White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of York.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/226933/

Version: Published Version

Article:
Mandrik, Olena, Whyte, S, Kunst, Natalia et al. (2025) Modelling the impact of Multi Cancer Early Detection tests: a review of natural history of disease models. Medical Decision Making. 1013–1024. ISSN: 1552-681X

https://doi.org/10.1177/0272989X251351639

Reuse
This article is distributed under the terms of the Creative Commons Attribution-NonCommercial (CC BY-NC) licence. This licence allows you to remix, tweak, and build upon this work non-commercially, and any new works must also acknowledge the authors and be non-commercial. You don't have to license any derivative works on the same terms. More information and the full terms of the licence here:
https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Check for updates

Original Research Article

MDM Medical Decision Making

Medical Decision Making
1–12
© The Author(s) 2025

Articlereuseguidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0272989X251351639
journals.sagepub.com/home/mdm
Sage

Modeling the Impact of Multicancer Early Detection Tests: A Review of Natural History of Disease Models

Olena Mandrik, Sophie Whyte, Natalia Kunst, Annabel Rayner, Melissa Harden, Sofia Dias, Katherine Payne, Stephen Palmer, and Marta O. Soares

Introduction. The potential for multicancer early detection (MCED) tests to detect cancer at earlier stages is currently being evaluated in screening clinical trials. Once trial evidence becomes available, modeling will be necessary to predict the effects on final outcomes (benefits and harms), account for heterogeneity in determining clinical and cost-effectiveness, and explore alternative screening program specifications. The natural history of disease (NHD) component will use statistical, mathematical, or calibration methods. This work aims to identify, review, and critically appraise the existing literature for alternative modeling approaches proposed for MCED that include an NHD component. **Methods.** Modeling approaches for MCED screening that include an NHD component were identified from the literature, reviewed, and critically appraised. Purposively selected (non-MCED) cancer-screening models were also reviewed. The appraisal focused on the scope, data sources, evaluation approaches, and the structure and parameterization of the models. **Results.** Five different MCED models incorporating an NHD component were identified and reviewed, alongside 4 additional (non-MCED) models. The critical appraisal highlighted several features of this literature. In the absence of trial evidence, MCED effects are based on predictions derived from test accuracy. These predictions rely on simplifying assumptions with unknown impacts, such as the stage-shift assumption used to estimate mortality impacts from predicted stage shifts. None of the MCED models fully characterized uncertainty in the NHD or examined uncertainty in the stage-shift assumption. **Conclusion.** There is currently no modeling approach for MCEDs that can integrate clinical study evidence. In support of policy, it is important that efforts are made to develop models that make the best use of data from the large and costly clinical studies being designed and implemented across the globe.

Highlights

- In the absence of trial evidence, published estimates of the effects of multicancer early detection (MCED) tests are based on predictions derived from test accuracy.
- These predictions rely on simplifying assumptions, such as the stage-shift assumption used to estimate mortality effects from predicted stage shifts. The effects of such simplifying assumptions are mostly unknown.
- None of the existing MCED models fully characterize uncertainty in the natural history of disease; none examine uncertainty in the stage-shift assumption.
- Currently, there is no modeling approach that can integrate clinical study evidence.

Corresponding Author
Marta O. Soares, Centre for Health Economics, University of York, Alcuin A Block, Heslington, York, UK; (marta.soares@york.ac.uk).

Keywords
Multicancer early detection; natural history of disease; models; calibration

Date received: March 4, 2025; accepted: May 16, 2025

Novel technologies have recently emerged that look for markers of cancer in blood, urine, saliva, or stool and have the potential to detect signals from multiple cancer types from a single sample. These are termed *multicancer early detection* (MCED) tests. Their use in screening asymptomatic persons has the potential to detect cancer at an earlier stage, when treatment is likely to be more effective and perhaps less costly.[1,2] However, policy makers have demanded evidence of mortality effects and a fuller examination of the potential harms and consequences of the test's imperfect accuracy (including of diagnostic resolution pathways), overdiagnosis, and the effect on existing screening programs.[3] The Galleri® test (GRAIL, Inc., Menlo Park, CA, USA) test is the blood multicancer test that is most advanced in the stage of clinical research, with a randomized clinical trial currently underway in the United Kingdom, the NHS-Galleri trial (NCT05611632), aiming to demonstrate the clinical effectiveness of the test in stage shifting advanced cancer in a population screening setting.[4]

To inform policy decisions on screening programs involving MCED tests, modeling will be required to 1) link evidence and predict expected effects over final outcomes (mortality, life expectancy, and quality-adjusted life-years), 2) appropriately reflect heterogeneity in the value of stage shifts across different cancer types to allow estimation of cost-effectiveness, and 3) allow alternative specifications for a screening program to be evaluated (e.g., different age and risk groups, alternative screening intervals, etc.). Modeling is therefore likely to underpin such policy evaluations of MCED tests. This may include statistical, mathematical, or calibration modeling to integrate cancer-screening data and infer the natural history of disease (NHD). It may also include decision modeling to predict results with alternative screening regimens and their longer-term clinical and cost-effectiveness.

Cancer-screening models typically include an NHD component that describes the prevalence of preclinical cancer (undiagnosed but detectable) and allows for examining the effect of important policy options, such as alternative specifications for the screening program. The NHD model component describes cancer progression through its preclinical stages over time (in the absence of the proposed screening test) and may also consider cancer onset and the competing risks of clinical detection (both incidental findings and symptomatic presentation) and mortality. The challenge in evaluating these NHD models arises from the fact that preclinical progression is unobserved. Empirical data, however, can still provide relevant information on preclinical cancer prevalence and progression supporting inference, where the data are used to infer the NHD model and help gain an understanding of the likely values of the NHD model parameters in the underlying population, using statistical and mathematical approaches[5] or calibration.[6,7] Besides alternative evaluation approaches, models in the general cancer screening literature[8,9] also use a variety of data sources and analytical methodologies, vary the core elements of the NHD that are modeled (they may or may not model cancer onset, the likelihood of clinical detection, and/or mortality), and vary whether and how within-tumor heterogeneity and overdiagnosis are modeled.

The objective of this article is to identify, review, and critically appraise the existing literature for alternative modeling approaches proposed for MCED that include an NHD component. As the literature and approaches in this area continue to develop and evolve, it is important to critically examine the range of modeling approaches that have been proposed for MCEDs and to assess the extent to which specific features of model structure and model evaluation can accommodate the complexity of multicancer modeling. While there has been extensive discussion and consideration of the appropriate study design to inform clinical utility,[10] we are not aware of any publications that have attempted to systematically identify and

Sheffield Centre for Health and Related Research, University of Sheffield, Sheffield, UK (OM, SW, AR); Centre for Health Economics, University of York, York, UK (NK, SP, MOS); Centre for Reviews and Dissemination, University of York, York, UK (MH, SD); Manchester Centre for Health Economics, School of Health Sciences, The University of Manchester, UK (KP). The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article. The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Financial support for this study was provided by NHS England. The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report. KP is supported by the National Institute for Health and Care Research (NIHR) Manchester Biomedical Research Centre (BRC) (NIHR203308).

Box 1 Glossary

Stage shift: change in the stage distribution attributed to screening

Sojourn time, *time to transition*, and *dwell time: Sojourn time* refers to the time spent in preclinical cancer, that is, the total time from cancer onset to clinical diagnosis or death without diagnosis. Cancer onset is defined as the time when cancer is potentially detectable by any medical test. This ensures that, for any particular analysis, sojourn time is independent of the individual screening tests. Sojourn time for a particular cancer stage is the time spent in that preclinical stage of cancer; specifically, it is the time until progression to the next stage, clinical detection, or death (whichever first).

Natural history of disease (NHD) models may be parameterized by using distributions that describe the times to each individual transition allowed in the model, for example, time for early cancer in the preclinical stage to progress to advanced preclinical cancer or time for preclinical cancer to be clinically detected.

Dwell time has been used in the literature to reflect time to stage progression, given the cancer does not get clinically detected at that stage or the individual does not die from other causes at that stage. Note that, because GRAIL models model only individuals who would be clinically diagnosed with cancers under current care, the term *dwell time* can be used interchangeably to represent sojourn time.

Inference: Inference (statistical) is the process of forming judgments about the parameters of a population on the basis of data obtained from (usually random) sampling. For NHD modeling, inference is often sought on multiple (populational) parameters describing the NHD. Inference uses data to gain an understanding of the likely values of the NHD model parameters in the underlying population.

Model identifiability: Identifiability is achieved when the number of observed quantities (the number of screen-detected and interval cancers across different screens) is larger than the number of model parameters.

Correlation in progression parameters: Uncorrelated (or independent) parameters describing the progression between stages assume that the time it takes for a cancer to progress between stage 1 and 2 is independent of the time it takes for the same cancer to progress between stages 2 and 3. These quantities may also be assumed correlated, meaning that a cancer with a lower time to progression between stage 1 and stage 2 would be expected to also present a lower time to progression in subsequent transitions. Correlation or independence can also apply to sojourn time.

Length time bias: Length bias occurs when tumors with a longer sojourn time (slow growing), and more likely to be screen detected, present a better prognosis. This implies that faster-progressing cancers (shorter sojourn time), and more likely to be missed, present worse prognosis. In this case, screen-detected cases may appear to present a survival benefit that is due only to the heterogeneity in detected cases.

Stage-shift assumption: This means that cases shifted to an earlier stage via screening are assumed to have the same survival as cases detected in an earlier stage without screening.

Lead-time assumption: Lead time is defined as the time between when a cancer is detected by screening and when it would have been detected without screening. The term *lead-time assumption* means mortality is not considered during lead time, and therefore, bringing forward diagnosis through screening does not bring forward harms such as those from more aggressive treatment.

Cancer overdiagnosis: We define overdiagnosis as the diagnosis, from screening, of a cancer that would not have been diagnosed under current care.

critique existing modeling approaches and specifically the extent to which they will be able appropriately integrate the findings of these clinical utility studies. The article is structured in the following way: the existing models are identified and described in the "Review of Models" section and critically appraised in the "Critical Appraisal" section, and the overall findings are discussed in the "Discussion" section. Box 1 provides a glossary of definitions that will be used throughout.

Review of Models

Methods

Literature search for MCED models. A scoping literature review was developed and undertaken to identify published models of MCEDs in relation to a comparator. The review included models of NHD that incorporate both detection rates and predicted stage distribution (stage shift) and that may also have extended these

models to quantify effects on mortality. The search methodology is reported in full in the online appendix.

Additional selected models. To further support the critical appraisal, additional selected models were added purposively. We included 2 multidisease (but non-MCED) models that were contemporary examples developed by elements of our research team (and were therefore understood in depth). We also included 2 single-disease models that were cited by relevant authors to support existing MCED models. For these additional models, the focus of the review was on the modeling mechanisms and related assumptions. These models provide background and context for the modeling assumptions made in MCED models and for any changes/extensions made.

Extraction. The review extracted information on the following aspects: 1) model structure, including the number and types of cancer, NHD parameterization, and modeling of screening impact; 2) how data were used within the models, including key NHD assumptions and data requirements; and 3) uncertainties related to NHD and how these were considered. From the extracted information, we identified the following features of the models reviewed:

- Scope: The population modeled (whether only individuals with clinically diagnosed cancer were modeled or the entire population eligible for screening, which would have allowed quantifications of overdiagnosis) and whether mortality effects were considered.
- Key data sources: Whether evidence on detection with screening was considered (e.g., clinical trial) or only evidence on cancer incidence under current care, and whether external evidence on preclinical progression parameters (elicited or from the literature) was used.
- Evaluation: Whether the model evaluation was based on prediction or on inference and whether it is evaluated at the cohort or individual level.

A predictive approach uses input evidence to directly describe model parameters and calculates expected cancer detection algebraically, with and without screening. NHD model parameters are prespecified using values or distributions (using external sources such as other evaluations or expert opinion) before running the model, which then outputs predictions. In contrast, inferential approaches use cancer diagnosis data from samples of individuals (for example, repeat screening data) to learn about NHD model parameters. Because sojourn time is not directly observable, the methods used differ from standard regressions and employ mathematical techniques such as deconvolution[11] or calibration.[6,7,12]

- Structure and parameterization: whether a common structure across cancer types is used; what level of disaggregation of cancer stages was used (i.e., whether individual stages were considered or whether they were aggregated, e.g., early v. late cancer); whether the impact of screening is predicted from test accuracy; whether mortality effect is predicted by applying mortality in clinically detected cancer to the screening stage distribution predicted by the model; what parameterization, distributional assumptions, and assumptions about the correlation between progression parameters were used; whether overdiagnosis (definition in Box 1) is quantified within the NHD model.

Results

The review identified 5 different MCED models with an NHD component: 4 funded by GRAIL (hereafter termed *GRAIL models*) and specifically related to the Galleri test[13–16] and 1 based on a hypothetical MCED (although using some inputs derived from the Galleri test).[17] Four additional models (non-MCED) were also reviewed: 2 multicancer models by Thomas[18] and Mandrik et al.[19] and 2 single-cancer models by Pinsky[11] and Skates and Singer.[20] The 9 models were reviewed; these are described in Table 1.

MCED models. Table 2 describes the key features of the models reviewed. These models are referred to by the name of the first author in the publication.

There are 4 GRAIL models: Hubbell, Sasieni, Tafazzoli, and Dai.[13–16] These use a common approach, referred to as the *interception model*, to determine the NHD and stage shift with the Galleri test, with the core methodology rooted in the Hubbell model. These also use a common set of evidence, including national cancer incidence statistics (by type, stage, age, and gender), expert or literature-derived sojourn time evidence,[13,21] and test sensitivity from diagnostic studies.

The NHD component of the GRAIL models focuses on individuals clinically diagnosed with cancer under standard care. The GRAIL models use a common NHD structure, assuming (Table 3) 1) disease progression across 4 stages (stages 1, 2, 3, and 4) without regression; 2) progression is sequential, with cancers moving through each stage until clinically detected; 3) sojourn times are exponentially distributed; 4) sojourn times are

Table 1 List of Models Reviewed

Model	Technology	*N* Cancers Modeled	Outcomes
GRAIL models			
Hubbell 2021[14]	Galleri test	19	Clinical
Sasieni 2023[15]	Galleri test	24	Clinical
Tafazzoli 2022[16]	Galleri test	23	Clinical and cost-effectiveness
Dai 2024[13]	Galleri test	25	Clinical
Other MCED			
Lange 2024[17]	Hypothetical MCED (based on Galleri test)	12	Clinical
Other multicancer			
Mandrik 2025[19]	Dipstick test for bladder and kidney cancer	2	Clinical and cost-effectiveness
Thomas 2025[18]	Imaging test for abdominal cancers	10	Clinical and cost-effectiveness
Other single-cancer			
Skates 1991[20]	Blood test (CA 125) for ovarian cancer	1	Clinical
Pinsky 2004[11]	CT screening for lung cancer	1	Clinical

CT, computed tomography; MCED, multicancer early detection.

independent between stages; 5) mean sojourn times may differ across cancer types (models consider groups of cancers with distinct mean sojourn times); and 6) there is no heterogeneity in sojourn times within tumor types beyond that expected by chance (i.e., the expected value of the sojourn time is equal for all individuals in the model). The NHD does not include the probability of cancer onset nor of clinical detection, and therefore, the main NHD parameters are the stage-specific sojourn, or dwell, times (see Box 1 for definitions).

All 4 GRAIL models consider stage shift as the main clinical benefit of screening. Stage shift is evaluated predictively, using test sensitivity to determine the likelihood of earlier detection. Mortality effects are also predicted under the stage-shift assumption and the lead-time assumption (Box 1), except in a scenario of the Tafazzoli model, which considers mortality during lead time.

None of the GRAIL models reviewed evaluated uncertainty probabilistically. In the context of predictive modeling, this would have entailed describing uncertainty in the input parameters and running probabilistic analysis to evaluate uncertainty over models' outputs. Also, models incorporate within-tumor heterogeneity only from the distribution of cancer diagnosis by age and sex. Further consideration of these aspects is provided in the "Critical Appraisal" and "Discussion" sections.

Although the GRAIL models are underpinned by the same core methodology proposed in Hubbell, there are a number of specific differences in terms of their parameterization and structural assumptions. The Sasieni model[15,22] applies the Hubbell model to UK cancer incidence and mortality data and examines structural extensions allowing consideration of differential survival of cfDNA-detectable cancers, alternative cohorts and screening regimens, and the possibility of nonsequential progression from stage I to IV only. Tafazzoli's model[16] integrates Hubbell's stage-shift matrices (i.e., the likelihood of a cancer clinically detected in a particular stage being detected by Galleri at each earlier cancer stage) within a cohort model of 50-y-old individuals tested annually with Galleri until the age of 79 y. In Tafazzoli's model, stage-shifted individuals in each model cycle are time shifted (shifted back in time to earlier cycles to account for an earlier time of diagnosis), based on cancer-specific sojourn times. Tafazzoli is the only Galleri model that incorporates overdiagnosis (but not explicitly in the NHD model) by increasing detection by a proportion that is applied as an input to the model and extends the evaluation to cost-effectiveness. Dai's model[13] uses the core assumptions of Hubbell's model but evaluates the model using individual patient simulation. It also describes sojourn times from empirically derived estimates sourced from other screening studies, rather than elicitation.

Our review identified only 1 MCED model that was not funded by GRAIL: this is Lange's model.[17] This model examines the impact of a hypothetical MCED (using the estimates for test sensitivity that are relevant to Galleri) on 12 cancer types. The model does not evaluate overdiagnosis or mortality (extensions to mortality have been further considered since publication, see https://cedarmodelingframework.shinyapps.io/mcedmodel/). It is based on the same type of evidence as the GRAIL models (age- and stage-specific clinical incidence data under current care) but applies an alternative NHD model that is more comprehensive in that it, in addition to preclinical progression (for which it uses the more aggregate classification of early v. late disease), also characterizes the probability of cancer onset, the age of cancer onset, and

Table 2 Key Features of the Natural History of Disease (NHD) Models Reviewed

	Scope		Structural			NHD Model Evaluation			Evidence for NHD		Uncertainty
Model	**Population Modeled in the NHD**	**Mortality Effects Included?**	**Common Structure across Cancer Types?**	**Disease Stages in NHD?**	**Clinically Diagnosed Cancer Mortality Used on Screen-Detected Cases?**	**Cohort Model (v. IPL)**	**Approach, NHD**	**Approach, Comparative Screening Outcomes**	**Detection Data on Screening?**	**External Preclinical Progression Evidence?**	**Uncertainty Evaluated (above Individual Variability, Where Relevant)?**
GRAIL models											
Hubbell 2020[14]	Incident	Yes	Yes	1, 2, 3, and 4	Yes	Cohort	Prediction	Prediction	No	Yes	No
Sasieni 2023[15]	Incident	Yes	Yes	1, 2, 3, and 4	Yes	Cohort	Prediction	Prediction	No	Yes	No
Tafazzoli 2022[16]	Incident	Yes	Yes	1, 2, 3, and 4	Yes	Cohort	Prediction	Prediction	No	Yes	No
Dai 2024[13]	Incident	Yes	Yes	1, 2, 3, and 4	Yes	IPL	Prediction	Prediction	No	Yes	No
Other MCED											
Lange 2024[17]	All	No	Yes	Early v. late	NA	Cohort	Inference, ML	Prediction	No	Yes	Yes
Other multidisease											
Mandrik 2025[19]	All	Yes	No	1, 2, 3, and 4	Yes	IPL	Inference, Bayesian calibration	Prediction	No	Yes (within priors)	Yes
Thomas 2025[18]	Screen detected	Yes	Yes	1, 2, 3, and 4	Yes	Cohort, multiple	Inference, calibration	Prediction	Yes, cases of cancer detected with screening; no data in the absence of screening	Yes	No
Other single disease											
Skates 1991[20]	Incident	Yes	NA	Early v. late	Yes	IPL	Prediction	Prediction	No	Yes	No
Pinsky 2004[11]	All	No	NA	Early v. late	NA	Cohort	Inference, ML		Yes, cases of screen-detected and interval cancers by screening round	No	Yes

IPL, coefficient of variation; individual patient level; ML, maximum likelihood estimation; NA, not applicable; NHD, natural history of disease.

Table 3 Assumptions over the Natural History of Disease (NHD) Model

	Cancer Onset		Preclinical Progression Parameter						
NHD Model Element	**Distribution**	**Heterogeneity**	**Parameterization**	**Distributions**	**Correlation?**	**Probability of Clinical Identification**	**Mortality Included in NHD Model**	**Overdiagnosis Included in NHD Model**	**Other**
GRAIL models	Not modeled		Individual parameters for progression between stages	Exponential	No	Not modeled	Yes, predictive	No	NA
Other MCED: Lange	Hypoexponential (fixed parameter m)	No	Fixed values for time in overall and late-stage preclinical disease	Exponential	No	Exponential	No	No	Clinical detection rates described by Poisson distribution as part of inference
Other multidisease: Mandrik	Annual probability as a function of age and other risk factors (cohort model component)	Yes	Individual parameters for progression between stages, using assumptions, informative priors, and constraints to ensure identification	Weibull (IPL model component)	Yes	Annual probability (cohort model component)	Yes	Yes	Bayesian calibration with multiple targets
Other multidisease: Thomas	Not modeled		Individual parameters for progression between stages	Triangular	No	Yes, for comparator arm; triangular distribution	Yes, predictive	Yes, by considering competing mortality	NA
Other single disease: Skates	Not modeled		Four-variate normal distribution; ratio of time in early v. late stage and constant CV for each stage were assumed constant	Four-variate normal distribution	Yes	Not modeled	Yes, predictive	No	NA
Other single disease: Pinsky	Cubic polynomial function of age	No	Single parameter for progression between early and late	Weibull (exponential as special case)	No	Weibull (exponential as special case)	Yes, predictive and not used in NHD model inference	Yes, predicted using NHD model estimates and external mortality estimates	NA

CV, coefficient of variation; IPL, individual patient level; MCED, multicancer early detection; ML, maximum likelihood estimation; NA, not applicable; NHD, natural history disease

the likelihood of clinical detection. Lange evaluates the underlying NHD model parameters using an inferential approach to describe clinical incidence rate data using a Poisson distribution; however, not all parameters of the model are identifiable based on age- and stage-specific incidence data. Therefore, given these data, the authors assumed fixed values for overall and late-stage sojourn times (user defined) allowing estimating all unknown parameters. It is unclear how inference over the early-stage sojourn times is, however, reached.

Selected non-MCED models

Selected multicancer screening models. Mandrik's model[19] examines the clinical and cost-effectiveness of a urine dipstick test in screening for bladder and kidney cancers. The NHD model structure includes cancer onset, preclinical cancer progression through cancer stages (1 to 4), cancer detection, and mortality. Heterogeneity is included by considering cancer onset to depend on age and smoking status and by considering a separate cancer pathway for nonfatal low-risk bladder cancers. Mandrik's model uses detection data for current care only (due to the absence of data for the screening test under evaluation) and summary evidence from the literature on the impact of risk factors, test sensitivity, and other elements. The model is Markovian for all transitions except for progression of preclinical cancers, which uses an individual patient time-to-event formulation. The NHD model was evaluated using Bayesian calibration (Metropolis–Hastings algorithm), an inferential calibration procedure that allows for uncertainty to be appropriately integrated. Due to the absence of screening data, and to ensure model identification, strong priors, assumptions, and constraints over the NHD parameters were used. A predictive approach anchored on test accuracy was used to project screening outcomes from test accuracy, overdiagnosis, and mortality impacts (from stage shifts)

Thomas' model[18] evaluates upper abdominal CT imaging for the screening of 10 cancers (alongside other abdominal diseases). It adopts a common structure across all cancers, with progression across stages 1 to 4. The model uses clinical incidence data with screening, combined with elicited estimates of test sensitivity. Despite not considering the probability of cancer onset, the model considers the age of onset in those who were screen detected. For the comparator arm, the model simulates what would have happened to the screen-detected individuals had they not been screened. In doing so, it considers the competing events of stage progression, clinical detection, and mortality in its structure. The model is a multicohort Markov model, considering various age and sex cohorts. The model conducts inference using a simplified non-Bayesian calibration (or fitting process), which does not consider uncertainty over the NHD, to evaluate outcomes for a cohort of unscreened individuals from elicited values describing stage-specific preclinical progression. Mortality impacts were predicted from stage shifts. By considering that those who would have been screen detected were at risk of death if unscreened, the comparator arm considers individuals dying with undiagnosed cancer and predicts a lower number of cancer cases than in the screening arm.

Single-cancer models cited by authors of existing MCED models. Skates' model[20] is cited in the GRAIL models in support of the proposed interception model. Skates examines the impact of screening for ovarian cancer with a blood biomarker using a predictive approach combining ovarian cancer incidence with preclinical progression times across 4 cancer stages. The key difference between this NHD model and the GRAIL NHD models are that Skates uses patient-level simulation (all GRAIL models except for Dai), a different parameterization of time to stage progression using log-normal distributions with fixed mean ratios between stages and a coefficient of variation, and accounts for correlation between stages. The impact of screening is predicted from biomarker levels, and the mortality impact is predicted using the stage-shift assumption and the lead time assumption while also assuming a proportion of patients are cured.

Pinsky's model[11] is the key reference cited by Lange. It uses the same structure as Lange but considers a range of distributions for the NHD parameters and imposes age dependency on time to cancer onset. Pinsky's model, however, uses screening trial data to achieve inference on the NHD via maximum likelihood estimation. In doing so, it carefully considers parameter identifiability from the data.

Critical appraisal. In this section, we critically appraise the existing MCED models for their key features, including how these accommodate the multicancer context, and highlight key uncertainties.

MCED effects are based on predictions rather than direct evidence. A critical feature of this evaluation problem is the current absence of data on cancer detection and mortality from screening with Galleri or other MCED. The NHD models therefore use similar data, namely, on cancer incidence data under current care and expected sojourn times, to back-calculate or infer undiagnosed cancer prevalence. The lack of screening data means that the accuracy of predictions and inferences in

MCED models will rely on the use of an appropriate NHD model and on the quality of the evidence underlying/supporting the NHD parameters.

MCED models apply simplifying assumptions. It is unclear where adding complexity may be most important. Existing MCED models, despite the similarity in the data included, have proposed a wide variety of modeling approaches for the NHD—from predictive to inferential models, cohort to individual-level simulation, more complex (or simpler) assumptions over the NHD. MCED models apply the most assumptions (see Table 3) despite many being shared with other models. This may be motivated by the multicancer context and the need to reduce parameterization and employ simpler evaluation approaches. There has been limited exploration of the impact of these simplifications, and it is unclear where additional complexity may add value.

Some of the simplifying assumptions allow the NHD to be evaluated algebraically as with the GRAIL models (i.e., exponentially distributed preclinical progression times with a common mean and independent across stages). However, models run as individual patient-level simulation, such as Skates and Mandrik, allow relaxing these assumptions and varying the level of variation (heterogeneity) in sojourn time, that is, the proportion of cases with extreme sojourn times. Existing explorations are insufficient to identify the likely sources and key impacts of heterogeneity but suggest important impacts (see, for example, Sasieni's scenario considering a proportion of very-fast-progressing cancers).

Overdiagnosis is not explicitly modeled in MCED models, and adding this may add complexity to modeling. One important potential harm of screening is overdiagnosis. Overdiagnosis has the potential to be explicitly estimated/predicted within an NHD model with a fuller structure that characterizes heterogeneity and includes cancer onset and mortality alongside preclinical progression and clinical detection. None of the MCED models have estimated/predicted overdiagnosis within the NHD model, presumably because of the reliance on a restricted structure and scope to allow evaluation from cancer incidence data; for example, Hubbell characterized only cancer progression and Lange also included cancer onset but not mortality. Of the broader models reviewed, those including a full NHD structure, such as Mandrik, included overdiagnosis, but none explicitly examined whether and how heterogeneity may affect overdiagnosis estimates.

Current MCED models do not appropriately characterize uncertainty in the NHD. Decisions in health are often made under uncertainty, and explicit descriptions of uncertainty help determine appropriate funding and research decisions. Uncertainty in model inputs can be described and propagated in prediction modeling; however, none of the predictive MCED models reviewed have done so. Since our review was conducted, GRAIL published an extension of the Tafazzoli model that includes probabilistic analysis,[23] although, in this analysis, none of the NHD parameters were assumed uncertain (e.g., sojourn times, mortality). Of the MCED models, only Lange considers uncertainty in the NHD by implementing an inferential procedure describing the cancer incidence data as uncertain. However, other important sources of uncertainty were not formally included in Lange's model, such as uncertainty over sojourn times, but can be examined by varying the choice of sojourn time inputs.

All models predict mortality effects using the stage-shift assumption. The stage-shift assumption is plausible only if cancers detected by a screening test do not differ systematically in their characteristics from clinically detected cancers. For example, if the higher ctDNA shedding expected in cancers detected by Galleri is associated with worse prognosis, the capacity of stage-shifted cancers to benefit may be smaller than expected. The Sasieni model examined hypothetical reductions in the capacity to benefit of stage-shifted cancers and showed that the effect can be significant. A number of publications have explored the accumulation of evidence, across screening trials, in support of the stage-shift assumption.[13,24,25] However, the validity of this assumption for particular multicancer tests is unknown until well-designed clinical research reports on the mortality effects. The NHS Galleri trial, at the time its primary endpoint reports, may not provide sufficient mortality evidence, and this is therefore likely to remain a key uncertainty for decision making.

Discussion

We identified, summarized, and critically appraised the NHD components of models of the clinical and/or economic impact of using MCED tests in a screening program and found that these models are characterized by the absence of screening data, by the limited use of inference, and by the limited characterization of uncertainty, heterogeneity, and overdiagnosis within the NHD. Our critical appraisal identified limitations of current MCED models and highlighted the limited exploration of the impact of modeling assumptions.

While recognizing the value of predictive models in anticipating future effects and making data-informed research and development decisions, our findings have important implications for the development of future models of the clinical and cost-effectiveness of MCED

screening programs used to inform clinical and policy decisions, which will need to incorporate clinical utility study evidence in support of these decisions. This requires an inferential approach, but, to date, no such approach has been developed to include screening data in the multicancer context. There is an extensive literature on inferential approaches used in the single-disease context, which includes 1) mathematical/statistical models that typically using a single main source of evidence and a clear specification of the model (NHD) with lower dimensionality (e.g., typically aggregating cancer stages for example) and 2) calibration models, typically using multiple sources of evidence (as calibration targets) and, perhaps for this reason, a higher dimensionality. In this article, we did not review this broader literature, but the future development of an inferential approach for MCEDs should draw on it.

MCED trials, like the NHS-Galleri trial, are likely to be powered on stage-shift outcomes aggregated over multiple cancer types, and estimates for each cancer type will need to be strengthened using modeling alongside additional external evidence. Model identifiability will need to carefully consider higher parameterizations (e.g., more detailed descriptions of between- and within-tumor heterogeneity) and the support of the evidence for structural simplifications in such descriptions and in the potential aggregation across cancer stages. GRAIL models disaggregate across the 4 cancer stages, but most mathematical approaches aggregate stages into early and advanced cancer or simply distinguish preclinical from clinical cancers. Uncertainty over the stage-shift assumption needs to be examined in further work in support of decision making. This should have appropriate consideration for the fact that screen-detected cancers may differ (in prognosis) from non–screen-detected cancers.

Computational burden is also of concern, as more complex models may compromise transparency and accessibility, particularly for calibration approaches, typically using individual-level simulation, applied in the multicancer context. Alternatives to individual-level simulation can be considered, such as the multicohort model structure exemplified in Thomas. It partitions the cohort into subcohorts based on relevant baseline characteristics, such as risk or demographic groups.

Other key considerations for future MCED model development relate to overdiagnosis and within-tumor heterogeneity. In what concerns overdiagnosis, there are important challenges in obtaining valid empirical estimates,[26] and therefore, decision making may initially need to consider estimates from modeling that require extensions to existing MCED modeling approaches (see the "Critical Appraisal" section). With regard to within-tumor heterogeneity, this is known to exist across several cancer types. Heterogeneity has been considered in the broader screening modeling literature structurally, for example, by adding states for indolent or slow-growing cancers[27] and in its contribution to overdiagnosis.[28] While describing heterogeneity depends on model specification,[29] it can lead to more accurate estimates but also increased uncertainty.[30,31] The NHS-Galleri trial will not provide characterization of within-tumor heterogeneity, so it is important to better understand its potential effects (on detection, overdiagnosis, and mortality), to support further evidence gathering in support of further model development.

Multicancer technologies are developing rapidly, and large and costly clinical studies are being designed and implemented across the globe. Recognizing the need to produce clinical and economic evidence suitable for consideration by committees deciding whether to introduce MCED-screening programs, it is important that similar efforts are made in the development of MCED models that make the best use of the available data and that the data required to fit those models from clinical studies are made widely available.

Acknowledgment

We thank GRAIL for providing access to their models and for related discussions.

Author Contributions

M. Harden designed the literature search strategy; N. Kunst conducted the search; O. Mandrik, A. Rayner, and S. Whyte extracted and summarized the models; M. O. Soares and S. Palmer conducted the critical appraisal; M. O. Soares drafted the article; all authors contributed to the write up and approved the final version.

Ethical Considerations

Not applicable.

Consent to Participate

Not applicable.

Patient Consent

Not applicable.

Consent for Publication

Not applicable.

ORCID iDs

Olena Mandrik https://orcid.org/0000-0003-3755-3031

Natalia Kunst https://orcid.org/0000-0002-2409-4246
Sofia Dias https://orcid.org/0000-0002-2172-0221
Marta O. Soares https://orcid.org/0000-0003-1579-8513

Data Availability

Not applicable.

Supplemental Material

Supplementary material for this article is available online at https://doi.org/10.1177/0272989X251351639.

References

1. McGarvey N, Bensink M, Jeske L, Louie KS, Navaratnam S. Increased healthcare costs by later stage cancer diagnosis. *BMC Health Serv Res*. 2022;22(1):1155.
2. Cancer Research UK. *Saving Lives, Averting Costs: An Analysis of the Financial Implications of Achieving Earlier Diagnosis of Colorectal, Lung and Ovarian Cancer*. London: Incisive Health and Cancer Research UK; 2014.
3. Turnbull C, Slavin T, Rahman N. GRAIL-Galleri: why the special treatment? *Lancet*. 2024;403(10425):431–2.
4. Neal RD, Din NU, Hamilton W, et al. Cell-free DNA-based multi-cancer early detection test in an asymptomatic screening population (NHS-Galleri): design of a pragmatic, prospective randomised controlled trial. *Cancers (Basel)*. 2022;14(19):4789.
5. Geurts SME, Massat NJ, Duffy SW. Quantifying the duration of the preclinical detectable phase in cancer screening: a systematic review. *Epidemiol Health*. 2022;44:e2022008.
6. Hazelbag CM, Dushoff J, Vijver Dvd, et al. Calibration of individual-based models to epidemiological data: a systematic review. *PLoS Comput Biol*. 2020;16(5):e1007893.
7. Stout NK, Goldie SJ, Etzioni R, et al. Calibration methods used in cancer simulation models and suggested reporting guidelines. *Pharmacoeconomics*. 2009;27(7):533–45.
8. Mendes D, Rabe KF, Caldas-de-Almeida JM, et al. Systematic review of model-based cervical screening evaluations. *BMC Cancer*. 2015;15:334.
9. Mandrik O, Purshouse K, Suijkerbuijk A, et al. Critical appraisal of decision models used for the economic evaluation of bladder cancer screening and diagnosis: a systematic review. *Pharmacoeconomics*. 2023;41(6):633–50.
10. Minasian LM, Lasonde B, Daly MB, et al. Study design considerations for trials to evaluate multicancer early detection assays for clinical utility. *J Natl Cancer Inst*. 2023;115(3):250–7.
11. Pinsky PF. An early- and late-stage convolution model for disease natural history. *Biometrics*. 2004;60(1):191–8.
12. Whyte S, Walsh C, Chilcott J. Bayesian calibration of a natural history model with application to a population model for colorectal cancer. *Med Decis Making*. 2011;31(4):625–41.
13. Dai JY, Hsu L, Moyer VA, Etzioni R. Clinical performance and utility: a microsimulation model to inform the design of screening trials for a multi-cancer early detection test. *J Med Screen*. 2024;31(3):140–9.
14. Hubbell E, Clarke CA, Aravanis AM, Berg CD. Modeled reductions in late-stage cancer with a multi-cancer early detection test. *Cancer Epidemiol Biomarkers Prev*. 2021;30(3):460–8.
15. Sasieni P, Shelton J, Bhatnagar G, et al. Modelled mortality benefits of multi-cancer early detection screening in England. *Br J Cancer*. 2023;129(1):72–80.
16. Tafazzoli A, Grigore B, Snowsill T, et al. The potential value-based price of a multi-cancer early detection genomic blood test to complement current single cancer screening in the USA. *Pharmacoeconomics*. 2022;40(11):1107–17.
17. Lange JM, Lewin A, Vennall GP, et al. Projecting the impact of multi-cancer early detection on late-stage incidence using multi-state disease modeling. *Cancer Epidemiol Biomarkers Prev*. 2024;33(6):830–7.
18. Thomas C. Cost-effectiveness of one-off upper abdominal CT screening as an add-on to lung cancer screening in England. *Br J Cancer*. Forthcoming 2025.
19. Mandrik O, Kellen E, Purshouse K, et al. Home urine dipstick screening for bladder and kidney cancer in high-risk populations in England: a microsimulation study of long-term impact and cost-effectiveness. *Pharmacoeconomics*. 2025;43(4):441–452.
20. Skates SJ, Singer DE. Quantifying the potential benefit of CA 125 screening for ovarian cancer. *J Clin Epidemiol*. 1991;44(4–5):365–80.
21. Schwartzberg L, Dhawan R, Kagan M, et al. Impact of early detection on cancer curability: a modified Delphi panel study. *PLoS One*. 2022;17(12):e0279227.
22. Sasieni P, Shelton J, Bhatnagar G, et al. Correction: modelled mortality benefits of multi-cancer early detection screening in England. *Br J Cancer*. 2024;131(12):1942–4.
23. Kansal AR, Liu D, Miller JD, et al. Cost-effectiveness of a multicancer early detection test in the US. *Am J Manag Care*. 2024;30(12):e352–8.
24. Feng X, Lin Y, Salazar MC, et al. Cancer stage compared with mortality as end points in randomized clinical trials of cancer screening: a systematic review and meta-analysis. *JAMA*. 2024;331(22):1910–7.
25. Owens L, Gulati R, Etzioni R. Stage shift as an endpoint in cancer screening trials: implications for evaluating multicancer early detection tests. *Cancer Epidemiol Biomarkers Prev*. 2022;31(7):1298–304.
26. Gulati R, Feuer EJ, Etzioni R. Conditions for valid empirical estimates of cancer overdiagnosis in randomized trials and population studies. *Am J Epidemiol*. 2016;184(2):140–7.
27. Weedon-Fekjaer H, Tretli S, Aalen OO. Estimating screening test sensitivity and tumour progression using tumour size and time since previous screening. *Stat Methods Med Res*. 2010;19(5):507–27.

28. Seigneurin A, Exbrayat C, Labarère J, et al. Overdiagnosis from non-progressive cancer detected by screening mammography: stochastic simulation study with calibration to population based registry data. *BMJ*. 2011;343:d7017.
29. Shen Y, Yang Y, Inoue LYT, Munsell MF, Miller AB. Estimating the frequency of indolent breast cancer in screening trials. *Stat Methods Med Res*. 2019;28(4):1261–71.
30. Uhry Z, Dejardin O, Grosclaude P, Bouvier AM, Launoy G. Multi-state Markov models in cancer screening evaluation: a brief review and case study. *Stat Methods Med Res*. 2010;19(5):463–86.
31. Welch HG, Black WC. Overdiagnosis in cancer. *J Natl Cancer Inst*. 2010;102(9):605–13.