PharmacoEconomics (2025) 43:483–497
https://doi.org/10.1007/s40273-025-01470-7

REVIEW ARTICLE

Towards Recommendations for Cost-Effectiveness Analysis of Predictive, Prognostic, and Serial Biomarker Tests in Oncology

Astrid Kramer[1] · Lucas F. van Schaik[2,3] · Daan van den Broek[4] · Gerrit A. Meijer[5] · Iñaki Gutierrez Ibarluzea[6] · Lorea Galnares Cordero[6] · Remond J. A. Fijneman[5] · Marjolijn J. L. Ligtenberg[7,8] · Ed Schuuring[9] · Wim H. van Harten[2,10] · Veerle M. H. Coupé[1] · Valesca P. Retèl[2,3] on behalf of the COIN Consortium

Accepted: 19 January 2025 / Published online: 8 February 2025

Abstract

Background Cost-effectiveness analysis (CEA) of biomarkers is challenging due to the indirect impact on health outcomes and the lack of sufficient fit-for-purpose data. Hands-on guidance is lacking.

Objective We aimed firstly to explore how CEAs in the context of three different types of biomarker applications have addressed these challenges, and secondly to develop recommendations for future CEAs.

Methods A scoping review was performed for three biomarker applications: predictive, prognostic, and serial testing, in advanced non-small cell lung cancer, early-stage colorectal cancer, and all-stage colorectal cancer, respectively. Information was extracted on the model assumptions and uncertainty, and the reported outcomes. An in-depth analysis of the literature was performed describing the impact of model assumptions in the included studies.

Results A total of 43 CEAs were included (31 predictive, 6 prognostic, and 6 serial testing). Of these, 40 utilized different sources for test and treatment parameters, and three studies utilized a single source. Test performance was included in 78% of these studies utilizing different sources, but this parameter was differently expressed across biomarker applications. Sensitivity analyses for test performance was only performed in half of these studies. For the linkage of test results to treatments outcomes, a minority of the studies explored the impact of suboptimal adherence to test results, and/or explored potential differences in treatment effects for different biomarker subgroups. Intermediate outcomes were reported by 67% of studies.

Conclusions We identified various approaches for dealing with challenges in CEAs of biomarker tests for three different biomarker applications. Recommendations on assumptions, handling uncertainty, and reported outcomes were drafted to enhance modeling practices for future biomarker cost-effectiveness evaluations.

1 Introduction

In oncology, biomarker tests are important to guide personalized treatment decisions by providing genetic and molecular information. These tests can provide a variety of information, enabling their application in various clinical decision problems across different cancer stages and tumor types. Applications of biomarkers during cancer treatment include predictive, prognostic, and serial testing. Predictive testing refers to the identification of markers for the selection of treatment (e.g. EGFR mutation for tyrosine kinase inhibitor treatment for lung cancer patients) [1]. Prognostic tests identify markers that allow for stratifying patients into subgroups at high- or low-risk to develop an event (e.g. circulating tumor DNA for the risk of recurrence after curative surgery) [2]. Serial testing refers to testing at multiple time points to monitor tumor evolution (e.g. carcinoembryonic antigen in colorectal cancer for surveillance in presumably cured patients to detect recurrence of disease and to start treatment, or to monitor metastatic patients to detect disease progression and to switch treatment) [3].

The accuracy of biomarker tests is evaluated in diagnostic accuracy studies. Decision makers, such as reimbursement agencies, often require evidence beyond test accuracy—the impact of biomarker tests on health outcomes (i.e. clinical utility) [4]. This is determined both by the ability of the test

Astrid Kramer and Lucas F. van Schaik: Shared first author.

Veerle M. H. Coupé and Valesca P. Retèl: Shared last author.

Extended author information available on the last page of the article

Key Points for Decision Makers

Estimating the cost effectiveness of novel biomarker tests is complicated due to challenges in the available evidence, linking of different evidence sources, and the interpretation of model outcomes.

In cost-effectiveness analysis (CEA), evidence for informing test and treatment parameters is mainly obtained from separate sources, requiring assumptions to link test results to treatment effects. These assumptions are crucial as they impact CEA results.

Reporting of intermediate outcomes describing the impact of the test, irrespective of the health outcomes of subsequent treatment, can enhance understanding of the mechanisms that play a role in the cost effectiveness of biomarker tests.

to inform treatment decisions (i.e. test accuracy) and by the effectiveness of subsequent treatments. Double randomized controlled trials (RCTs) are considered the gold standard to generate robust evidence for (reimbursement) decision making. In such trials, patients are randomized twice: first to a testing strategy, and second to a subsequent treatment based on the test result [5]. When designed well, they can prospectively evaluate both test accuracy and the impact on health outcomes. However, successfully performing such trials is challenging, due to the complex designs of these trials, difficulties in patient recruitment, and limited research funding for diagnostic tests [4, 6, 7]. Consequently, most existing evidence on biomarker tests focuses on test accuracy and is often derived from observational and/or retrospective studies, which is generally not directly suitable for policy making [5, 8].

Decision-analytic models can be used for estimating the long-term impact of biomarker tests on health outcomes and for conducting cost-effectiveness analysis. Compared with RCTs, decision models are a useful, cheap and time-saving method for exploring the potential long-term impact, as they can synthesize and link evidence from different sources on test accuracy and treatment effectiveness. However, linking evidence requires assumptions and introduces an additional layer of uncertainty [9–12]. These assumptions can vary between different biomarker applications due to different downstream consequences. Additionally, there is little guidance available on how to deal with challenges related to the evidence linkage for different types of biomarkers [11].

Another challenge in conducting cost-effectiveness analyses for biomarker tests is interpreting the results. The primary outcome, the incremental cost-effectiveness ratio (ICER), is typically expressed in cost per quality-adjusted life-year (QALY) gained. However, when assessing biomarker tests, the ICER captures not only the impact of the biomarker test itself, but also the impact of subsequent treatments. This dual influence makes it more difficult to accurately interpret cost-effectiveness analysis of biomarkers [4, 10–12]. In addition, it has been argued that alongside health benefit and costs, other factors can be relevant for decision making, such as the capacity of laboratories [8, 13–16]. However, reporting additional outcomes besides the ICER is currently not standard practice for cost-effectiveness analysis research.

Several publications have discussed the challenges in conducting cost-effectiveness analysis in the field of personalized medicine and have provided guidance [9–13]. However, most of these provide more general guidance, and only a few focus specifically on biomarkers or other diagnostic tests [10, 11, 17]. These publications do not include the differences between various biomarker applications, as they mainly address challenges specific to predictive biomarker tests.

Therefore, our study aimed to observe how previous studies have dealt with the specific challenges described above by an in-depth exploration of published cost-effectiveness analyses for three biomarker applications (predictive, prognostic, and serial testing), focusing on (1) the model assumptions and uncertainty and (2) the reporting of additional outcomes. Building on the observations and lessons learned from our review, our second objective was to propose a set of recommendations that may provide guidance to future investigators conducting cost-effectiveness analysis of biomarker tests in the context of these three biomarker applications.

This study is initiated from the Dutch multidisciplinary "Circulating Tumor DNA on the Road to Implementation in the Netherlands" (COIN) consortium, which aims to facilitate the controlled, evidence-based introduction of circulating tumor DNA (ctDNA) testing into the Dutch healthcare system. Consequently, occasionally a ctDNA perspective is taken in this study. However, recognizing that the described challenges related to cost-effectiveness analysis are not unique to ctDNA, we adopted a broader perspective to evaluate the challenges associated with biomarkers in general, while distinguishing between different types of biomarker applications.

2 Methods

A scoping review was performed for three different biomarker applications (predictive, prognostic, and serial testing), which are described in the following section. This

review included an in-depth exploration of the literature findings. Building on the lessons learned during the in-depth exploration and discussions with experts, recommendations for future cost-effectiveness research were proposed.

2.1 Biomarker Applications

Three different applications of biomarker testing in cancer care were selected to explore differences in the cost-effectiveness analysis approach and in the (downstream) consequences of a biomarker test: predictive, prognostic, and serial testing (Table 1). A schematic visualization of the biomarker applications, their downstream consequences, and the clinical settings is shown in Fig. 1. The clinical settings of non-small cell lung cancer (NSCLC) and colorectal cancer (CRC) were selected based on available expertise within the COIN consortium and the findings of a previous study examining ctDNA applications with clinical potential that may be implemented in the future [18]. In addition, the availability of studies was checked in an initial literature scan for the selected clinical settings.

The biomarker application 'predictive testing' involves tumor profiling in advanced NSCLC patients for the identification of targets to select targeted treatments. Currently, many healthcare systems offering targeted treatments already perform biomarker tests in this setting [19]. However, there are substantial national and international differences between European hospitals in terms of which tests have been adopted, as many different methods and assays (e.g. single-gene, multi-gene tests, either tissue- or blood-based) are available for the identification of targets [20, 21].

The biomarker application 'prognostic testing' involves biomarker testing to establish a patient's prognosis (low- or high-risk of recurrence) after surgery with curative intent for stage II/III CRC to inform adjuvant treatment decisions. If a patient has a poor prognosis, an intensified treatment plan might be desired, whereas if a prognosis is good, treatment could potentially be de-escalated. Currently, the selection for adjuvant treatment differs among countries, but is generally based on prognostic clinicopathological factors and a biomarker test for mismatch repair status [22].

The third application (serial testing) involves sequential testing in all stages of CRC for monitoring patients over time to identify progression or recurrence of disease. In CRC, this is based on imaging techniques (e.g. CT scan or MRI) and the measurement of the biomarker carcinoembryonic antigen in blood [22, 23]. Serial testing can involve follow-up programs with a variety of tests and/or procedures complementing each other.

2.2 Scoping Review

A separate literature search was performed in PubMed between June and December 2023 for each biomarker application. The search strategies included relevant MeSH terms and title and/or abstract keywords to identify relevant articles. No publication date restrictions were applied to these searches. All searches contained search terms to identify cost-effectiveness analyses. Additionally, each search strategy contained separate search terms related to the specified population for each biomarker application, and search terms to identify the relevant biomarker application and tests (Online Resource 1, see electronic supplementary material [ESM]). The search queries were optimized by cross-checking with a recently published systematic review on health economic evidence for liquid biopsies [24].

The eligibility criteria were identical for all three biomarker applications. Studies were included only if they reported cost-effectiveness outcomes (both costs and survival measures, such as [quality-adjusted] life years), evaluated a test, and matched the clinical setting described in Table 1. For serial testing, imaging tests were also included, because they are typically included in follow-up programs and their evaluation involves similar modeling challenges to biomarker tests. Studies were excluded when they were not published in English, did not match the clinical setting, no cost-effectiveness outcomes were reported, or when the full text was not available. Titles and abstracts were screened by two independent reviewers (AK, LS) based on the eligibility criteria. Disagreements were resolved through consensus. Subsequently, full-text articles were assessed in an identical procedure to the title and abstract screening.

Table 1 Description of the selected biomarker applications in the scoping review

Application	Goal of test/biomarker	Treatment decision	Test frequency	Population
Predictive testing	Identify target	Select (targeted) treatment	Single timepoint	Advanced NSCLC
Prognostic testing	Identify patients at high-risk for recurrence	Treat with adjuvant therapy	Single timepoint	Non-metastatic CRC
Serial testing	Identify recurrence	Treat recurrence	Multiple timepoints	All stages CRC
	Identify progression	Adapt treatment		

CRC colorectal cancer, *NSCLC* non-small cell lung cancer

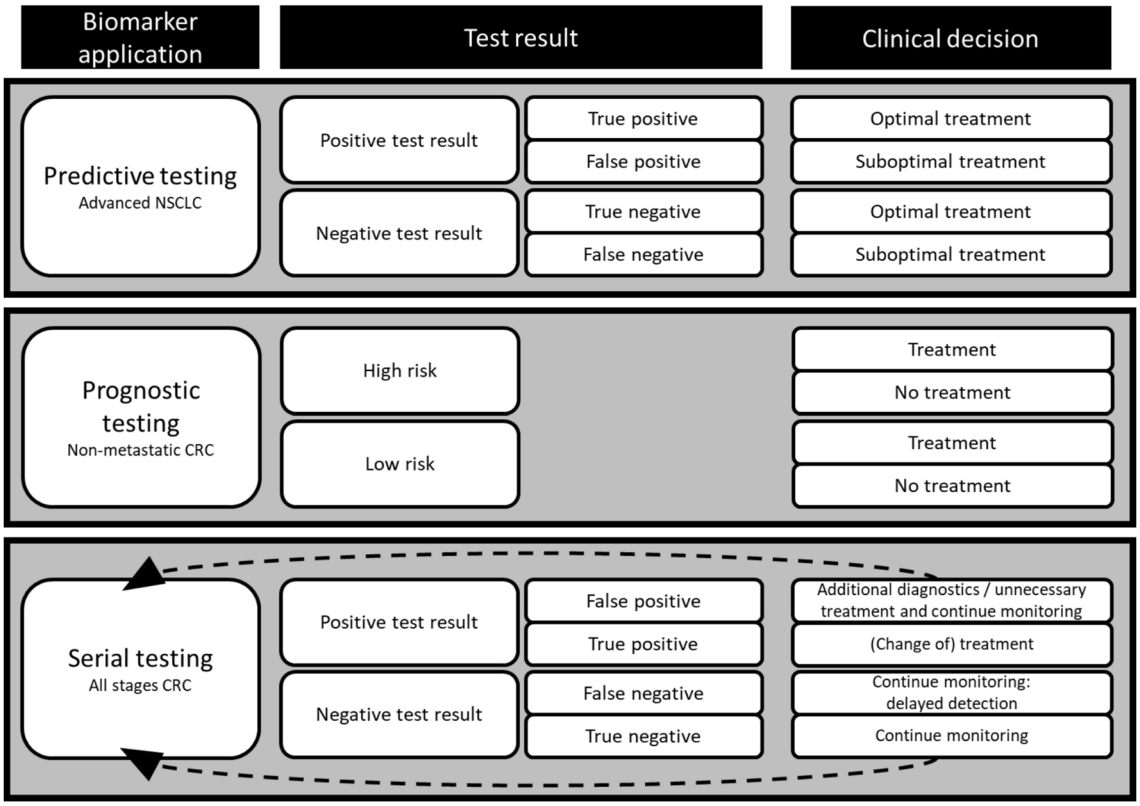

Fig. 1 A simplified schematic overview of the biomarker applications and their potential impact on subsequent clinical decisions. This figure illustrates how biomarker tests inform clinical decision making across various applications. Predictive testing involves identifying a biomarker where the test result indicates the presence (positive result) or absence (negative result) of the target biomarker. The test results can be further classified as true or false, depending on the clinical sensitivity and specificity of the test. Prognostic testing predicts the risk of disease recurrence within a patient population, often due to biological differences between subgroups. A prognostic biomarker stratifies patients into high-risk or low-risk groups. The greater the difference in recurrence risk between these subgroups, the more effective the biomarker is in accurately categorizing patients. Serial testing involves repeated testing to detect disease recurrence or progression. Like predictive testing, both positive and negative results in serial testing can be classified as true or false, based on the sensitivity and specificity. *CRC* colorectal cancer, *NSCLC* non-small cell lung cancer

The data extraction process was conducted by the same reviewers, AK and LS. Of the included studies, 30%, distributed across the biomarker applications, were independently extracted and subsequently discussed by AK and LS to ensure consistency and extraction of all relevant information. Subsequently, the data extraction process was standardized and the remaining studies were divided between AK and LS. Any ambiguities were discussed when encountered. In addition to general study characteristics, the data extraction process collected information on (i) model assumptions and uncertainty, and (ii) the reported outcomes. The model assumptions and uncertainty consisted of the evidence base for the input parameters, the link between the test and treatment consequences, and the performed sensitivity analyses. The reported outcomes consisted of all outcomes that were reported in the included cost-effectiveness studies. An overview of the extracted items can be found in Online Resource 2 (see ESM).

2.3 Development of Recommendations

An expert roundtable was organized to discuss the findings of the scoping literature review and the preliminary recommendations formulated by the core research team (AK, LS, WvH, VR, VC). Experts who participated in the roundtable included three members of the Dutch national advisory committee for diagnostic tests in oncology (i.e. cieBOD), two clinicians with (research) experience with biomarker tests, and two members of the advisory committee on expensive drug reimbursement from Dutch healthcare insurers (i.e. CieBAG). After the meeting, a summary of the discussion including the revised recommendations was sent to the experts for additional feedback.

3 Results

3.1 Included Publications

After removal of duplicates, a total of 86, 157, and 110 records were identified for predictive, prognostic, and serial testing, respectively. Of these, 55, 151, and 104 studies were excluded after abstract screening and the assessment of the full-text as they did not fulfill the eligibility criteria. Reasons for exclusion were non-English papers, other intervention or population, no cost-effectiveness outcomes reported, and full-text not available. Ultimately, 43 papers were included in total: 31 papers for predictive testing [25–54], 6 for prognostic testing [55–60], and 6 for serial testing [61–66]. The flow diagram of the study selection process of the screening procedure is shown in Online Resource 3 (see ESM).

Table 2 presents the summary characteristics of the included studies for the three biomarker applications. Different testing strategies were compared between the biomarker applications, but also a wide variety of testing strategies were observed within each biomarker application. Note that the included studies for serial testing were generally older compared with the other biomarker applications, as five of the six studies were published before 2005. In general, these older studies tended to report less comprehensive information on the methodology and the data sources. A more detailed overview of the included studies, including the extracted data per study, is listed in Online Resource 4 (see ESM).

3.2 Literature Findings

3.2.1 Model Assumptions and Uncertainty

Of the 43 included studies, 38 (88%) relied on distinct sources for the input data used for the test and treatment parameters (28 predictive testing, 6 prognostic testing, and 4 serial testing). These studies required linkage of test results to treatment evidence, which will be further described below. A further three (7%) studies, all predictive testing, informed their cost-effectiveness analysis with sources describing the combined effect of testing and subsequent treatments ('end-to-end' evidence). These three studies were not required to link test results to treatment parameters in the model. One study incorporated data from four RCTs on immunotherapy with biomarker-stratified trial designs [26]. The two other studies utilized real-world data (RWD) to inform their analyses (i.e. one national registry and one prospective observational cohort) [39, 51]. In the remaining two (5%) studies it was unclear what evidence was used as input for each parameter (both serial testing) [61, 62].

3.2.1.1 Input for Biomarker Test Studies linking sources for test parameters to different sources for subsequent treatment effects utilized a variety of input parameters related to the biomarker test to inform their models. Most studies (78%) included the test performance, although how the test performance was expressed differed across the three biomarker applications which will be further discussed below. Besides test performance, 15/43 (35%) studies also included other parameters related to the test, such as success rates of tests or biopsies, turnaround time or lead time for disease progression.

For predictive testing, test performance expressed as sensitivity and/or specificity was explicitly included in 21/28 (75%) studies. Of these 21 studies, ten studies derived the evidence for these parameters from retrospective evidence, four from prospective evidence and the remaining seven studies had a mixture of evidence, relied exclusively on expert opinion, or did not clearly report the source. Studies that did not include test performance-related parameters informed their cost-effectiveness analysis with the prevalence of mutations or the positivity rate of tests.

For prognostic testing, test performance was expressed as the difference in recurrence risk between prognostic subgroups, and was included in all six (100%) studies. Four of the six (66%) studies used a hazard ratio for recurrence risk between subgroups, one (17%) used a continuous scale

Table 2 Summary characteristics of included studies for the three biomarker applications

Predictive testing	Total studies	31 [25–54]
	Range publication year (median)	2012–2023 (2020)
	Testing strategies	Single-gene tests, multi-gene tests, liquid or tissue based
Prognostic testing	Total studies	6 [55–60]
	Range publication year (median)	2014–2022 (2021)
	Testing strategies	Immunohistochemistry, genetic tests, clinicopathological risk factors
Serial testing	Total studies	6 [61–66]
	Range publication year (median)	1990–2019 (2004)
	Testing strategies	Follow-up strategies including blood tests, imaging, colonoscopies

relative risk which was dependent on the prognostic score, and one (17%) used time to recurrence distributions for different prognostic subgroups. Five (85%) studies based this parameter on prospective evidence, and one (17%) on retrospective evidence.

For serial testing, test performance was expressed as sensitivity and/or specificity, and was explicitly included in 4/6 (66%) studies. Two of these studies based this parameter on prospective evidence, one study on expert opinion, and the last study did not clearly report the evidence source. The remaining two (33%) studies did not clearly report how and if it was incorporated in their cost-effectiveness analysis.

Some testing strategies included multiple tests, which were performed either in parallel and/or in sequence. For predictive testing, 23/28 (82%) studies included multiple tests. Of these 23 studies, 13 (57%) studies included multiple tests performed in parallel and 19 (83%) performed tests in sequence; 12 (52%) of these studies explicitly reported on the relationship between these multiple tests. They mostly assumed that mutations were mutually exclusive. In addition, two studies incorporated a correlation between PD-L1 and other biomarker status in their models. In prognostic testing, only 2/6 (33%) studies included multiple tests. These tests were performed in parallel, and no relationship between test results was discussed. In serial testing, all six (100%) studies included multiple tests. All of these studies included sequential testing, and four (66%) also included strategies with multiple tests performed in parallel. The relationship between these tests was not reported in five (80%) of these studies. The study that reported on the relationship between tests pooled the sensitivity and specificity of all tests at one time point to estimate the combined testing performance.

3.2.1.2 Assumptions About the Adherence to the Test Result Studies that used different evidence sources for test and treatment parameters were required to make assumptions to link a test result to the treatment effectiveness. One of the underlying assumptions includes to what extent the test result is always followed in the subsequent treatment decisions. Of the 38 studies using different evidence sources, most studies assumed in the base-case analysis that clinicians perfectly adhered to the test results in making the subsequent treatment decision (26/28 [93%] for predictive testing, 3/6 [50%] for prognostic testing, 4/4 [100%] for serial testing). However, this assumption was often not explicitly mentioned.

3.2.1.3 Assumptions About the Different Treatment Effects for Different Biomarker Subgroups A second underlying assumption for linking test and treatment parameters from different sources concerns the treatment effectiveness for different subgroups with a different test result. For predictive and serial testing, this related to the difference in treatment effectiveness in patients with true- or false-positive test results (and true-/false-negative test results). For predictive testing, the impact of false-positive and false-negative test results was incorporated in 11/28 (39%) studies, while in serial testing, this was explicitly addressed in 2/4 (50%) studies. For prognostic testing, a differentiation in treatment effects between prognostic subgroups (low and high risk) would indicate that besides a prognostic effect, the biomarker also has some predictive effects. Two of six (33%) studies assumed different treatment effects in different prognostic subgroups in the base case analysis. In addition, one other study stated that they did not assume a different treatment effect in their model, as existing evidence had demonstrated that there was no difference between the subgroups [56].

3.2.1.4 Exploring the Uncertainty Almost all included studies, both studies utilizing different evidence sources and end-to-end sources, conducted sensitivity analyses, including scenario analyses, probabilistic, and one- or two-way sensitivity analyses (42/43). Among the studies that included test performance, 17/21 (81%), 5/6 (83%), and 2/4 (50%) studies explored the impact of test performance in predictive, prognostic, and serial testing, respectively. The impact of the cost of testing was less often explored, with 15/31 (48%) exploring cost for predictive testing, 4/6 (66%) for prognostic testing, and cost was not explored in any of the studies for serial testing. The impact of suboptimal adherence to the test results was explored in sensitivity analyses in 2/31 (6%), 2/6 (33%), and 0/6 (0%) studies, for predictive, prognostic, and serial testing, respectively. The uncertainty around different treatment effects for different biomarker subgroups was assessed in 5/31 (13%) for predictive testing, 3/6 (50%) for prognostic testing, and 0/6 (0%) for serial testing in sensitivity analyses.

3.2.2 Reported Model Outcomes

3.2.2.1 Long-Term Outcomes (of Test and Subsequent Treatment(s)) and Intermediate Outcomes (of Diagnostic Test Phase) All included studies reported long-term cost outcomes and clinical outcomes in terms of survival. Besides long-term outcomes, 67% of studies reported intermediate outcomes, which are outcomes that provide information on the impact of the test, without yet incorporating the effects and costs of subsequent treatments (22/31 [71%] for predictive testing, 4/6 [66%] for prognostic testing, 3/6 [50%] for serial testing). Costs related to the testing procedure only (i.e. costs of testing) were reported in 17/31 [55%], 2/6 [33%], and 0/6 [0%] studies for predictive, prognostic, and serial testing, respectively. Various other intermediate outcomes were reported by 16/31 (52%) studies for predic-

tive testing, 3/6 (50%) studies for prognostic testing, and 3/6 (50%) studies for serial testing. Especially within predictive testing, a wide range of short-term outcomes were identified (Fig. 2).

3.2.2.2 Cost-Effectiveness Ratios Most studies included cost-effectiveness ratios for long-term outcomes (24/31 for predictive testing, 6/6 for prognostic testing, 6/6 for serial testing) (e.g. cost/QALY). Few studies also reported cost-efficiency ratios for intermediate outcomes (i.e. related to the impact of the test only) (5/31 [16%] for predictive testing, 0/6 [0%] for prognostic testing, 0/6 [0%] for serial testing).

3.3 Lessons Learned and Observations From the Scoping Review

3.3.1 Lessons Learned Regarding Model Assumptions and Uncertainty

Observation 1: Most studies utilized different evidence sources for the input of test and treatment parameters.

The most robust evidence for the clinical utility of a biomarker and the subsequent treatment (decision) can be obtained through RCTs. Double randomized RCTs, in which patients undergo two levels of randomization: (1) an initial randomization to the test, and (2) a subsequent randomization within each arm to the subsequent treatment based on the biomarker result, is the ideal trial design that allows for the evaluation of both the test and the treatment [5]. However, such trials are challenging to perform due to practical and sometimes ethical concerns. This is reflected in our scoping review, where none of the studies used evidence from a double-randomized RCT, and most used different evidence sources for the test and treatment parameters. Of the three studies that used a single source for these parameters, two relied on RWD.

Studies using an end-to-end source for test and treatment often combined these into a single test-treatment parameter, as can be seen in Steuten et al. [51] and Loubière et al. [39]. As a consequence, these studies did not require assumptions to link test outcomes to treatment effects. In addition, if a study uses a single evidence source for both test and treatment parameters, the data is derived from the same population, avoiding bias that can occur when linking multiple sources of potentially different populations. An advantage of using real-world test and treatment data is that it better reflects the real clinical pathway, and implicitly includes other relevant testing aspects, such as the timing of testing or test adherence. However, studies using a single source, particularly when this concerns RWD, also have several drawbacks: RWD tends to be more susceptible to bias, comparator data can be more difficult to obtain, and there is less flexibility to evaluate multiple testing strategies or conduct extensive sensitivity analyses.

Using different evidence sources for test and treatment parameters enables more stepwise modeling of all clinical actions and greater flexibility in the analysis, allowing for the evaluation of a broader range of strategies and more sensitivity and scenario analyses. While robust cost-effectiveness analyses can be conducted using multiple data sources, researchers should remain aware of potential pitfalls and implications of linking evidence. The following observations, lessons learned, and recommendations are particularly relevant for studies utilizing different data sources.

During the round table discussion, experts indicated that the use of a different patient population may result in a different test performance and/or treatment efficacy, thereby introducing bias in the cost-effectiveness analysis. Therefore, the first recommendation is that 'the intended population

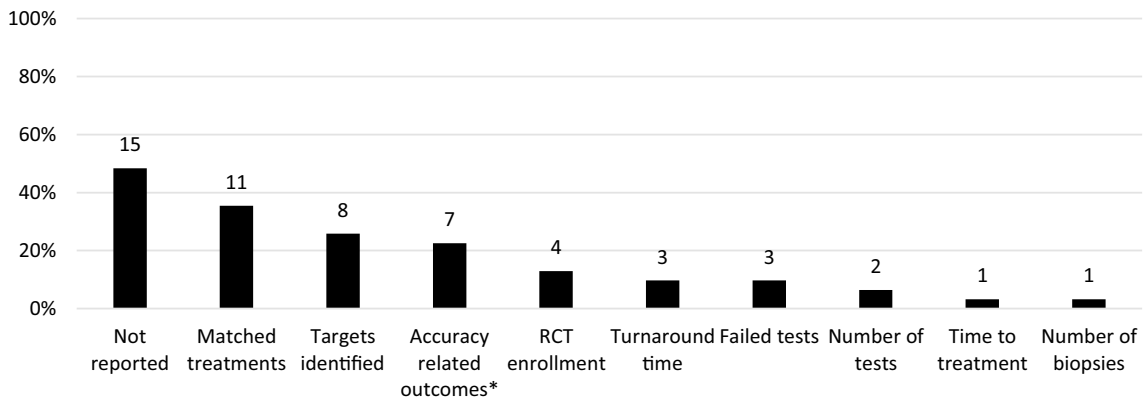

Fig. 2 Reported intermediate outcomes for predictive testing cost-effectiveness analyses. *Accuracy-related outcomes include true positives and negatives, false positives and negatives, suboptimal received treatments (treatments based on false negatives and false positives) and correct treatment decisions. *RCT* randomized controlled trial

and biomarker test application in the economic evaluation should align with the evidence sources' (Recommendation 1, Table 3). To clarify, if the cost effectiveness of a biomarker test X that identifies biomarker Y in a patient population Z is evaluated, test parameters should be informed by evidence in which biomarker test X is used to identify biomarker Y in patient population Z.

> Observation 2: Test performance is included in most studies but expressed in different parameters across biomarker applications, and the relationship between multiple tests is not always considered.

When evaluating the cost effectiveness of different testing strategies, the key differences between the tests lie in how well they identify a target (predictive testing), high-risk patients (prognostic testing), or disease recurrence/progression (serial testing). In our scoping review, we found that most studies included a parameter for test performance. In predictive and serial testing, this was primarily incorporated as sensitivity and/or specificity, while in prognostic testing, this was incorporated as the difference in recurrence risk between prognostic subgroups. If test performance was not incorporated, we observed that evidence linkage was simplified by assuming a 100% test accuracy. For example, Simons et al. compared different testing strategies using only the prevalence of alterations [49, 50]. Omitting test performance in their model may have contributed to similar identified alteration rates across the compared testing strategies.

Explicitly incorporating test performance allows for a more accurate comparison of testing strategies and their characteristics, as demonstrated in the study by Hofmarcher et al. [36]. They accounted for differences in sensitivity and specificity between biomarker tests, with a notable difference in specificity between PCR (86%) and next-generation sequencing (NGS) (100%). This contributed to improved treatment allocation in the NGS-based testing strategy. To avoid oversimplification when evaluating testing strategies, we propose to '*explicitly consider test performance in cost-effectiveness analysis*' (Recommendation 2, Table 3). This enhances the comparison of test strategies, enables modeling of downstream consequences of inaccurate or suboptimal test results, and allows for the reporting of intermediate outcomes related to the test.

The majority of studies in predictive and serial testing evaluated strategies that involved a combination of tests. When multiple tests are conducted, their results may be interdependent. For predictive testing, most studies dealt with this by assuming that mutations were mutually exclusive. Only one study included a source containing evidence on the likelihood of co-occurrence of multiple targets [41]. In serial testing, little consideration was given to the correlation between outcomes of tests performed in parallel or in sequence, while this is particularly relevant in this context, because follow-up programs often include a variety of tests. Therefore, we recommend to *'consider the interdependency between different tests at the same or at sequential time points, and explicitly report the underlying assumptions'* (Recommendation 3, Table 3).

> Observation 3: Most studies that included the test performance analyzed its impact through sensitivity analyses, whereas only approximately half of the studies varied the cost of testing.

In the studies that performed sensitivity analysis for either or both test performance and test costs, the influence of these parameters seemed to vary between clinical applications and patient populations. In predictive testing, test performance and costs were often not among the most influential factors, as test costs were typically overshadowed by expensive (targeted) treatments. In prognostic testing, varying test performance had limited impact in the studies. Two studies demonstrated that varying the costs of testing impacted their conclusions, changing the preferred strategy [57, 58]. In serial testing, the study from Wanis et al. showed that varying the test performance affected the preferred testing

Table 3 Proposed recommendations for cost-effectiveness analysis for biomarker tests

Model assumptions and uncertainty	1. The intended population and biomarker test application in the economic evaluation should align with the evidence sources
	2. Explicitly consider test performance in cost-effectiveness analysis
	3. Consider the interdependency between different tests at the same or at sequential time points, and explicitly report the underlying assumptions
	4. Explore the impact of specifically the test costs and the test performance in sensitivity analyses
	5. Explore the impact of suboptimal adherence to the test results through sensitivity analyses
	6. When using different sources of evidence for test and treatment parameters, consider potential differences in treatment effects between biomarker subgroups
Reported outcomes	7. Besides the standard long-term outcomes (of test and subsequent treatment(s)), also report intermediate outcomes (of the diagnostic phase) to provide more insight into downstream consequences
	8. Report incremental cost-efficiency ratios for relevant intermediate outcomes

strategy [66]. The impact of test costs was not explored in any of the studies on serial testing, despite the fact that this impact is multiplied over time in this biomarker application due to the repetitive nature of testing.

The evolving landscape of biomarker applications can result in advances in technologies improving the test performance and decreasing costs over time. Performing sensitivity analyses for these parameters is therefore highly informative, guiding future research and further test development. Several examples illustrated how these analyses contributed to the robustness of cost-effectiveness analyses. Therefore, we propose to '*explore the impact of specifically the test costs and the test performance in sensitivity analyses'* (Recommendation 4, Table 3).

Observation 4: Most studies assumed a perfect adherence between test results and subsequent clinical decisions.

In our scoping review, we observed that the impact of suboptimal adherence to the test result was only considered in a minority of studies in either the base-case or sensitivity analysis, while this can also significantly impact the results of cost-effectiveness studies. To illustrate, we observed in the study of Jongeneel et al. that the preferred strategy changed in the sensitivity analysis in which they assumed real-world adherence compared with perfect adherence [59]. Data for the sensitivity analysis reflecting a scenario for real-word adherence cannot be obtained from RCTs, thus was generally obtained from RWD sources [59, 60], or expert opinion [48]. Considering that clinical practice does not perfectly adhere to guidelines and/or test results, we propose to '*explore the impact of (suboptimal) adherence to the test results through sensitivity analyses'* (Recommendation 5, Table 3).

Observation 5: A minority of the included studies explored different treatment effects in different biomarker subgroups.

In the scoping review, we observed that studies using different sources for test and treatment parameters did not always explicitly consider that different biomarker subgroups can respond differently to the same treatment. For predictive and serial testing, this implies that patients with true- or false-positive (or negative) test results (may) exhibit a different response to the same treatment. Within predictive testing, evidence to inform the false-positive biomarker subgroup was often lacking, which multiple studies solved by assuming a treatment effect equal to best supportive care in these patients. One study informed the effectiveness of treatment in false-positive patients based on an RCT that evaluated targeted treatment in both wild-type and mutation-positive patients [29]. In this study, false positives, in patients who were assumed to have a treatment effect observed in wild-type patients, had a substantial impact on the overall survival (OS) and led to high additional costs due to misclassified patients. The application of adjusted treatment responses in serial testing can be illustrated by the work of Wanis et al., where false negatives led to missed diagnoses and delayed detections [66]. In addition, false positives led to extra costs for diagnostic workup. Conversely, Gazelle et al. included test sensitivity, but not specificity in their analysis, which limited their ability to account for the effects of false positives [63].

Prognostic biomarker tests stratify patients into subgroups by differentiating between high and low risk for recurrence. Two studies acknowledged that high-quality evidence informing the effectiveness of treatments in differentiated prognostic subgroups was not (yet) available for their prognostic biomarker tests [55, 60]. They both emphasize the role of prospective trials to examine whether a prognostic biomarker also has predictive value, indicating a different treatment effect in low- and high-risk subgroups. When such trials have not yet been performed, it can be worthwhile to explore the impact of a potential predictive value of prognostic biomarkers. These two studies explored the scenario in which high-risk patients responded better to treatment compared with low-risk patients [55, 60], which is beneficial for the prognostic biomarker of interest. On the other hand, Jongeneel et al. explored the impact of the biomarker-identified high-risk group being resistant to treatment [58]. This sensitivity analysis showed that an alternative testing strategy would be preferred in this situation. Therefore, we propose 'to consider potential differences in treatment effects for different biomarker subgroups' (Recommendation 6, Table 3). Note that the inclusion of test performance is a requirement for studies to incorporate these different treatment effects, as otherwise the biomarker subgroups cannot be differentiated.

3.3.2 Lessons Learned Regarding the Reported Outcomes

Observation 6: 67% of included studies reported intermediate outcomes.

Model-based cost-effectiveness analyses can provide long-term outcomes such as the total costs, life-years or QALYs, which are often seen some of the most important outcomes for decision makers. However, included studies across biomarker applications solely reporting these long-term outcomes provided limited insight into the underlying mechanisms driving these outcomes. Wolff et al. reported both long-term and intermediate outcomes, demonstrating that the intermediate outcomes revealed complementary insights [53]. While the long-term outcomes indicated a modest health benefit at higher costs, the intermediate outcomes showed a substantial increase in the number of patients receiving a diagnostically correct treatment, along

with reductions in turnaround time, test costs, and the number of unsuccessful tests. Thus, while demonstrating that the increase in costs was driven by treatment costs only, they highlighted the importance of reporting intermediate outcomes to better understand the mechanisms that play a role.

In the scoping review, a variety of intermediate clinical outcomes were identified. We classified them into three distinct types of outcomes (performance, efficiency, opportunity) and complemented the identified outcomes with suggestions from the experts (Table 4). Intermediate outcomes related to test performance and costs of testing can and should always be reported, but the specific outcomes depend on the biomarker application and the input parameters. For example, the quantification of the number of false test results showing the impact of the (in)accuracy of a test is only possible when sensitivity and specificity are included. Furthermore, depending on the aim of the cost-effectiveness analysis and the clinical setting, other intermediate outcomes can be relevant to report, such as the efficiency of laboratory procedures or new opportunities (e.g. clinical trial enrollment). To illustrate, four studies on predictive testing reported intermediate outcomes related to time (e.g. turnaround time of the test or time to treatment) [25, 27, 30, 53]. This can be particularly informative and relevant for institutional decision makers when they have to deal with time or capacity constraints. Note that modelers should not only report positive intermediate outcomes, but should also report negative ones as well. To provide a better understanding into the mechanisms that play a role in the cost effectiveness of biomarker tests, we propose to 'report intermediate outcomes (of the diagnostic phase), alongside the standard long-term outcomes' (Recommendation 7, Table 3) to provide more insight into downstream consequences.

Observation 7: Few studies reported intermediate cost-efficiency ratios.

Intermediate cost-efficiency ratios were only reported in predictive testing studies. Similar to reporting intermediate outcomes, reporting the incremental cost-efficiency ratio for intermediate outcomes and diagnostic costs provides useful insights about the diagnostic process for (institutional) decision makers. This can be illustrated by the results of Schluckebier et al., who present incremental 'diagnostic cost per correct case identified' (diagnostic cost of roughly $1000 per correct case identified), besides the long-term ICER including both diagnostic and treatment costs (roughly $200,000 total costs per QALY gained) [47]. By disentangling the effects of tests and treatments, the authors demonstrate that identifying correct diagnosed cases does not necessarily lead to cost-effective treatments. No cost-efficiency ratios were reported for prognostic and serial biomarker applications, while these ratios could also provide additional insights (e.g. the incremental cost per detected recurrence for serial testing). Therefore, we propose our final recommendation: 'Report incremental cost-efficiency ratios for relevant intermediate outcomes' (Recommendation 8, Table 3). It is worth noting that no universally accepted thresholds can exist for such outcomes, and may not be desirable as they would differ between indications and clinical settings. Therefore, the interpretation of such cost-efficiency ratios should be made with care.

4 Discussion

In this study, we explored if and how published cost-effectiveness analyses of biomarker tests addressed challenges related to evidence linkage and to the evaluation of the impact of biomarker tests by conducting an in-depth exploration of (i) model assumptions and uncertainty, and (ii) the reported outcomes across three biomarker applications:

Table 4 Different types of clinical outcomes classified as intermediate- and long-term model outcomes, providing relevant example outcomes for the three biomarker applications

	Predictive testing	Prognostic testing	Serial testing
Long-term clinical outcomes			
	OS PFS Quality of life	OS PFS Quality of life	OS PFS Quality of life
Intermediate clinical outcomes			
Performance	Biomarker matched treatments Targets identified Test accuracy-related outcomes *(e.g. false positive rate)*	Treatment rate (in risk groups)	Recurrence/progression rate Treatment rate Test accuracy-related outcomes *(e.g. false positive rate)* Lead time
Efficiency	Turnaround time Failure rate	Turnaround time Failure rate	Turnaround time Failure rate
Opportunity	Trial enrollment		

OS overall survival, *PFS* progression-free survival

predictive, prognostic, and serial testing. Most studies derived evidence for the test and treatment effectiveness from different sources, requiring assumptions to link test results to treatment effects, which were rarely explicitly considered in the analyses or explored in sensitivity analyses. Of these studies, we found that test performance was not always explicitly included in the analyses, although this is a requirement for such studies to incorporate downstream consequences for each biomarker subgroup. When included, the test performance was expressed differently across biomarker applications. Regarding the reported cost-effectiveness outcomes that are relevant for biomarker tests, our results showed that half of the studies reported intermediate outcomes (of the diagnostic test phase) alongside long-term outcomes. These studies showed that reporting a combination of intermediate- and long-term outcomes can enhance the understanding of the (broad) impact and the downstream consequences of biomarker tests. Based on the literature findings and in-depth exploration, we provide eight recommendations tailored to three biomarker applications, focusing on (i) the model assumptions and uncertainty, and (ii) the reported outcomes. These recommendations can help to improve modeling practices for future cost-effectiveness analyses of biomarker tests.

Our scoping review identified different numbers of relevant studies for the biomarker applications of interest. Only six papers were included for both prognostic and serial testing, compared with 31 for predictive testing. The high volume of publications on predictive testing is likely due to the increased attention that expensive targeted treatments receive from both regulatory bodies and the pharmaceutical industry, as well as their established role in clinical practice. Companion diagnostics, a form of predictive testing, are closely linked to these expensive targeted treatments. For serial testing, most papers were published before 2005. This was unexpected as, for example, several promising prospective studies for ctDNA testing have been published indicating the potential use of both prognostic and serial testing in CRC [67–70]. The lack of cost-effectiveness analyses for serial testing could be due to evidence gaps for pending clinical questions and difficulties in modeling (e.g. complex diagnostic pathways, number of treatment lines, or challenges in detectability thresholds and observing disease progression). This should, however, not be seen as a reason to refrain from performing cost-effectiveness analyses. Even at an early development stage, while considering current evidence gaps, they can provide valuable insights and guide future research. The lack of cost-effectiveness analyses in early research phases of biomarkers can delay the timely assessment of effective biomarkers and increases the risk of uncontrolled introduction into clinical practice.

In our recommendations, we underscore the significance of reporting intermediate outcomes in cost-effectiveness analyses to provide enhanced insights into the impact in the diagnostic test phase, especially when limited data is available. Intermediate outcomes can facilitate in the interpretation of long-term outcomes, as they provide additional insight into the relationships that translate test parameters via intermediate outcomes to long-term outcomes. This increases the understanding of how biomarkers may contribute to the cost effectiveness of the test-treatment pathway. While it fell outside the scope of our recommendations, achieving reimbursement for novel biomarker applications was a prominent topic during the round table discussion. From a policy perspective, long-term outcomes (e.g. ICER, OS, QALY) are often primary outcomes in national reimbursement decisions. However, during the round table discussion it was also mentioned that in specific cases, such as when a new test is a technical variant of an existing test, demonstrating similarity in test performance may suffice. Although national reimbursement bodies, in most cases, deem intermediate outcomes insufficient for decision making, the additional insights they offer can be valuable for other stakeholders, including institutional decision makers and researchers.

The valuation of these intermediate outcomes may vary among stakeholders and no universal willingness-to-pay thresholds exist for such outcomes. For instance, hospital administrators also may prioritize efficiency outcomes in their decisions, such as capacity and turnaround time, while patients may place greater value on opportunity outcomes, such as trial enrollment. This stakeholder-dependent valuation adds complexity to interpreting these outcomes and assessing the benefits of biomarker tests. Besides quantifiable outcomes that can be included in cost-effectiveness analyses, other factors that are not so easily included in cost-effectiveness analyses may contribute to the benefits of biomarker tests, such as organizational benefits of using a widely applicable biomarker test for multiple purposes, or the benefit of more certainty in clinical decision making [13–16].

Our recommendations build upon and complement existing literature concerning the challenges and recommendations for cost-effectiveness analyses in precision oncology. Vellekoop, Annemans, and Bouttel, and their colleagues have offered broad guidelines for performing cost-effectiveness analyses in the topic of personalized medicine, whereas our focus lies in providing more specific guidance for cost-effectiveness analysis of biomarker tests, tailored to specific applications [9, 12, 13]. Kip et al. developed a checklist of aspects to consider in cost-effectiveness analyses of biomarkers, irrespective of their application [10]. These aspects were also explored in our review, and in a systematic review focusing on liquid biopsies by Fagery et al., which reported comparable findings to ours [24]. To illustrate, most included studies in their review assessed test performance,

while the adherence to test results was seldom addressed. Considering that both studies identified similar challenges, we suggest that researchers adhere to the checklist of Kip et al. and use our lessons learned for more context on how published cost-effectiveness analyses have addressed or incorporated different topics on the checklist for the three different types of biomarker applications.

Additionally, Shinkins et al. discussed the challenges associated with linking test outcomes to treatment effects specifically for predictive biomarkers, highlighting the consequences of different assumptions [11]. Our research reveals that this linking process is often overlooked across all three included applications, and its execution varies significantly over the biomarker applications, especially regarding the estimates or assumptions for the treatment effectiveness in biomarker subgroups. For predictive testing, we suggest explicitly considering differentiating treatment effects for false positives and negatives, while for prognostic testing the treatment differentiation should be considered for the stratified risk groups. For serial testing, consequences of false positives and negatives should be considered beyond the treatment effect, as false-positive results can lead to additional diagnostics and false-negative results to delayed detection of recurrence and delayed initiation of treatments.

One of the strengths of our study is our comprehensive approach, examining several clinical biomarker applications across different tumor types, not just predictive testing, which is the main focus of existing guidance. Additionally, we incorporated expert opinions from various decision-making perspectives, including clinical, molecular, and health insurance authorities. Our study also has limitations. Our scoping search might have been too narrow, as we focused on studies that aimed to evaluate a test, potentially missing studies that aimed to evaluate a treatment while still including a companion diagnostic in their analysis. A second limitation is that, especially for serial testing, some of the included studies were older, of limited quality, and provided limited detail on their methodology. Despite these limitations, our study provides valuable insights and actionable recommendations for improving cost-effectiveness analyses for biomarker tests, aiming for a more comprehensive evaluation of the impact of biomarker tests in clinical practice.

5 Conclusion

Our study has identified various approaches for dealing with challenges in cost-effectiveness analyses of biomarker tests. We propose eight recommendations to improve future modeling practices, addressing modeling choices and downstream consequences of biomarker tests. Implementing these recommendations will enhance comprehensive and accurate evaluations specific to the biomarker application, ultimately improving the evaluation process which will facilitate the implementation of (cost-effective) biomarker tests.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40273-025-01470-7.

Acknowledgements We would like to thank Stijn Kleerebezem for his contribution in the early phases of this project and the experts for their participation in the roundtable discussion and their input.

Declarations

Funding The COIN project is funded by the ZonMw ' Personalised Medicine' programme (project number 848101011), PGDX and CZ and this research was funded by the CAN.HEAL project through the European Commission EU4Health Program 2021–2027 under Grant No. 101080009.

Conflict of interest A.K., L.F.S., D.B., I.G.I., L.G.C., W.H.H., V.M.H.C., and V.P.R. declare no conflict of interest. R.J.A.F reports public private partnership consortia grants in collaboration with Labcorp (Personal Genome Diagnostics), Delfi Diagnostics, Solvias (Cergentis BV), MERCK BV, outside the submitted work. In addition, R.J.A.F. has several patents pending**.** E.S. reports lectures for Bio-Rad, Seracare, Roche, Biocartis, Illumina, Lilly, Janssen Cilag (Johnson&Johnson), AstraZeneca and Agena Bioscience; he is consultant in advisory boards for MSD/Merck, GSK, AstraZeneca, Astellas Pharma, Sysmex, Roche, Novartis, Bayer, BMS, Lilly, Amgen, Illumina, Agena Bioscience, Janssen Cilag (Johnson&Johnson), Sinnovisionlab, Diaceutics, CC Diagnostics; and received research grants from Biocartis, Invitae-ArcherDX, AstraZeneca, Agena Bio-science, BMS, Bio-Rad, Roche, Boehringer Ingelheim, CC Diagnostics, SNN/EFRO and Abbott (all paid to UMCG account); and travel reimbursements from Bio-Rad, Abbott, Illumina, Agena Bioscience, Roche, IQNPath and BioRAD. M.J.L.L. has had advisory roles (institutional) with AstraZeneca, Bristol-Myers Squibb, GlaxoSmithKline, Illumina, Janssen, Lilly, Merck Sharp & Dohme and Roche. G.A.M. is co-founder and board member (CSO) of CRCbioscreen BV, CSO of Health-RI (Dutch National Health Data infrastructure for research & innovation), and member of the supervisory board of IKNL (Netherlands Comprehensive Cancer Organisation). He has a research collaboration with CZ Health Insurances (cash matching to ZonMw grant) and he has research collaborations with Exact Sciences, Sysmex, Sentinel Ch. SpA, Personal Genome Diagnostics (PGDX), DELFi and Hartwig Medical Foundation; these companies provide materials, equipment and/or sample/genomic analyses.

Ethics approval According to the Dutch Law on Medical Scientific Research (WMO), this study did not require approval by a medical ethics committee.

Informed consent All invited experts that participated in the expert roundtable consented before participating.

Data availability Extracted literature findings are available in Online Resource 4 in the ESM. Transcripts of the expert roundtable will not be shared to ensure anonymity of the participants.

Author contributions AK, VC and VR developed the ideas and the methods for this study. AK and LS conducted the screening, the data extraction and the data analysis of the literature. AK, LS, VC, VR, and WH developed the first version of the recommendations. These recommendations were revised by all authors, after which AK and LS drafted

the first version of the manuscript. The manuscript was reviewed in multiple rounds by all authors.

References

1. Lindeman NI, Cagle PT, Aisner DL, Arcila ME, Beasley MB, Bernicker EH, et al. Updated molecular testing guideline for the selection of lung cancer patients for treatment with targeted tyrosine kinase inhibitors: guideline from the College of American Pathologists, the International Association for the Study of Lung Cancer, and the Association for Molecular Pathology. J Mol Diagn. 2018;20(2):129–59.
2. Faulkner LG, Howells LM, Pepper C, Shaw JA, Thomas AL. The utility of ctDNA in detecting minimal residual disease following curative surgery in colorectal cancer: a systematic review and meta-analysis. Br J Cancer. 2023;128(2):297–309.
3. Goldstein MJ, Mitchell EP. Carcinoembryonic antigen in the staging and follow-up of patients with colorectal cancer. Cancer Invest. 2005;23(4):338–51.
4. Bossuyt PM, Reitsma JB, Linnet K, Moons KG. Beyond diagnostic accuracy: the clinical utility of diagnostic tests. Clin Chem. 2012;58(12):1636–43.
5. Merlin T, Farah C, Schubert C, Mitchell A, Hiller JE, Ryan P. Assessing personalized medicines in Australia: a national framework for reviewing codependent technologies. Med Decis Mak. 2013;33(3):333–42.
6. Pletcher MJ, Pignone M. Evaluating the clinical utility of a biomarker: a review of methods for estimating health impact. Circulation. 2011;123(10):1116–24.
7. Hot A, Bossuyt PM, Gerke O, Wahl S, Vach W, Zapf A. Randomized test-treatment studies with an outlook on adaptive designs. BMC Med Res Methodol. 2021;21(1):110.
8. Horgan D, Ciliberto G, Conte P, Baldwin D, Seijo L, Montuenga LM, et al. Bringing greater accuracy to Europe's Healthcare Systems: the unexploited potential of biomarker testing in oncology. Biomed Hub. 2020;5(3):1–42.
9. Vellekoop H, Huygens S, Versteegh M, Szilberhorn L, Zelei T, Nagy B, et al. Guidance for the harmonisation and improvement of economic evaluations of personalised medicine. Pharmacoeconomics. 2021;39(7):771–88.
10. Kip MM, IJzerman MJ, Henriksson M, Merlin T, Weinstein MC, Phelps CE, et al. Toward alignment in the reporting of economic evaluations of diagnostic tests and biomarkers: the AGREEDT checklist. Med Decis Mak. 2018;38(7):778–88.
11. Shinkins B, Allen AJ, Karichu J, Garrison LP, Monz BU. Evidence synthesis and linkage for modelling the cost-effectiveness of diagnostic tests: preliminary good practice recommendations. Appl Health Econ Health Policy. 2024;22(2):131–44.
12. Annemans L, Redekop K, Payne K. Current methodological issues in the economic assessment of personalized medicine. Value Health. 2013;16(6):S20–6.
13. Bouttell J, Heggie R, Oien K, Romaniuk A, VanSteenhouse H, von Delft S, Hawkins N. Economic evaluation of genomic/genetic tests: a review and future directions. Int J Technol Assess Health Care. 2022;38(1): e67.
14. Doble B. Budget impact and cost-effectiveness: can we afford precision medicine in oncology? Scand J Clin Lab Invest. 2016;76(sup245):S6–11.
15. Anonychuk A, Beastall G, Shorter S, Kloss-Wolf R, Neumann P, editors. A framework for assessing the value of laboratory diagnostics. Healthcare Management Forum; 2012. Los Angeles: SAGE Publications Sage CA.
16. van Schaik LF, Engelhardt EG, Wilthagen EA, Steeghs N, Coves AF, Joore MA, et al. Factors for a broad technology assessment of comprehensive genomic profiling in advanced cancer, a systematic review. Crit Rev Oncol Hematol. 2024;24(1):104441.
17. Snowsill T. Modelling the cost-effectiveness of diagnostic tests. Pharmacoeconomics. 2023;41(4):339–51.
18. Kramer A, Rubio-Alarcón C, van den Broek D, Vessies DC, Van't Erve I, Meijer GA, et al. A scenario-drafting study to explore potential future implementation pathways of circulating tumor DNA testing in oncology. Mol Oncol. 2024;18(11):2730–42.
19. Planchard D, Popat S, Kerr K, Novello S, Smit E, Faivre-Finn C, et al. Metastatic non-small cell lung cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2018;29:192–237.
20. Hofman P, Calabrese F, Kern I, Adam J, Alarcão A, Alborelli I, et al. Real-world EGFR testing practices for non-small-cell lung cancer by thoracic pathology laboratories across Europe. ESMO Open. 2023;8(5): 101628.
21. Steeghs EM, Groen HJ, Schuuring E, Aarts MJ, Damhuis RA, Voorham QJ, et al. Mutation-tailored treatment selection in non-small cell lung cancer patients in daily clinical practice. Lung Cancer. 2022;167:87–97.
22. Argilés G, Tabernero J, Labianca R, Hochhauser D, Salazar R, Iveson T, et al. Localised colon cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2020;31(10):1291–305.
23. Cervantes A, Adam R, Roselló S, Arnold D, Normanno N, Taïeb J, et al. Metastatic colorectal cancer: ESMO clinical practice guideline for diagnosis, treatment and follow-up☆. Ann Oncol. 2023;34(1):10–32.
24. Fagery M, Khorshidi HA, Wong SQ, Vu M, IJzerman M. Health economic evidence and modeling challenges for liquid biopsy assays in cancer management: a systematic literature review. Pharmacoeconomics. 2023;41(10):1229–48.
25. Cell-free circulating tumour DNA blood testing to detect EGFR T790M mutation in people with advanced non-small cell lung cancer: a health technology assessment. Ont Health Technol Assess Ser. 2020;20(5):1–176.
26. Aguiar PN Jr, Perry LA, Penny-Dimri J, Babiker H, Tadokoro H, de Mello RA, Lopes GL Jr. The effect of PD-L1 testing on the cost-effectiveness and economic impact of immune checkpoint inhibitors for the second-line treatment of NSCLC. Ann Oncol. 2017;28(9):2256–63.
27. Arriola E, Bernabé R, Campelo RG, Biscuola M, Enguita AB, López-Ríos F, et al. Cost-effectiveness of next-generation sequencing versus single-gene testing for the molecular diagnosis of patients with metastatic non-small-cell lung cancer from the perspective of Spanish reference centers. JCO Precis Oncol. 2023;7: e2200546.

28. Atherly AJ, Camidge DR. The cost-effectiveness of screening lung cancer patients for targeted drug sensitivity markers. Br J Cancer. 2012;106(6):1100–6.
29. Cheng MM, Palma JF, Scudder S, Poulios N, Liesenfeld O. The clinical and economic impact of inaccurate EGFR mutation tests in the treatment of metastatic non-small cell lung cancer. J Pers Med. 2017;7(3).
30. de Alava E, Pareja MJ, Carcedo D, Arrabal N, García JF, Bernabé-Caro R. Cost-effectiveness analysis of molecular diagnosis by next-generation sequencing versus sequential single testing in metastatic non-small cell lung cancer patients from a south Spanish hospital perspective. Expert Rev Pharmacoecon Outcomes Res. 2022;22(6):1033–42.
31. Djalalov S, Beca J, Hoch JS, Krahn M, Tsao MS, Cutz JC, Leighl NB. Cost effectiveness of EML4-ALK fusion testing and first-line crizotinib treatment for patients with advanced ALK-positive non-small-cell lung cancer. J Clin Oncol Off J Am Soc Clin Oncol. 2014;32(10):1012–9.
32. Doble B, John T, Thomas D, Fellowes A, Fox S, Lorgelly P. Cost-effectiveness of precision medicine in the fourth-line treatment of metastatic lung adenocarcinoma: an early decision analytic model of multiplex targeted sequencing. Lung Cancer. 2017;107:22–35.
33. Englmeier F, Bleckmann A, Brückl W, Griesinger F, Fleitz A, Nagels K. Clinical benefit and cost-effectiveness analysis of liquid biopsy application in patients with advanced non-small cell lung cancer (NSCLC): a modelling approach. J Cancer Res Clin Oncol. 2023;149(4):1495–511.
34. Ezeife DA, Spackman E, Juergens RA, Laskin JJ, Agulnik JS, Hao D, et al. The economic value of liquid biopsy for genomic profiling in advanced non-small cell lung cancer. Ther Adv Med Oncol. 2022;14:17588359221112696.
35. Harvey MJ, Cunningham R, Sawchyn B, Montesion M, Reddy P, McBride A, Chawla AJ. Budget impact analysis of comprehensive genomic profiling in patients with advanced non-small-cell lung cancer. JCO Precis Oncol. 2021;5:1611–24.
36. Hofmarcher T, Malmberg C, Lindgren P. A global analysis of the value of precision medicine in oncology—the case of non-small cell lung cancer. Front Med (Lausanne). 2023;10:1119506.
37. Johnston KM, Sheffield BS, Yip S, Lakzadeh P, Qian C, Nam J. Comprehensive genomic profiling for non-small-cell lung cancer: health and budget impact. Curr Oncol. 2020;27(6):e569–77.
38. Lieberthal RD, Dudash K, Axelrod R, Goldfarb NI. An economic model to value companion diagnostics in non-small-cell lung cancer. Per Med. 2013;10(2):139–47.
39. Loubière S, Drezet A, Beau-Faller M, Moro-Sibilot D, Friard S, Wislez M, et al. Cost-effectiveness of KRAS, EGFR and ALK testing for decision making in advanced nonsmall cell lung carcinoma: the French IFCT-PREDICT.amm study. Eur Respir J. 2018;51(3):1701467.
40. Lu S, Yu Y, Fu S, Ren H. Cost-effectiveness of ALK testing and first-line crizotinib therapy for non-small-cell lung cancer in China. PLoS ONE. 2018;13(10): e0205827.
41. Nadal E, Bautista D, Cabezón-Gutiérrez L, Ortega AL, Torres H, Carcedo D, et al. Clinical and economic impact of current ALK rearrangement testing in Spain compared with a hypothetical no-testing scenario. BMC Cancer. 2021;21(1):689.
42. Narita Y, Matsushima Y, Shiroiwa T, Chiba K, Nakanishi Y, Kurokawa T, Urushihara H. Cost-effectiveness analysis of EGFR mutation testing and gefitinib as first-line therapy for non-small cell lung cancer. Lung Cancer. 2015;90(1):71–7.
43. Paolini D, Tiseo M, Demma F, Furneri G, Dionisi M, Akkermans M, Marchetti A. Ventana ALK (D5F3) in the detection of patients affected by anaplastic lymphoma kinase-positive non-small-cell lung cancer: clinical and budget effect. Clin Lung Cancer. 2018;19(5):e735–43.
44. Patel YP, Husereau D, Leighl NB, Melosky B, Nam J. Health and budget impact of liquid-biopsy-based comprehensive genomic profile (CGP) testing in tissue-limited advanced non-small cell lung cancer (aNSCLC) patients. Curr Oncol. 2021;28(6):5278–94.
45. Rojo F, Conde E, Torres H, Cabezón-Gutiérrez L, Bautista D, Ramos I, et al. Clinical and economic impact of "ROS1-testing" strategy compared to a "no-ROS1-testing" strategy in advanced NSCLC in Spain. BMC Cancer. 2022;22(1):292.
46. Romanus D, Cardarella S, Cutler D, Landrum MB, Lindeman NI, Gazelle GS. Cost-effectiveness of multiplexed predictive biomarker screening in non-small-cell lung cancer. J Thorac Oncol. 2015;10(4):586–94.
47. Schluckebier L, Caetano R, Garay OU, Montenegro GT, Custodio M, Aran V, Gil FC. Cost-effectiveness analysis comparing companion diagnostic tests for EGFR, ALK, and ROS1 versus next-generation sequencing (NGS) in advanced adenocarcinoma lung cancer patients. BMC Cancer. 2020;20(1):875.
48. Signorovitch J, Zhou Z, Ryan J, Anhorn R, Chawla A. Budget impact analysis of comprehensive genomic profiling in patients with advanced non-small cell lung cancer. J Med Econ. 2019;22(2):140–50.
49. Simons M, Retèl VP, Ramaekers BLT, Butter R, Mankor JM, Paats MS, et al. Early cost effectiveness of whole-genome sequencing as a clinical diagnostic test for patients with inoperable stage IIIB, C/IV non-squamous non-small-cell lung cancer. Pharmacoeconomics. 2021;39(12):1429–42.
50. Simons M, Uyl-de Groot CA, Retèl VP, Mankor JM, Ramaekers BLT, Joore MA, van Harten WH. Cost-effectiveness and budget impact of future developments with whole-genome sequencing for patients with lung cancer. Value Health. 2023;26(1):71–80.
51. Steuten L, Goulart B, Meropol NJ, Pritchard D, Ramsey SD. Cost effectiveness of multigene panel sequencing for patients with advanced non-small-cell lung cancer. JCO Clin Cancer Inform. 2019;3:1–10.
52. Westwood M, Joore M, Whiting P, van Asselt T, Ramaekers B, Armstrong N, et al. Epidermal growth factor receptor tyrosine kinase (EGFR-TK) mutation testing in adults with locally advanced or metastatic non-small cell lung cancer: a systematic review and cost-effectiveness analysis. Health Technol Assess. 2014;18(32):1–166.
53. Wolff HB, Steeghs EMP, Mfumbilwa ZA, Groen HJM, Adang EM, Willems SM, et al. Cost-effectiveness of parallel versus sequential testing of genetic aberrations for stage IV non-small-cell lung cancer in the Netherlands. JCO Precis Oncol. 2022;6: e2200201.
54. Zou D, Ye W, Hess LM, Bhandari NR, Ale-Ali A, Foster J, et al. Diagnostic value and cost-effectiveness of next-generation sequencing-based testing for treatment of patients with advanced/metastatic non-squamous non-small-cell lung cancer in the United States. J Mol Diagn. 2022;24(8):901–14.
55. Alarid-Escudero F, Schrag D, Kuntz KM. CDX2 biomarker testing and adjuvant therapy for stage II colon cancer: an exploratory cost-effectiveness analysis. Value Health. 2022;25(3):409–18.
56. Alberts SR, Yu TM, Behrens RJ, Renfro LA, Srivastava G, Soori GS, et al. Comparative economics of a 12-gene assay for predicting risk of recurrence in stage II colon cancer. Pharmacoeconomics. 2014;32(12):1231–43.
57. Chaudhari VS, Issa AM. Cost-effectiveness of precision molecular diagnostic tests for stage II colorectal cancer. Ann Transl Med. 2022;10(23):1260.
58. Jongeneel G, Greuter MJE, Kunst N, van Erning FN, Koopman M, Medema JP, et al. Early cost-effectiveness analysis of risk-based selection strategies for adjuvant treatment in stage II colon cancer:

the potential value of prognostic molecular markers. Cancer Epidemiol Biomarkers Prev. 2021;30(9):1726–34.
59. Jongeneel G, Greuter MJE, van Erning FN, Koopman M, Vink GR, Punt CJA, Coupé VMH. Model-based effectiveness and cost-effectiveness of risk-based selection strategies for adjuvant chemotherapy in Dutch stage II colon cancer patients. Therap Adv Gastroenterol. 2021;14:1756284821995715.
60. To YH, Degeling K, Kosmider S, Wong R, Lee M, Dunn C, et al. Circulating tumour DNA as a potential cost-effective biomarker to reduce adjuvant chemotherapy overtreatment in stage II colorectal cancer. Pharmacoeconomics. 2021;39(8):953–64.
61. Borie F, Combescure C, Daurès JP, Trétarre B, Millat B. Cost-effectiveness of two follow-up strategies for curative resection of colorectal cancer: comparative study using a Markov model. World J Surg. 2004;28(6):563–9.
62. Borie F, Daurès JP, Millat B, Trétarre B. Cost and effectiveness of follow-up examinations in patients with colorectal cancer resected for cure in a French population-based study. J Gastrointest Surg. 2004;8(5):552–8.
63. Gazelle GS, Hunink MG, Kuntz KM, McMahon PM, Halpern EF, Beinfeld M, et al. Cost-effectiveness of hepatic metastasectomy in patients with metastatic colorectal carcinoma: a state-transition Monte Carlo decision analysis. Ann Surg. 2003;237(4):544–55.
64. Kievit J, van de Velde CJ. Utility and cost of carcinoembryonic antigen monitoring in colon cancer follow-up evaluation. A Markov analysis. Cancer. 1990;65(11):2580–7.
65. Norum J, Olsen JA. A cost-effectiveness approach to the Norwegian follow-up programme in colorectal cancer. Ann Oncol. 1997;8(11):1081–7.
66. Wanis KN, Maleyeff L, Van Koughnett JAM, P HDC, Ott M, Leslie K, et al. Health and economic impact of intensive surveillance for distant recurrence after curative treatment of colon cancer: a mathematical modeling study. Dis Colon Rectum. 2019;62(7):872–81.
67. Henriksen T, Demuth C, Frydendahl A, Nors J, Nesic M, Rasmussen M, et al. Unraveling the potential clinical utility of circulating tumor DNA detection in colorectal cancer—evaluation in a nationwide Danish cohort. Ann Oncol. 2024;35(2):229–39.
68. Kotani D, Oki E, Nakamura Y, Yukami H, Mishima S, Bando H, et al. Molecular residual disease and efficacy of adjuvant chemotherapy in patients with colorectal cancer. Nat Med. 2023;29(1):127–34.
69. Tie J, Cohen JD, Lahouel K, Lo SN, Wang Y, Kosmider S, et al. Circulating tumor DNA analysis guiding adjuvant therapy in stage II colon cancer. New Engl J Med. 2022;386(24):2261–72.
70. van't Erve I, Medina JE, Leal A, Papp E, Phallen J, Adleff V, et al. Metastatic colorectal cancer treatment response evaluation by ultra-deep sequencing of cell-free DNA and matched white blood cells. Clin Cancer Res. 2023;29(5):899–909.

Authors and Affiliations

Astrid Kramer[1] · Lucas F. van Schaik[2,3] · Daan van den Broek[4] · Gerrit A. Meijer[5] · Iñaki Gutierrez Ibarluzea[6] · Lorea Galnares Cordero[6] · Remond J. A. Fijneman[5] · Marjolijn J. L. Ligtenberg[7,8] · Ed Schuuring[9] · Wim H. van Harten[2,10] · Veerle M. H. Coupé[1] · Valesca P. Retèl[2,3] on behalf of the COIN Consortium

✉ Valesca P. Retèl
v.retel@nki.nl

[1] Department of Epidemiology and Data Science, Amsterdam University Medical Centers, Amsterdam, The Netherlands

[2] Division of Psychosocial Research and Epidemiology, The Netherlands Cancer Institute, Amsterdam, The Netherlands

[3] Erasmus School of Health Policy and Management, Erasmus University Rotterdam, Rotterdam, The Netherlands

[4] Department for Laboratory Medicine, The Netherlands Cancer Institute, Amsterdam, The Netherlands

[5] Department of Pathology, The Netherlands Cancer Institute, Amsterdam, The Netherlands

[6] Basque Office for Health Technology Assessment-Osteba, Bilbao, Spain

[7] Department of Human Genetics, Radboudumc, Nijmegen, The Netherlands

[8] Department of Pathology, Radboudumc, Nijmegen, The Netherlands

[9] Department of Pathology and Medical Biology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

[10] Department of Health technology and Services Research, University of Twente, Enschede, The Netherlands