JNCI: Journal of the National Cancer Institute, 2024, 116(2), 189–193

https://doi.org/10.1093/jnci/djad227

Advance Access Publication Date: November 6, 2023

Commentary

Commentary

Revisiting the standard blueprint for biomarker development to address emerging cancer early detection technologies

Ruth Etzioni, PhD,[1,*] Roman Gulati, MS,[1] Christos Patriotis, PhD,[2] Carolyn Rutter, PhD,[1] Yingye Zheng, PhD,[1] Sudhir Srivastava, PhD,[2] Ziding Feng, PhD[1]

[1]Division of Public Health Sciences, Fred Hutchinson Cancer Center, Seattle, WA, USA

[2]Division of Cancer Prevention, National Cancer Institute, Bethesda, MD, USA

*Correspondence to: Ruth Etzioni, PhD, Division of Public Health Sciences, Fred Hutchinson Cancer Center, 1100 Fairview Ave N, M2-B230, Seattle, WA 98109-1024, USA (e-mail: retzioni@fredhutch.org).

Abstract

Novel liquid biopsy technologies are creating a watershed moment in cancer early detection. Evidence supporting population screening is nascent, but a rush to market the new tests is prompting cancer early detection researchers to revisit the standard blueprint that the Early Detection Research Network established to evaluate novel screening biomarkers. In this commentary, we review the Early Detection Research Network's Phases of Biomarker Development (PBD) for rigorous evaluation of novel early detection biomarkers and discuss both hazards and opportunities involved in expedited evaluation. According to the PBD, for a biomarker-based test to be considered for population screening, 1) test sensitivity in a prospective screening setting must be adequate, 2) the shift to early curable stages must be meaningful, and 3) any stage shift must translate into clinically significant mortality benefit. In the past, determining mortality benefit has required lengthy randomized screening trials, but interest is growing in expedited trial designs with shorter-term endpoints. Whether and how best to use such endpoints in a manner that retains the rigor of the PBD remains to be determined. We discuss how computational disease modeling can be harnessed to learn about screening impact and meet the needs of the moment.

We are at a watershed moment in cancer early detection, with novel liquid biopsy technologies creating the possibility to screen for multiple cancers using a simple blood test. Studies to date have shown that such multicancer early detection tests can identify cancers in people who already have a diagnosis (1-3). Industry messaging and public anticipation are driving a perception that widespread deployment of the tests is an urgent matter [eg, Klein et al. (4)]; indeed, several products are being marketed ahead of United States Food and Drug Administration approval while we await evidence of clinical utility, and a congressional bill has been introduced to allow Medicare coverage of the new tests once they have been approved and shown to be clinically effective (5,6). There is now growing concern among early detection researchers that the tests will become widely available before they have been properly vetted for benefit and harm.

In this commentary, we examine the road from demonstrating the diagnostic performance of a new early detection biomarker to producing convincing evidence of its value for population screening. We explain why a measured, sequential approach has become standard and explore whether this standard can and should be adapted to meet the needs of the moment.

Cancer early detection has always been controversial because it involves intervening in a predominantly healthy population to benefit a relative few. A cancer screening test is not a simple matter; even if the test itself is not costly or harmful, it can lead to downstream interventions that are. In the past decade, research studies and the popular press focused far more on the harms of cancer screening than on its benefits. Newspaper articles warned about overdiagnosis and overtreatment (7,8), and research studies highlighted how the risk of unnecessary biopsies is compounded dramatically under regular screening (9).

We previously commented on the valuable lessons that the history of early detection holds for the field today (10). The key lesson bears repeating: It is not just about the test's diagnostic performance. Bridging preliminary evidence that a new screening test can detect cancer to establishing that the population screening can materially reduce deaths in a sustainable fashion requires a whole sequence of pieces to fall into place.

First, the promising performance of a screening test in known cases and noncases must lead to adequate discrimination in the prospectively screened (intended-use) population. In practice, degradation of screening test sensitivity in the prospective setting is expected because the case mix will be skewed toward cases that are earlier in their natural history and may include a fraction that are clinically insignificant (11,12). Moreover, in this setting, sensitivity will depend on what happens after a positive test; if accurate confirmation testing is not readily accessible, the sensitivity of the entire screening episode will be further degraded (13).

Received: February 14, 2023. Revised: June 7, 2023. Accepted: October 25, 2023

Second, screening in the intended-use population must lead to material movement of the timing of diagnosis to an earlier, more treatable point in the disease's natural history. Because we most often conceptualize early vs late diagnosis in terms of clinical stage of disease, we generally refer to this condition as screening that produces an adequate "stage shift." This shift can happen, however, only if the cancer provides adequate opportunity—through a sufficiently long early-stage duration—to be detected at this point. Thus, the natural history of the target cancer will factor critically into the ability of a screening program to change the fate of the cancers detected. Learning a cancer's natural history, however, requires incidence data with and without screening, which is not available for most types of cancer. Whether a candidate test can achieve adequate stage shift is virtually impossible to establish for these cancers without prospective evaluation.

Third, any stage shift that screening produces should translate into an adequate and sustainable reduction in cancer mortality. Whether this will be the case depends on the cancer—the expected mortality reduction from a given stage shift is highly variable across cancers—and on the implementation of the screening program, including any subsequent diagnostic and treatment interventions. In practice, access to confirmation testing can be heterogeneous, and changes in available treatments can affect the effectiveness of screening. Thus, several factors determine whether a promising stage shift will result in a clinically significant and sustainable mortality reduction.

Establishing that a new test meets these 3 requirements while also controlling adverse outcomes, such as unnecessary biopsies and overdiagnosis, involves a corresponding sequence of studies that typically take many years to complete. In 2001, the Early Detection Research Network of the National Cancer Institute established the Phases of Biomarker Development (PBD) to codify these studies and specify criteria for progression from one phase to the next (14,15).

The PBD blueprint begins with discovery (phase 1) and assessment of the discriminative performance in known cancer cases and noncases (phase 2). It progresses to evaluation of prediagnosis performance using stored serum samples (phase 3) and ascertainment of the change in incidence and stage induced by biomarker-based screening in prospective cohort studies (phase 4). Phase 4 studies may also inform researchers about screening test performance in the prospective setting, but simple empirical estimates of sensitivity from such studies are frequently overly optimistic (16,17). The final phase is the randomized screening trial, with disease-specific mortality as a primary endpoint (phase 5). Screening trials not only avoid selection bias resulting from random assignment of screening but also examine the collective contributions of screening, diagnostic confirmation testing, and treatment to delivering mortality benefit. The demonstration of significant mortality benefit in a randomized screening trial has become established as a condition for acceptability of a new test; the United States Preventive Services Task Force (USPSTF) generally requires such evidence as a prerequisite to recommending population screening.

To a certain extent, the PBD sequence of studies is being followed with the new cancer tests. Several retrospective (phase 2) and prospective (phase 4) studies of test performance have been or are currently being conducted (11,18). Phase 3 studies are rare because of the specimen volume the new multicancer tests require (eg, 20 mL of blood for 1 product) (19). A recent phase 3 study presented findings regarding detection rates up to 3 years before clinical diagnosis based on specimens from a population cohort study that drew 36 mL of blood per participant (20). At this time, few phase 5 trials are in process, and test developers are making the case that trials with disease-specific mortality as a primary endpoint are too lengthy, costly, and complex (4,21).

Randomized screening trials are indeed lengthy, costly, and complex, particularly in the average-risk population. The rarity of disease-specific mortality in this population means that such trials must enroll a large number of participants; further, the time required to observe enough of these events in an initially asymptomatic cohort means that long follow-up is generally needed. Given their expense and duration, randomized trials can generally evaluate only 1 or 2 screening strategies. Further, because diagnostic and treatment practices often continue to evolve, the trial results may be less relevant to contemporary practice when they finally become available. Given the sheer number of liquid biopsy–based tests currently under development, it is unrealistic to conduct a screening trial for each one. To this end, there is growing interest among cancer researchers in expediting screening test evaluation in a manner that gains efficiency but retains the intent of the PBD to rigorously assess the mortality benefit that may be reasonably expected from a novel cancer screening biomarker (22).

One approach that has been gaining attention, with strong support from industry, is to use short-term outcomes in screening trials for the new tests. A prime candidate is the reduction in the incidence of late-stage disease, which has been proposed as a surrogate or provisional endpoint for mortality benefit (4,23). At this time, there is no clear consensus for how this approach would be operationalized; for example, would conclusions about screening benefit be made based on this endpoint, or would the stage shift be used to predict the reduction in disease-specific mortality? We strongly recommend against the first option. We have previously shown that a given reduction in late-stage incidence does not imply the same reduction in disease mortality across cancers; a seemingly dramatic reduction in late-stage incidence would be expected to produce only a modest mortality benefit for some cancers (24). In short, we do not know what might constitute a clinically significant reduction in late-stage incidence. Given the challenges of deimplementing cancer screening programs once they have been initiated, basing screening decisions on a seemingly favorable reduction that does not lead to a reasonable mortality benefit could be suboptimal for both clinical and policy purposes.

The second option—using the mortality reduction predicted by late-stage incidence rates as an endpoint—may be more reasonable. Indeed, studies have previously proposed using this endpoint in breast and colorectal cancer (CRC) screening trials; results suggested advantages over the mortality endpoint in terms of both timeliness and statistical power (25,26). This approach will still require further investigation, however, into different ways to predict mortality benefit, given a stage shift and their validity across cancers. For example, we have shown that accounting for prognostic subtype when substituting early-stage for late-stage survival among cases shifted from a late stage to an early stage by screening may affect the predicted mortality reduction (27).

Might other approaches be harnessed to expedite evaluation of novel screening tests and produce evidence to support well-founded population screening decisions?

Real-world data have been suggested as a potentially useful evidence source because the tests disseminate in the clinic and produce data on their use and consequences in practice (28). There are likely to be many challenges, however, to the development of reliable evidence from real-world data beyond the primary challenge of ensuring that key variables are collected in a consistent manner in the clinical setting. First, selection bias in real-world data is a well-understood problem that cannot always be remedied. Such bias will almost certainly be a major issue in the evaluation of benefits of novel early detection tests given their costs and other barriers to their use as well as to accurate confirmation testing. Long-term outcomes will take years to accumulate, and short-term outcomes will be subject to the same concerns as those pertaining to surrogate endpoints for screening trials. In addition, data sharing issues will need to be addressed or alternatives to data sharing adopted to permit analysis of large, representative datasets. If real-world data can be made available, they will be most useful for assessments of short-term outcomes: patterns and predictors of novel test use, access issues with respect to both screening and confirmation testing, and diagnostic performance in the intended-use setting.

Computational disease modeling is an established approach to learn from and extrapolate beyond the empirical results of screening studies. A computational model for studying cancer screening is a mathematical representation of the events in disease progression that drive screening outcomes, such as disease onset, metastasis, diagnosis, and survival. Once the rates of the relevant events have been estimated, the model permits projecting virtually (eg, by simulation) the impact of screening and treatment on key clinical outcomes, such as late-stage incidence and mortality. Models have been used to expand the range of screening strategies beyond those studied in actual trials through simulated trials that examine a range of screening ages and intervals, biopsy referral criteria, and strategies tailored to disease risk (29-36). In a sense, modeling has become an informal sixth phase of the PBD, and the USPSTF and other national guidelines panels have relied on modeling to inform their policies for lung, breast, cervical, and colorectal cancer screening (37). The most recent USPSTF guidelines for both lung cancer and CRC screening were directly informed by modeling. In the case of lung cancer, the modeling studies showed that changing the eligibility criteria from 30 to 20 pack-year smoking histories dramatically increased the fraction of the population eligible for screening and the life-years saved while only modestly increasing harms (36). In the case of CRC, the modeling studies showed that strategies that started screening at age 45 years increased life-years saved and yielded fewer CRC cases and deaths than similar strategies that started screening at age 50 or 55 years (38). The new USPSTF lung cancer screening guidelines expand eligibility to 20 or more pack-year smoking histories (39), and the new CRC screening guidelines recommend beginning screening at age 45 years (40).

Although computational modeling cannot replace every screening trial, a calibrated model of the disease process—one that replicates observed results when simulating existing studies—is a powerful tool. Indeed, a calibrated model permits rigorous and transparent projections that may increase efficiencies and even eliminate the need for some trials. Here, we consider 2 ways in which modeling could be used to expedite the evaluation of novel tests, but there are likely to be many more.

First, as we have already discussed, a model that has been calibrated to stage-specific incidence in a trial could be harnessed to predict the corresponding mortality reduction. A model that has been calibrated to incidence and mortality patterns could go further, predicting these outcomes beyond the trial duration. We previously used modeling to predict the long-term mortality benefit and ratio of overdiagnoses to lives saved under prostate cancer screening based on results obtained under limited follow-up from the European Randomized Study of Screening for Prostate Cancer (41). Although long-term empirical data do not always exist to validate such projections, their availability in 2 CRC screening studies permitted verification that their long-term results matched those projected by the models (42,43). This application of models not only bridges from the trial to the policy setting, which necessarily requires quantifying outcomes over a long-term (ideally lifetime) horizon, but also opens the door to potentially shortening trial durations through judicious blending of empirical and model-based results. Prediction of the mortality reduction given the observed late-stage incidence in screen vs control groups, validated in 1 setting (eg, annual testing) and used to anticipate effects in another setting (eg, biennial testing), would be an example of such a blended model.

Second, a model of screening could, in principle, build from long-term studies of an older test to project outcomes of screening using a test with different performance characteristics. Modeling studies [eg, Knudsen et al. (38)] have projected outcomes of newer stool-based tests for CRC by superimposing these tests, given their sensitivity, on existing models of CRC natural history that were calibrated to adenoma prevalence data and CRC incidence rates in the United States (44). This application of models could reduce the need for trials of novel tests when their performance in the prospective screening setting has been well estimated and calibrated models of screening for the tests' target cancers are available.

Naturally, modeling is subject to challenges and limitations. Models require extensive, high-quality data for adequate calibration; can be difficult to estimate, even when such data are available (45); and often make unverifiable assumptions about natural history, screening performance, or screening benefit. The science of modeling has advanced over the past several decades, however, and methods for mitigating these limitations have been developed and continue to evolve (46). Notable advances include metrics for independently developed models to examine the validity of unobservable quantities (such as sojourn times) (47,48), efficient algorithms to calibrate models with potentially many parameters to multiple data targets (49), and methods to propagate uncertainty in model inputs to uncertainty in policy preferences based on model outputs (50). As data from studies of novel screening tests become available, they will facilitate rigorous development of models for cancers with unknown natural histories and provide new opportunities to validate existing natural history models.

In conclusion, marketing pitches for the new tests argue that lack of screening for many cancers has created an urgency to deploy novel tests as quickly as possible. The real urgency, though, is driven by concerns that the tests may be released to an unsuspecting public before researchers can establish that the tests will do more good than harm. The early detection research community must now address the conflict between the established blueprint of the PBD and the pressure to accelerate evaluation of novel screening tests. Will we continue to require a significant reduction in disease mortality from a randomized study to greenlight a new test? Or are we willing to expand the scope of what constitutes adequate evidence to recommend such a test? If the latter course is followed, then building a rigorous program of objective analytical and modeling studies to learn from and extend the results of PBD studies would be a worthwhile investment.

Data availability

No new data were generated or analyzed in this manuscript.

Author contributions

Ruth Etzioni, PhD (Conceptualization; Funding acquisition; Writing—original draft; Writing—review & editing), Roman Gulati, MS (Funding acquisition; Visualization; Writing—original draft; Writing—review & editing), Christos Patriotis, PhD (Writing—review & editing), Carolyn Rutter, PhD (Writing—review & editing), Yingye Zheng, PhD (Writing—review & editing), Sudhir Srivastava, PhD (Writing—review & editing), Ziding Feng, PhD (Conceptualization; Funding acquisition; Writing—review & editing).

Funding

This work was supported by the Rosalie and Harold Rea Brown Endowed Chair (to R.E.) and the National Cancer Institute (grant No. R35CA274442 to R.E., R50CA221836 to R.G., and U24CA086368 to the Early Detection Research Network Data Management and Coordinating Center).

Conflicts of interest

The authors have no conflicts of interest to declare.

Acknowledgements

The funder had no role in the design of the study; the collection, analysis, or interpretation of the data; or the writing of the manuscript and decision to submit it for publication.

References

1. Cohen JD, Li L, Wang Y, et al. Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science*. 2018;359(6378):926-930.
2. Cristiano S, Leal A, Phallen J, et al. Genome-wide cell-free DNA fragmentation in patients with cancer. *Nature*. 2019;570(7761):385-389.
3. Liu MC, Oxnard GR, Klein EA, et al.; CCGA Consortium. Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. *Ann Oncol*. 2020;31(6):745-759.
4. Klein EA, Madhavan S, Beer TM, et al. Dying to find out: the cost of time at the dawn of the multicancer early detection era. *Cancer Epidemiol Biomarkers Prev* 2023;32(8):1003-1010., 10.1158/1055-9965.EPI-22-1275.
5. Welch HG, Dey T. Testing whether cancer screening saves lives: implications for randomized clinical trials of multicancer screening. *JAMA Intern Med*. 2023;183(11):1255-1258. doi:10.1001/jamainternmed.2023.3781.
6. Deverka PA, Douglas MP, Phillips KA. Multicancer screening tests: anticipating and addressing considerations for payer coverage and patient access. *Health Aff (Millwood)*. 2022;41(3):383-389.
7. Parker-Pope T. *Mammogram's Role as Savior Is Tested*. New York Times. October 25, 2011, 2011. https://archive.nytimes.com/well.blogs.nytimes.com/2011/10/24/mammograms-role-as-savior-is-tested/. Accessed August 15, 2023.
8. Welch HG. *Cancer Survivor or Victim of Overdiagnosis?* New York Times. November 21, 2012, 2012. https://www.nytimes.com/2012/11/22/opinion/cancer-survivor-or-victim-of-overdiagnosis.html. Accessed August 15, 2023.
9. Croswell JM, Kramer BS, Kreimer AR, et al. Cumulative incidence of false-positive results in repeated, multimodal cancer screening. *Ann Fam Med*. 2009;7(3):212-222.
10. Etzioni R, Gulati R, Weiss NS. Multi-cancer early detection: learning from the past to meet the future. *J Natl Cancer Inst* 2022;114(3):349-352. doi:10.1093/jnci/djab168.
11. LeeVan E, Pinsky P. Predictive performance of cell-free nucleic acid-based multi-cancer early detection tests: a systematic review. *Clin Chem*. 2023. doi:10.1093/clinchem/hvad134.
12. Schrag D, Beer TM, McDonnell CH 3rd, et al. Blood-based tests for multicancer early detection (PATHFINDER): a prospective cohort study. *Lancet*. 2023;402(10409):1251-1260.
13. Hakama M, Auvinen A, Day NE, et al. Sensitivity in cancer screening. *J Med Screen*. 2007;14(4):174-177.
14. Pepe MS, Etzioni R, Feng Z, et al. Phases of biomarker development for early detection of cancer. *J Natl Cancer Inst*. 2001;93(14):1054-1061.
15. Feng Z, Pepe MS. Adding rigor to biomarker evaluations-EDRN experience. *Cancer Epidemiol Biomarkers Prev*. 2020;29(12):2575-2582.
16. Chubak J, Burnett-Hartman AN, Barlow WE, et al. Estimating cancer screening sensitivity and specificity using healthcare utilization data: defining the accuracy assessment interval. *Cancer Epidemiol Biomarkers Prev*. 2022;31(8):1517-1520.
17. Lange J, Zhao Y, Gogebakan KC, et al. Test sensitivity in a prospective cancer screening program: a critique of a common proxy measure. *Stat Methods Med Res*. 2023;32(6):1053-1063.
18. Nadauld LD, McDonnell CH, 3rd, Beer TM, et al. The PATHFINDER study: assessment of the implementation of an investigational multi-cancer early detection test into clinical practice. *Cancers (Basel)*. 2021;13(14):3501.
19. https://www.galleri.com/patient/faqs. Accessed August 15, 2023.
20. Patel A, Dur CAC, Alexander G, et al. Methylated DNA biomarkers and incident cancer in the American Cancer Society (ACS) Cancer Prevention Study-3 (CPS-3) cohort. *J Clin Oncol*. 2023;41(suppl 16):3004-3004.
21. Putcha G, Gutierrez A, Skates S. Multicancer screening: one size does not fit all. *J Clin Oncol Precis Oncol*. 2021;5:574-576.
22. Minasian LM, Pinsky P, Katki HA, et al. Study design considerations for trials to evaluate multicancer early detection assays for clinical utility. *J Natl Cancer Inst*. 2023;115(3):250-257.
23. Raoof S, Lee RJ, Jajoo K, et al. Multicancer early detection technologies: a review informed by past cancer screening studies. *Cancer Epidemiol Biomarkers Prev*. 2022;31(6):1139-1145.
24. Owens L, Gulati R, Etzioni R. Stage shift as an endpoint in cancer screening trials: implications for evaluating multi-cancer early detection tests. *Cancer Epidemiol Biomarkers Prev*. 2022;31(7):1298-1304., 10.1158/1055-9965.EPI-22-0024.
25. Day NE, Duffy SW. Trial design based on surrogate end points—application to comparison of different breast screening frequencies. *J R Stat Soc Ser A (Stat Soc)*. 1996;159(1):49-60.
26. Cuzick J, Cafferty FH, Edwards R, et al. Surrogate endpoints for cancer screening trials: general principles and an illustration using the UK Flexible Sigmoidoscopy Screening Trial. *J Med Screen*. 2007;14(4):178-185.
27. Owens L, Gogebakan KC, Menon U, et al. Short-term endpoints for cancer screening trials: does tumor subtype matter? *Cancer Epidemiol Biomarkers Prev*. 2023;32(6):741-743.

28. Tunis S, Parmar MKB, Burris HA, et al. Approaches and data needed for real-world evaluation of multicancer early detection tests. *J Clin Oncol*. 2023;41(suppl 16):e15069.
29. Zauber AG, Lansdorp-Vogelaar I, Knudsen AB, et al. Evaluating test strategies for colorectal cancer screening: a decision analysis for the U.S. Preventive Services Task Force. *Ann Intern Med*. 2008;149(9):659-669.
30. Gulati R, Gore JL, Etzioni R. Comparative effectiveness of alternative prostate-specific antigen-based prostate cancer screening strategies: model estimates of potential benefits and harms. *Ann Intern Med*. 2013;158(3):145-153.
31. de Koning HJ, Meza R, Plevritis SK, et al. Benefits and harms of computed tomography lung cancer screening strategies: a comparative modeling study for the U.S. Preventive Services Task Force. *Ann Intern Med*. 2014;160(5):311-320.
32. Knudsen AB, Zauber AG, Rutter CM, et al. Estimation of benefits, burden, and harms of colorectal cancer screening strategies: modeling study for the US Preventive Services Task Force. *JAMA*. 2016;315(23):2595-2609.
33. Mandelblatt JS, Stout NK, Schechter CB, et al. Collaborative modeling of the benefits and harms associated with different U. S. breast cancer screening strategies. *Ann Intern Med*. 2016;164 (4):215-225.
34. Pashayan N, Morris S, Gilbert FJ, et al. Cost-effectiveness and benefit-to-harm ratio of risk-stratified screening for breast cancer: a life-table model. *JAMA Oncol*. 2018;4(11):1504-1510.
35. Hendrix N, Gulati R, Jiao B, et al. Clarifying the trade-offs of risk-stratified screening for prostate cancer: a cost-effectiveness study. *Am J Epidemiol*. 2021;190(10):2064-2074., 10.1093/aje/kwab155.
36. Meza R, Jeon J, Toumazis I, et al. Evaluation of the benefits and harms of lung cancer screening with low-dose computed tomography: modeling study for the US Preventive Services Task Force. *JAMA*. 2021;325(10):988-997.
37. Habbema JD, Wilt TJ, Etzioni R, et al. Models in the development of clinical practice guidelines. *Ann Intern Med*. 2014;161 (11):812-818.
38. Knudsen AB, Rutter CM, Peterse EFP, et al. Colorectal cancer screening: an updated modeling study for the US Preventive Services Task Force. *JAMA*. 2021;325(19):1998-2011.
39. Krist AH, Davidson KW, Mangione CM, et al.; US Preventive Services Task Force. Screening for lung cancer: US Preventive Services Task Force recommendation statement. *JAMA*. 2021;325(10):962-970.
40. Davidson KW, Barry MJ, Mangione CM, et al.; US Preventive Services Task Force. Screening for colorectal cancer: US Preventive Services Task Force recommendation statement. *JAMA*. 2021;325(19):1965-1977.
41. Shoag JE, Nyame YA, Gulati R, et al. Reconsidering the trade-offs of prostate cancer screening. *N Engl J Med*. 2020;382 (25):2465-2468.
42. DeYoreo M, Lansdorp-Vogelaar I, Knudsen AB, et al. Validation of colorectal cancer models on long-term outcomes from a randomized controlled trial. *Med Decis Making*. 2020;40 (8):1034-1040.
43. van den Berg DMN, Nascimento de Lima P, Knudsen AB, et al.; Cisnet-Colon Group. NordICC trial results in line with expected colorectal cancer mortality reduction after colonoscopy: a modeling study. *Gastroenterology*. 2023;165(4):1077-1079.e2.
44. Kuntz KM, Lansdorp-Vogelaar I, Rutter CM, et al. A systematic comparison of microsimulation models of colorectal cancer: the role of assumptions about adenoma progression. *Med Decis Making*. 2011;31(4):530-539.
45. Ryser MD, Gulati R, Eisenberg MC, et al. Identification of the fraction of indolent tumors and associated overdiagnosis in breast cancer screening trials. *Am J Epidemiol*. 2019;188 (1):197-205.
46. Alarid-Escudero F, Gulati R, Rutter C. Validation of microsimulation models used for population health policy. In: Apostolopoulos Y, Lich K, Lemke M, eds. *Complex Systems and Population Health*. New York, NY: Oxford University Press; 2020.
47. van Ballegooijen M, Rutter CM, Knudsen AB, et al. Clarifying differences in natural history between models of screening: the case of colorectal cancer. *Med Decis Making*. 2011;31(4):540-549.
48. de Kok I, Burger EA, Naber SK, et al. The impact of different screening model structures on cervical cancer incidence and mortality predictions: the Maximum Clinical Incidence Reduction (MCLIR) methodology. *Med Decis Making*. 2020;40 (4):474-482.
49. Rutter CM, Ozik J, DeYoreo M, et al. Microsimulation model calibration using incremental mixture approximate bayesian computation. *Ann Appl Stat*. 2019;13(4):2189-2212.
50. Lempert RJ. Robust Decision Making (RDM). In: Marchau VAWJ, Walker WE, Bloemen PJTM, et al., eds. *Decision Making under Deep Uncertainty: From Theory to Practice*. Cham: Springer International Publishing; 2019:23-51.