):565–574, 2006.

[9] Charlene F Belanger, Charles H Hennekens, Bernard Rosner, Frank E Speizer, et al. The nurses' health study. *Am J Nurs*, 78(6):1039–1040, 1978.

[10] Béatrice Lauby-Secretan, Chiara Scoccianti, Dana Loomis, Yann Grosse, Franca Bianchini, and Kurt Straif. Body fatness and cancer—viewpoint of the iarc working group. *New England journal of medicine*, 375(8):794–798, 2016.

[11] Simona Iodice, Sara Gandini, Patrick Maisonneuve, and Albert B Lowenfels. Tobacco and the risk of pancreatic cancer: a review and meta-analysis. *Langenbeck's archives of surgery*, 393 (4):535–545, 2008.

[12] Karoline B Kuchenbaecker, John L Hopper, Daniel R Barnes, Kelly-Anne Phillips, Thea M Mooij, Marie-José Roos-Blom, Sarah Jervis, Flora E Van Leeuwen, Roger L Milne, Nadine Andrieu, et al. Risks of breast, ovarian, and contralateral breast cancer for brca1 and brca2 mutation carriers. *Jama*, 317(23):2402–2416, 2017.

[13] Aung Ko Win, Mark A Jenkins, James G Dowty, Antonis C Antoniou, Andrew Lee, Graham G Giles, Daniel D Buchanan, Mark Clendenning, Christophe Rosty, Dennis J Ahnen, et al. Prevalence and penetrance of major genes and polygenes for colorectal cancer. *Cancer Epidemiology, Biomarkers & Prevention*, 26(3):404–412, 2017.

[14] Lorelei A. Mucci, Jacob B. Hjelmborg, Jennifer R. Harris, Kamila Czene, David J. Havelick, Thomas Scheike, Rebecca E. Graff, Klaus Holst, Sören Möller, Robert H. Unger, Christina McIntosh, Elizabeth Nuttall, Ingunn Brandt, Kathryn L. Penney, Mikael Hartman, Peter Kraft, Giovanni Parmigiani, Kaare Christensen, Markku Koskenvuo, Niels V. Holm, Kauko Heikkilä, Eero Pukkala, Axel Skytthe, Hans-Olov Adami, Jaakko Kaprio, and for the Nordic Twin Study of Cancer (NorTwinCan) Collaboration. Familial Risk and Heritability of Cancer Among Twins in Nordic Countries. *JAMA*, 315(1):68–76, January 2016. ISSN 0098-7484. doi: 10.1001/jama.2015.17703. URL `https://doi.org/10.1001/jama.2015.17703`.

[15] Amit Sud, Clare Turnbull, and Richard Houlston. Will polygenic risk scores for cancer ever be clinically useful? *npj Precision Oncology*, 5(1):40, May 2021. ISSN 2397-768X. doi: 10.1038/s41698-021-00176-1. URL `https://www.nature.com/articles/s41698-021-00176-1`. Publisher: Nature Publishing Group.

[16] Cathryn M. Lewis and Evangelos Vassos. Polygenic risk scores: from research tools to clinical instruments. *Genome Medicine*, 12(1):44, May 2020. ISSN 1756-994X. doi: 10.1186/s13073-020-00742-5. URL `https://doi.org/10.1186/s13073-020-00742-5`.

[17] Lisa Bastarache, Rory J. Tinker, Bryce A. Schuler, Lucas Richter, John A. Phillips, William W. Stead, Gillian W. Hooker, Josh F. Peterson, and Douglas M. Ruderfer. Characterizing trends in clinical genetic testing: A single-center analysis of EHR data from 1.8 million patients over two decades. *The American Journal of Human Genetics*, 112(5):1029–1038, May 2025. ISSN 0002-9297, 1537-6605. doi: 10.1016/j.ajhg.2025.03.009. URL `https://www.cell.com/ajhg/abstract/S0002-9297(25)00108-9`. Publisher: Elsevier.

[18] Ian J Jacobs, Usha Menon, Andy Ryan, Aleksandra Gentry-Maharaj, Matthew Burnell, Jatinderpal K Kalsi, Nazar N Amso, Sophia Apostolidou, Elizabeth Benjamin, Derek Cruickshank, et al. Ovarian cancer screening and mortality in the uk collaborative trial of ovarian cancer screening (ukctocs): a randomised controlled trial. *The Lancet*, 387(10022):945–956, 2016.

[19] Deborah Levine, Douglas L Brown, Rochelle F Andreotti, Beryl Benacerraf, Carol B Benson, Wendy R Brewster, Beverly Coleman, Paul DePriest, Peter M Doubilet, Steven R Goldstein, et al. Management of asymptomatic ovarian and other adnexal cysts imaged at us: Society of radiologists in ultrasound consensus conference statement. *Radiology*, 256(3):943–954, 2010.

[20] Sepideh Saadatmand, Reini Bretveld, Sabine Siesling, and Madeleine MA Tilanus-Linthorst. Influence of tumour stage at breast cancer detection on survival in modern times: population based study in 173 797 patients. *Bmj*, 351, 2015.

[21] Gustav K von Schulthess, Hans C Steinert, and Thomas F Hany. Integrated pet/ct: current applications and future directions. *Radiology*, 238(2):405–422, 2006.

[22] Centers for Disease Control and Prevention. Breast cancer screening guidelines for women, 2024. URL `https://www.cdc.gov/breast-cancer/about/index.html`. NOVEMBER 5, 2024.

[23] Joshua D Cohen, Lu Li, Yuxuan Wang, Christopher Thoburn, Bahman Afsari, Ludmila Danilova, Christopher Douville, Ammar A Javed, Fay Wong, Austin Mattox, et al. Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science*, 359(6378): 926–930, 2018.

[24] Ellen Heitzer, Imran S Haque, Charles ES Roberts, and Michael R Speicher. Current and future perspectives of liquid biopsies in genomics-driven oncology. *Nature Reviews Genetics*, 20(2): 71–88, 2019.

[25] David C Grossman, Susan J Curry, Douglas K Owens, Michael J Barry, Karina W Davidson, Chyke A Doubeni, John W Epling, Alex R Kemper, Alex H Krist, Ann E Kurth, et al. Screening for ovarian cancer: Us preventive services task force recommendation statement. *Jama*, 319(6): 588–594, 2018.

[26] Andrew MD Wolf, Elizabeth TH Fontham, Timothy R Church, Christopher R Flowers, Carmen E Guerra, Samuel J LaMonte, Ruth Etzioni, Matthew T McKenna, Kevin C Oeffinger, Ya-Chen Tina Shih, et al. Colorectal cancer screening for average-risk adults: 2018 guideline update from the american cancer society. *CA: a cancer journal for clinicians*, 68(4):250–281, 2018.

[27] Avram M Cooperman, Mazen E Iskandar, Michael G Wayne, and Justin G Steele. Prevention and early detection of pancreatic cancer. *Surgical Clinics*, 98(1):1–12, 2018.

[28] Florent Mouliere, Dineika Chandrananda, Anna M Piskorz, Elizabeth K Moore, James Morris, Lise Barlebo Ahlborn, Richard Mair, Teodora Goranova, Francesco Marass, Katrin Heider, et al. Enhanced detection of circulating tumor dna by fragment size analysis. *Science translational medicine*, 10(466):eaat4921, 2018.

[29] Matthias Mann and Ole N Jensen. Proteomic analysis of post-translational modifications. *Nature biotechnology*, 21(3):255–261, 2003.

[30] Jürgen Cox and Matthias Mann. Quantitative, high-resolution proteomics for data-driven systems biology. *Annual review of biochemistry*, 80(1):273–299, 2011.

[31] Aleksandra A Petelski, Edward Emmott, Andrew Leduc, R Gray Huffman, Harrison Specht, David H Perlman, and Nikolai Slavov. Multiplexed single-cell proteomics using scope2. *Nature protocols*, 16(12):5398–5425, 2021.

[32] Bogdan Budnik, Hossein Amirkhani, Mohammad H Forouzanfar, and Ashkan Afshin. Novel proteomics-based plasma test for early detection of multiple cancers in the general population. *BMJ oncology*, 3(1):e000073, 2024.

[33] Maria Bueno Alvez, Fredrik Edfors, Kalle von Feilitzen, Martin Zwahlen, Adil Mardinoglu, Per-Henrik Edqvist, Tobias Sjöblom, Emma Lundin, Natallia Rameika, Gunilla Enblad, et al. Next generation pan-cancer blood proteome profiling using proximity extension assay. *Nature Communications*, 14(1):4308, 2023.

[34] Wei Feng, Joanne C Beer, Qinyu Hao, Ishara S Ariyapala, Aparna Sahajan, Andrei Komarov, Katie Cha, Mason Moua, Xiaolei Qiu, Xiaomei Xu, et al. Nulisa: a proteomic liquid biopsy platform with attomolar sensitivity and high multiplexing. *Nature communications*, 14(1):7238, 2023.

[35] Yun Huang, Karolina Sulek, Sara E Stinson, Louise Aas Holm, Min Kim, Kajetan Trost, Kourosh Hooshmand, Morten Asp Vonsild Lund, Cilius E Fonvig, Helene Bæk Juel, et al. Lipid profiling identifies modifiable signatures of cardiometabolic risk in children and adolescents with obesity. *Nature Medicine*, 31(1):294–305, 2025.

[36] Shashi Gupta, Matthew J Westacott, Deborah G Ayers, Sophie J Weiss, Penn Whitley, Christopher Mueller, Daniel C Weaver, Daniel J Schneider, Anis Karimpour-Fard, Lawrence E Hunter, et al. Plasma proteome of growing tumors. *Scientific Reports*, 13(1):12195, 2023.

[37] Deb Schrag, Tomasz M Beer, Charles H McDonnell, Lincoln Nadauld, Christina A Dilaveri, Robert Reid, Catherine R Marinac, Karen C Chung, Margarita Lopatin, Eric T Fung, et al. Blood-based tests for multicancer early detection (pathfinder): a prospective cohort study. *The Lancet*, 402(10409):1251–1260, 2023.

[38] Brian D Nicholson, Jason Oke, Pradeep S Virdee, Dean A Harris, Catherine O'Doherty, John ES Park, Zaed Hamady, Vinay Sehgal, Andrew Millar, Louise Medley, et al. Multi-cancer early detection test in symptomatic patients referred for cancer investigation in england and wales (symplify): a large-scale, observational cohort study. *The Lancet Oncology*, 24(7):733–743, 2023.

[39] Elyse LeeVan, Amanda L Skarlupka, Christos Patriotis, Wendy S Rubinstein, Paul F Pinsky, Wade Bolton, Anthony Dickherber, Daniel C Edelman, Lyndsay N Harris, Hormuzd A Katki, et al. Framework to select multi-cancer detection assays in the national cancer institute's vanguard study. *Cancer Epidemiology, Biomarkers & Prevention*, pages OF1–OF7, 2025.

[40] Chyke A Doubeni, Douglas A Corley, Virginia P Quinn, Christopher D Jensen, Ann G Zauber, Michael Goodman, Jill R Johnson, Shivan J Mehta, Tracy A Becerra, Wei K Zhao, et al. Effectiveness of screening colonoscopy in reducing the risk of death from right and left colon cancer: a large community-based study. *Gut*, 67(2):291–298, 2018.

[41] Richard C Wender, Otis W Brawley, Stacey A Fedewa, Ted Gansler, and Robert A Smith. A blueprint for cancer screening and early detection: Advancing screening's contribution to cancer control. *CA: a cancer journal for clinicians*, 69(1):50–79, 2019.

[42] Mingyang Song and Edward Giovannucci. Preventable incidence and mortality of carcinoma associated with lifestyle factors among white adults in the united states. *JAMA oncology*, 2(9): 1154–1161, 2016.

[43] Cheryl L Rock, Cynthia Thomson, Ted Gansler, Susan M Gapstur, Marjorie L McCullough, Alpa V Patel, Kimberly S Andrews, Elisa V Bandera, Colleen K Spees, Kimberly Robien, et al. American cancer society guideline for diet and physical activity for cancer prevention. *CA: a cancer journal for clinicians*, 70(4):245–271, 2020.

[44] Barbara Ratner and Anna Bonislawski. *Early Detection: Catching Cancer When It's Curable*. OR Books, New York, 2024. ISBN 9781682193525.

[45] Joshua C Mandel, David A Kreda, Kenneth D Mandl, Isaac S Kohane, and Rachel B Ramoni. Smart on fhir: a standards-based, interoperable apps platform for electronic health records. *Journal of the American Medical Informatics Association*, 23(5):899–908, 2016.

[46] Michael Lane and M Marie. The adoption of single sign-on and multifactor authentication in organisations–a critical evaluation using toe framework. *Information in Motion*, 7:161, 2010.

[47] Mukund Sundararajan, Ankur Taly, and Qiqi Yan. Axiomatic attribution for deep networks. In *International conference on machine learning*, pages 3319–3328. PMLR, 2017.

[48] Sheila R Reddy, Michael S Broder, Eunice Chang, Caleb Paydar, Karen C Chung, and Anuraag R Kansal. Cost of cancer management by stage at diagnosis among medicare beneficiaries. *Current Medical Research and Opinion*, 38(8):1285–1294, 2022.

[49] George W Torrance. Measurement of health state utilities for economic appraisal: a review. *Journal of health economics*, 5(1):1–30, 1986.

[50] Liam Rose, Ganesh Rajasekar, Anjali Nambiar, Alexa Pohl, Kathryn J Ruddy, Katherine Arnow, Manali Patel, and Arden M Morris. Estimated out-of-pocket costs for patients with common cancers and private insurance. *JAMA Network Open*, 8(7):e2521575–e2521575, 2025.

[51] Davide Placido, Bo Yuan, Jessica X Hjaltelin, Chunlei Zheng, Amalie D Haue, Piotr J Chmura, Chen Yuan, Jihye Kim, Renato Umeton, Gregory Antell, et al. A deep learning algorithm to predict risk of pancreatic cancer from disease trajectories. *Nature medicine*, 29(5):1113–1122, 2023.

[52] Kai Jia, Steven Kundrot, Matvey B. Palchuk, Jeff Warnick, Kathryn Haapala, Irving D. Kaplan, Martin Rinard, and Limor Appelbaum. A pancreatic cancer risk prediction model (prism) developed and validated on large-scale us clinical data. *EBioMedicine*, 98:104888, 2023. doi: 10.1016/j.ebiom.2023.104888.

[53] Jiheum Park, Jason Patterson, Jose M Acitores Cortina, Tian Gu, Chin Hur, and Nicholas Tatonetti. Enhancing ehr-based pancreatic cancer prediction with llm-derived embeddings. *npj Digital Medicine*, 8(1):465, 2025.

[54] U.S. Food and Drug Administration. Clinical decision support software: Guidance for industry and food and drug administration staff. `https://www.fda.gov/regulatory-information/search-fda-guidance-documents/clinical-decision-support-software`, September 2022. Final guidance, Center for Devices and Radiological Health (CDRH).

[55] U.S. Food and Drug Administration. Marketing submission recommendations for a predetermined change control plan for artificial intelligence/machine learning (ai/ml)-enabled device software functions: Guidance for industry and food and drug administration staff. `https://www.fda.gov/regulatory-information/search-fda-guidance-documents/marketing-submission-recommendations-predetermined-change-control-plan-artificial-intelligen` April 2023. Final guidance, Center for Devices and Radiological Health (CDRH).

[56] U.S. Food, Drug Administration, Health Canada, Medicines, and Healthcare products Regulatory Agency. Predetermined change control plan (pccp) for artificial intelligence/machine learning (ai/ml)-enabled medical devices: Guiding principles. `https://www.fda.gov/medical-devices/software-medical-device-samd/predetermined-change-control-plan-guiding-principles`, 2023. Joint principles published by FDA, Health Canada, and UK MHRA; accessed March 25, 2025.

[57] U.S. Food, Drug Administration, Health Canada, Medicines, and Healthcare products Regulatory Agency. Good machine learning practice for medical device development: Guiding principles. `https://www.fda.gov/medical-devices/software-medical-device-samd/good-machine-learning-practice-medical-device-development-guiding-principles`, 2021. Joint guidance by the FDA, Health Canada, and UK MHRA; published in October 2021, accessed on March 25, 2025.

[58] Kaiser Family Foundation. Hospitals by ownership type. `https://www.kff.org/state-health-policy-data/state-indicator/hospitals-by-ownership/`, 2023.

[59] Zachary Levinson, Jamie Godwin, and Tricia Neuman. Hospital margins rebounded in 2023, but rural hospitals and those with high medicaid shares were struggling more than others. kff. `https://www.kff.org/health-costs/hospital-margins-rebounded-in-2023-but-rural-hospitals-and-those-with-high-medicaid-shares-` 2024.

[60] Daniel P Kessler and William Wygal. Non-profit hospital governance, conduct, and ceo pay. *INQUIRY: The Journal of Health Care Organization, Provision, and Financing*, 62: 00469580251366975, 2025.

[61] Dilek Anuk, Seref Bugra Tuncer, Mine Özkan, and Hülya Yazıcı. Factors associated with psychological distress during genetic counseling in high-risk women with breast cancer in turkey. *Supportive Care in Cancer*, 32(6):359, 2024.

[62] Zhuoer Xie, Neil Wenger, Annette L Stanton, Karen Sepucha, Celia Kaplan, Lisa Madlensky, David Elashoff, Jacqueline Trent, Antonia Petruse, Liliana Johansen, et al. Risk estimation, anxiety, and breast cancer worry in women at risk for breast cancer: A single-arm trial of personalized risk communication. *Psycho-oncology*, 28(11):2226–2232, 2019.

Supplementary Material

A Stakeholders, governance, and adoption pathway

While adoption of AI driven risk stratification has overall benefits for the healthcare system; careful consideration of individual stakeholders is crucial to facilitate adoption of this approach. The payer, provider, and patient framework is used here to analyze key stakeholders, coupled with a discussion on regulators to provide a wholistic view of the landscape for adoption.

Regulators FDA. Risk models that only *inform* clinicians may qualify as *non-device* clinical decision support and fall outside FDA device regulation [54]. For AI/ML devices, the FDA now support *Predetermined Change Control Plans (PCCPs)* that pre-specify how the model may be updated (e.g., retraining, threshold changes) without submitting a new marketing application each time [55, 56]. In practice, a risk-stratification tool intended for prospective use should (i) document intended use and clinical claims, (ii) predefine performance targets and monitoring, and (iii) include a change-management plan aligned with FDA good machine learning practice [57].

Payers/Insurers. Stage at cancer diagnosis is a dominant determinant of total cost among medicare beneficiaries, first-year costs are several-fold higher for stage IV compared to stage I across multiple cancer types [48]. Earlier diagnosis also reduces patient out-of-pocket burden and financial toxicity [50]. For payers, risk-stratified surveillance is attractive if it demonstrably shifts stage at diagnosis and lowers cost per cancer detected and per early-stage case, relative to the status quo.

Self-pay patients with financial means may pursue screening regardless of risk group, while those without means may remain unable to do so. Because private and public insurers finance the majority of U.S. healthcare (30% private and 39% medicare/medicaid), payer adoption is essential for real-world impact.

Healthcare systems In the U.S., hospitals operate under different modes of ownership and incentives, approximately 58% are non-profit, 25% are for-profit, and the remainder are government-facilities [58]. For-profit hospitals typically report substantially higher operating margins that non-profits, allowing them to prioritize high-revenue services such as advanced oncology treatments [59]. While non-profit and public hospitals must emphasize community benefit, they still require positive operating margins to sustain infrastructure, innovation, and even executive compensation [60]. Across ownership types, the viability of risk-stratified surveillance is determined by diagnostic capacity, reimbursement arrangements, pathways for follow-up testing and treatment. If alert volume exceeds available capacity, it can create bottleneck or compromise care quality. By contrast, earlier diagnosis and reduced costs provide a compelling business case for surveillance programs.

Patients. For patients, stratified risk communication needs to be combined with transparency, shared decision-making, and accessible follow up options. Being flagged with high-risk can create anxiety and psychological burden, which requires careful framing and supportive counseling [61]. At the same time, actionable information can empower patients to pursue preventive measures such as lifestyle modification. When information is coupled with adequate counseling, cancer anxiety and psychological distress have been found to resolve, even decreasing below prior baselines [62]. Equitable implementation also requires addressing barriers such as transportation, co-payments, language access and the ability to take time off work, which disproportionately affect lower-income groups. Reporting patient-centered outcomes such as time to diagnosis, time to treatment initiation, patient reported experience can help ensure that model improve not only detection but also quality of care. Importantly, stratification programs should avoid increasing disparities by ensuring that outreach and follow-up are equally available across demographic and socioeconomic group.

B PDAC risk prediction training details

Transformer architecture consists of a multi-head self-attention (MHSA) layer, residual connections, and feed-forward layers. A time-to-event regression head maps the transformer output to cumulative risk probabilities across five intervals (0-6, 0-12, 0-24, 0-36, 0-60). The total loss is a sum of a binary cross-entropy across all intervals. Model training was performed on an NVIDIA A100 GPU, with an average training time of 12 hours for 100 epochs, monitoring convergence every five epochs. We use a single MHSA layer with 16 heads and dimensionality 64, ReLU activations, and an initial learning rate of 0.001. To handle irregular longitudinal sequences, positional encodings are based on

the difference between timestamp of each event and the last recorded event in the patient trajectory. Each patient trajectory is set to a maximum length of 350 codes, with padding applied to shorter sequences and truncation for longer ones. The model is evaluated on a retrospective held-out set of 1 million patients, including 1,000 PDAC cases.

C Attribution to identify top markers

We aggregate IG scores across all patient trajectories to obtain global attribution that highlight the events that contribute most to model predictions. For each trajectory, IG scores are computed per event and used to rank events within that patient. These ranks are then accumulated across cancer cases predicted as high risk, and events are sorted by their overall cumulative rank. The top ten medication and diagnosis codes are reported in the Figure S1.

	Diagnosis Codes
1	Essential (primary) hypertension
2	Non-insulin-dependent diabetes mellitus, Insulin-dependent diabetes mellitus, Other specified diabetes mellitus
3	Disorders of lipoprotein metabolism and other lipidaemias, Other metabolic disorders, Disorders of glycoprotein metabolism
4	Acute pancreatitis
5	Chronic ischaemic heart disease
6	Psoriasis, Parapsoriasis, Pityriasis rosea Other papulosquamous disorders, Other dermatitis
7	Benign neoplasm of other and ill-defined parts of digestive system
8	Other chronic obstructive pulmonary disease
9	Polyarthrosis, Other arthrosis, Coxarthrosis
10	Glaucoma, Other congenital malformations of eye

	Medication Codes
1	Lisinopril
2	Simvastatin
3	Hydrochlorothiazide
4	Aspirin
5	Metformin
6	Hydrocortisone
7	Nitroglycerin
8	Docusate
9	Insulin
10	Vitamin B12

Figure S1: Top ten most informative diagnoses and medications in pre-cancer patient trajectories determined using attribution score as discussed in Section C