Review

Maria Olsen*, Mona Ghannad, Christianne Lok and Patrick M. Bossuyt

Shortcomings in the evaluation of biomarkers in ovarian cancer: a systematic review

https://doi.org/10.1515/cclm-2019-0038
Received January 11, 2019; accepted March 9, 2019

Abstract

Background: Shortcomings in study design have been hinted at as one of the possible causes of failures in the translation of discovered biomarkers into the care of ovarian cancer patients, but systematic assessments of biomarker studies are scarce. We aimed to document study design features of recently reported evaluations of biomarkers in ovarian cancer.

Methods: We performed a systematic search in PubMed (MEDLINE) for reports of studies evaluating the clinical performance of putative biomarkers in ovarian cancer. We extracted data on study designs and characteristics.

Results: Our search resulted in 1026 studies; 329 (32%) were found eligible after screening, of which we evaluated the first 200. Of these, 93 (47%) were single center studies. Few studies reported eligibility criteria (17%), sampling methods (10%) or a sample size justification or power calculation (3%). Studies often used disjoint groups of patients, sometimes with extreme phenotypic contrasts; 46 studies included healthy controls (23%), but only five (3%) had exclusively included advanced stage cases.

Conclusions: Our findings confirm the presence of suboptimal features in clinical evaluations of ovarian cancer biomarkers. This may lead to premature claims about the clinical value of these markers or, alternatively, the risk of discarding potential biomarkers that are urgently needed.

Keywords: biomarkers; biomarker development; clinical evaluations; ovarian cancer; study designs.

Key message: This review shows that design shortcomings in the clinical evaluations of ovarian cancer biomarkers are frequent. These include limited sample size and the recruitment of multiple, disjoint groups. Such shortcomings may hinder successful translation of ovarian cancer biomarkers.

Introduction

Epithelial ovarian cancer (EOC) is the gynecologic malignancy with the highest mortality rate. With an overall 5-year survival of 95% for early stages and only 30% for advanced disease, efforts to change the survival rate in ovarian cancer has led to minor improvements over the past 25 years. Of the different histological EOC subtypes, high grade serous adenocarcinoma is the most frequent. Ovarian cancer is often asymptomatic or has nonspecific symptoms in early stage disease. As 70–80% of patients are diagnosed with advanced disease, prognosis is typically poor [1]. Using biomarkers for detection at an early curative stage is therefore a pressing unmet clinical need [2]. Biomarkers can also be used to evaluate treatment and to detect recurrence of EOC.

Considerable investments in ovarian cancer biomarker research have been made in the last decades. Despite claims from numerous studies, few markers have been successfully implemented in practice since the discovery of CA-125 [3].

The bench-to-bedside process of biomarker development is a complex and multistep process. It has several distinct phases, ranging from the discovery and analytical validation, to clinical marker evaluation and final

***Corresponding author: Maria Olsen,** Amsterdam University Medical Centers, Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Amsterdam Public Health Research Institute, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands, E-mail: m.olsen@amc.nl
Mona Ghannad: Amsterdam University Medical Centers, Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Amsterdam Public Health Research Institute, Amsterdam, The Netherlands; and Centre de Recherche Épidémiologie et Statistique Sorbonne Paris Cité, Université Paris Descartes, Centre d'épidémiologie Clinique, Hôpital Hôtel-Dieu, Paris, France
Christianne Lok: Center Gynaecologic Oncology Amsterdam, Location Antoni van Leeuwenhoek/Netherlands Cancer Institute, Department of Gynaecologic Oncology, Amsterdam, The Netherlands
Patrick M. Bossuyt: Amsterdam University Medical Centers, Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Amsterdam Public Health Research Institute, Amsterdam, The Netherlands

implementation. Each phase holds different primary objectives, methods and study designs [4–8]. Discovery studies usually show an association between marker values and clinical entities. In contrast, evaluations of clinical performance will be used to inform clinical decision making, as in recommendations for using the biomarker to guide further testing, start treatment and choice of treatment.

To properly inform decision-making, a clinical evaluation of a biomarker would include a single group of consecutive participants, recruited in a clinical setting, identified by pre-defined and clear eligibility criteria, preferably from multiple centers, to facilitate generalizability and with a sufficiently large sample size for precise estimates, justified by a power calculation [9–11].

Shortcomings in the design of clinical evaluation studies have been hinted at as one of the possible causes of failures in translation of discovered biomarkers into the care of ovarian cancer patients. This has been described mostly in commentaries, based on anecdotal evidence, but more systematic assessments of biomarker studies are scarce. The use of sub-optimal designs features may introduce bias in the estimated performance of a marker or limit the applicability of study findings, subsequently leading to unjustified optimism or premature rejection, contributing to translational failure [12–16].

Here, we report a systematic review of study design features used in recent evaluations of the clinical performance of ovarian cancer biomarkers.

Methods

Literature search

We performed a search on 22.12.2016 for reports of studies evaluating biomarkers in ovarian cancer in PubMed (MEDLINE). The search was limited to 2015 to obtain recent studies already indexed in MEDLINE.

The search strategy was developed in collaboration with a medical information specialist (RS) (Supplementary material, Table 1). Based on sample sizes from similar systematic reviews, we aimed to include 200 studies [17].

Study selection

Articles were eligible if they reported a primary clinical study, evaluating one or more biomarkers, and included adult women diagnosed, screened, treated or monitored for any type of ovarian cancer. To distinguish clinical evaluation studies from studies of other phases (primarily discovery studies) we defined a clinical study as a study that included the assessment of a previously discovered biomarker and reported a clinical performance measure that could be used to inform clinical decision-making.

We relied on the 1998 National Institutes of Health definition of a biomarker [18], including not only markers from body fluids but also imaging markers, such as ultrasound, CT, MRI and other modalities. Screening of titles and abstracts and full text evaluations was done in duplicate by two independent reviewers (MG and MO). Disagreements were solved through discussion; a third reviewer (PB) was consulted if consensus was not reached.

Data extraction

The study features were identified from previous commentaries, studies, checklists and quality assessment tools [3, 11, 12, 19–22] (Table 1). Data extraction was performed with a dedicated form by one reviewer (MO); unclear items were discussed with two other reviewers (MG and PB). Extraction guidance, as used in data-extraction, is provided in Supplementary material, Table 2.

Statistics

We calculated the proportion of studies with each respective feature, presented as estimates and 95% confidence intervals (CIs). We used Fisher's exact test to evaluate differences and a Kruskal-Wallis test for differences in sample size between subgroups. Two-sided p-values below 0.05 were considered as pointing to statistically significant differences. Calculations were performed in R (version i386 3.4.3).

Results

Search and study selection

Our search resulted in 1026 articles, of which 516 (49%) reports were considered potentially eligible after screening titles and abstracts, and 329 eligible (32%) after reading the full text (Figure 1). Of these, we evaluated the first 200, in chronological order of publication, starting January 1st 2015 to the most recent. The evaluated studies had been published in 95 journals from January 2015 until January

Table 1: Frequencies of study design features.

Design features and collection characteristics	Total	95% CI	Prognostic and predictive	Diagnostic	Other	p-Value
Intended use	n = 200		n = 140 (70%)	n = 36 (18%)	n = 24 (12%)	
Reporting of eligibility	34 (17%)	[12%–23%]	30/140 (22%)	3/36 (8%)	1/24 (4%)	p = 0.04
Reporting of sampling method	19 (10%)	[6%–14%]	12/140 (9%)	5/36 (14%)	2/24 (8%)	p = 0.55
Protocol	12 (6%)	[3%–10%]	6/140 (4%)	4/36 (11%)	2/24 (8%)	p = 0.16
Power calculation	5 (3%)	[1%–6%]	2/140 (1%)	1/36 (3%)	2/24 (8%)	p = 0.10
Multi-group	66 (33%)	[27%–40%]	37/140 (26%)	12/36 (33%)	16/24 (67%)	p < 0.01
Single-group	113 (57%)	[49%–64%]	91/140 (65%)	15/36 (42%)	7/24 (29%)	p < 0.01
Unclear	21 (11%)	[7%–16%]	12/140 (9%)	9/36 (25%)	1/24 (4%)	p = 0.02
Healthy controls	46 (23%)	[17%–30%]	17/140 (12%)	10/36 (28%)	19/24 (79%)	p < 0.01
Exclusively advanced stages as cases	5 (3%)	[1%–6%]	5/140 (4%)	0/36	0/24	p = 0.78
Multi-center	93 (47%)	[39%–54%]	60/140 (43%)	16/36 (44%)	17/24 (71%)	p = 0.04
Single-center	93 (47%)	[39%–54%]	72/140 (51%)	16/36 (44%)	5/24 (21%)	p = 0.02
Unclear	14 (7%)	[4%–12%]	8/140 (6%)	4/36 (11%)	2/24 (8%)	p = 0.46
Primary data	14 (7%)	[4%–12%]	7/140 (5%)	4/36 (11%)	3/24 (13%)	p = 0.19
Secondary data	182 (91%)	[86%–95%]	133/140 (95%)	29/36 (81%)	20/24 (83%)	p < 0.01
Routinely collected	130/182 (71%)	[64%–78%]	95/133 (71%)	22/29 (76%)	13/20 (65%)	p = 0.74
Including retrospective data	176 (88%)	[83%–92%]	129/140 (92%)	27/36 (75%)	20/24 (88%)	p = 0.01
Including prospective data	21 (11%)	[7%–16%]	11/140 (8%)	7/36 (19%)	3/24 (4%)	p = 0.10
Unclear	3 (2%)	[0%–4%]	0/140 (0%)	2/36 (8%)	1/24 (4%)	p = 0.03
Median sample size	156		132	145	657	p < 0.01
Min–max	13–50,078		13–6556	26–2665	31–50,078	
IQR	97–357		89–214	100–309	227–2366	
Smallest sample size used in analysis (median)	28 (of 102)		34 (of 70)	20 (of 20)	12 (of 12)	

The Table shows results for the total studies (n = 200) and in subgroups of intended use. Testing for differences between subgroups where performed by Fisher's exact test (two-sided), Kruskal-Wallis test for medians, and binomial test for 95% CI. "Others" include risk stratification [11], screening [4], monitoring [1] and studies with multiple use [8].

2016 and with a distribution ranging from one to 13 articles per journal (Supplementary material, Table 3) within both pre-clinical/translational and clinical journals.

The largest group of studies reported on prognostic and predictive biomarkers (70%). The second largest group consisted of studies describing markers for diagnostic purposes (18%) (Table 1). Across applications, we found a variety of different types of biomarkers and biomarker profiles including but not limited to clinical (risk) factors, as BMI and menopausal status, genetic profiles/mutations, as BRCA1/2, protein biomarkers, as CA-125 and HE4, clinical risk scores, as ROMA and RMI. The most frequently evaluated biomarkers were CA-125, HE4 and risk scores, evaluated either alone or in combinations. E-cadherin and clinical prognostic factors were also among the most frequently evaluated biomarkers (Supplementary material, Table 4).

The most frequently reported performance measures expressed the strength of associations, for example as hazard ratios or odds ratios, often accompanied by Kaplan-Meier survival analysis (54%). Other studies reported classification statistics, such as the area under the receiver operating characteristic (ROC) curve and ROC-statistics (24%).

Study design features

Recruitment of study participants

To evaluate the validity and applicability of the performance measures, study reports should include clear eligibility criteria and the methods for recruiting study participants. Of the 200 included study reports, 34 (17%) explicitly reported eligibility criteria and 19 (10%) sampling methods. Only 12 articles (6%) referred to an existing protocol (Table 1). As illustrated in Supplementary Table 5, the information provided on the identification and selection of study participants was often limited (Supplementary Table 5, Ex. 1) and even less detailed in analyses based on registries (Supplementary Table 5, Ex. 2).

Whenever the study group was described in study reports (n = 59, 30%), this was often done in rather broad and general terms, such as "sampled from the general

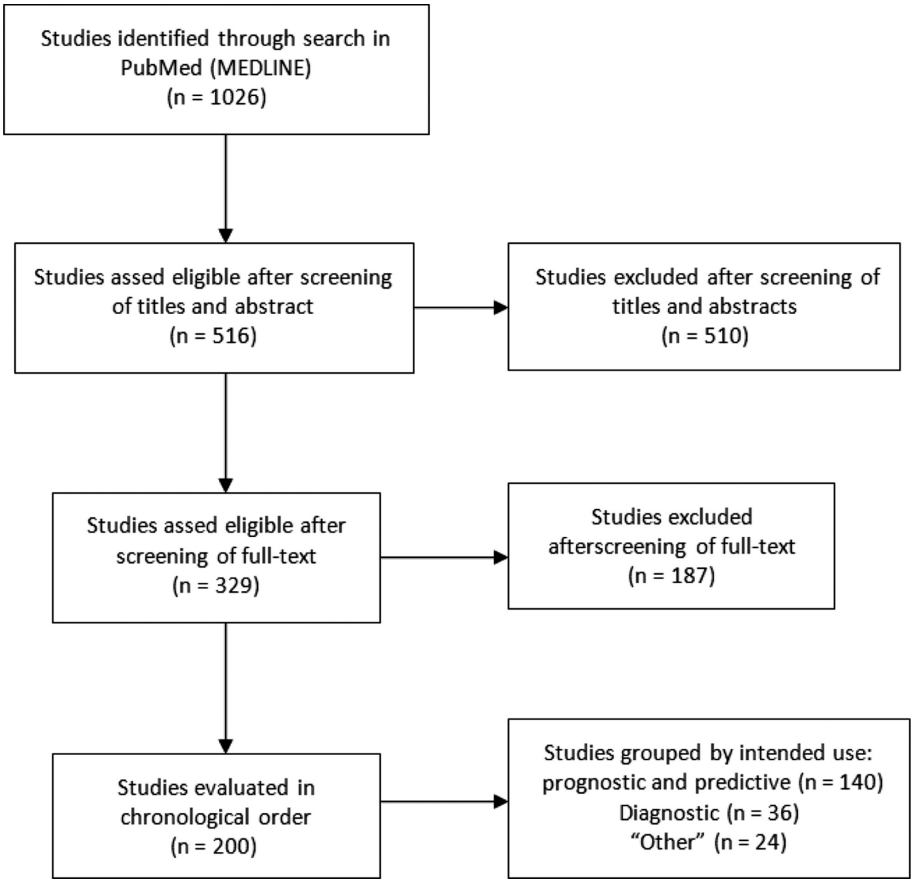

Figure 1: Flow diagram of studies.
Shows search results and study flow, including the distribution of intended use among the evaluated studies.

population" (n = 1) or "in women/patients with ovarian cancer/tumor" (n = 10). In other cases, this was described by nationality (n = 8), subtype (n = 18) or symptom(s) (n = 3). In contrast, a few studies had a description very specific to treatment or outcome (n = 6).

Single vs. multiple groups

In evaluations of the clinical performance of biomarkers, study participants should represent the intended use population. Of the 200 studies in our sample, 113 (57%) had indeed included a single group of study participants (i.e. groups of comparison originated from one single study group). In contrast, 66 (33%) studies had recruited patients in multiple, disjoint groups (i.e. groups of comparison originated from separate study groups). Forty-six studies (23%) reported on healthy controls, although the definition of a healthy control varied between studies (Supplementary Table 5, Ex. 3, 4). The groups that were included, other than ovarian cancer patients, ranged from patients with benign conditions to participants with other diseases and conditions, also referred to as "controls" (Supplementary Table 5, Ex. 5, 6). In one study, patients served as their own control (Supplementary Table 5, Ex. 7). At the other end of the spectrum, five (3%) studies had exclusively included patients with advanced stages (III–IV), which was not entirely consistent with the stated target population and study objective (Supplementary Table 5, Ex. 8).

Single center vs. multicenter

If data for clinical evaluation are collected in a single center, there may be a concern about a lack of generalizability; multi-center studies with prospective data collection are therefore preferred. We found that samples and data had often been acquired from a single center (93 studies; 47%). The majority of the studies (182; 91%) relied on previously collected samples (Table 1). Of these, 130 studies (71%) used samples collected during

routine clinical care (Supplementary Table 5, Ex. 1) while 31 (17%) used data from external registries of molecular data, of which 21 (68%) had used The Cancer Genome Atlas (TCGA) registry (Supplementary Table 5, Ex. 2). Most studies analyzed retrospectively collected data (176 studies; 88%) only 21 (11%) had collected data prospectively (Table 1).

Sample size

The number of patients in biomarker studies should be high enough to arrive at sufficiently precise estimates or to have enough power to test statistical hypotheses. In this review, the median sample size was 156 patients, ranging from 13 to 50,078, with an interquartile range from 97 to 357. Only five (3%) studies justified sample size, for example, by reporting a power calculation such as "A preliminary power analysis was performed to determine the number of patients needed to generate solid, meaningful data using Cochran's formulas. Based on this model under a 90% confident level, 0.5 standard deviation and ±10% CI, 68 EOC patients are needed to obtain confident results." [23] or justified by a sample sizes used in previous, similar studies such as "The number of sequenced individuals is within an acceptable range used previously to obtain significant results." [24]

Subgroup analysis

To assess whether frequencies of the design features differed between groups of biomarker studies defined by intended use, we classified the studies into nine groups (Supplementary material, Table 6). We found significant differences depending on the intended use of the biomarker in reporting of eligibility criteria, multi-group and single-groups, use of healthy controls, multi-center and single center, use of secondary and retrospective collected data and median sample size. Studies of biomarkers used for purposes other than prognostic, predictive or diagnostic more often included multiple groups, healthy controls, were often larger and designed as multi-center trials. In contrast, prognostic and predictive studies more frequently reported eligibility criteria and used a single group in their design.

In the 200 studies, we found one (0.5%) multi-center study that had recruited a single group of ovarian cancer patients (no separate controls) and clearly reported eligibility criteria, sampling method and sample size justification.

Discussion

In general, the field of biomarker research and medical tests is less well developed than the evaluation of pharmaceuticals and other interventions [8]. Despite the relatively large volume of published studies in ovarian cancer biomarker research, many putative markers have not been translated into clinical use [3, 12]. Shortcomings and deficiencies in study design have been suggested as a partial explanation for this translational failure. Our analysis of recently published evaluations of putative biomarkers provides systematic evidence for this hypothesis. Most studies in our sample were limited in size, performed in a single center and had often recruited multiple, disjoint groups of ovarian cancer cases and non-cancer controls.

As defined by Ransohoff and Gourlay, 2010, bias is "a systematic difference between the compared groups", which can give rise to differences caused by other factors than the one in question [25]. To this end, several authors, for example, have stressed the importance of identifying and selecting appropriate study participants and samples: those that represent the target population for the intended use. Failure to do so can lead to selection bias [4, 10, 13, 14, 25, 26].

Despite the many initiatives to improve reporting and transparency, such as the reporting guidelines – Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK), Strengthening the Reporting of Observational Studies in Epidemiology (STROBE), Biospecimen Reporting for Improved Study Quality (BRISQ), Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) and Standards for Reporting of Diagnostic Accuracy Studies (STARD) [27–31], we found that eligibility criteria and sampling methods were rarely reported As a consequence, we were not able to analyze in detail if the group of study participants actually matched the intended use population. Such incomplete reporting not only hampers secondary research but also the direct usefulness of a study report in clinical practice. However, the issues surrounding incomplete and not-transparent reporting have been addressed and documented elsewhere, by several other authors [32–34].

The use of multiple groups rather than a single group of study participants – preferably a consecutive series of patients – has been identified as a major source of bias in marker evaluations. Meta-epidemiological research has shown that the additional inclusion of other groups, in particular, the recruitment of healthy controls, is prone to lead to an overestimation of performance in diagnostic studies [11, 20, 26]. We found that one in three ovarian cancer marker studies relied on multiple, disjoint groups.

Almost one in four included some form of healthy controls. This may be surprising, as screening was not the intended use of most biomarkers, and application of the biomarker would not involve the testing of asymptomatic persons. The inclusion of healthy controls may be justified in the marker discovery phase, or for providing proof-of-principle, but the correct classification of these healthy, asymptomatic participants is not informative about the performance of the marker in clinical applications.

A majority of studies had used secondary and routinely collected data and many relied on retrospectively collected data. For the initial discovery phases, such convenient and readily accessible data and bio-specimen may be used. For a clinical evaluation, however, the data collection setting and conditions may not correspond to the clinical question [35, 36]. Single center studies were also relatively frequent, potentially limiting the generalizability of procedures and findings.

With a median sample size of slightly more than a hundred patients, most studies were relatively small, and, in particular, without sample size justification, the uncertainty around the estimated performance measures may still be considerable, hampering strong conclusions about the value of putative markers, or the lack thereof.

We investigated the shortcomings of studies on biomarkers in ovarian cancer, as in this framework their introduction may cover unmet clinical needs, such as early diagnosis or timely recognition of relapse [37]. However, as the selected design features in our study are generic for studies that evaluate biomarkers, we believe that similar shortcomings exist in biomarker evaluations in other cancers as well.

The included studies were published in a variety of different journals and we found only one study that were free of deficiencies. For these reasons, we believe that our results reflects the general practice in biomarker evaluations rather than being related to the journals in which the studies were published.

Proposals for diagnostic, prognostic and predictive biomarker studies have been made before [10]. An impressive number of authors, statisticians and others have written about the designs and analysis of biomarker evaluations. Many of the design limitations that we observed could therefore be explained by a lack of awareness in biomedical research. This could be addressed through more extensive training, promoting the use of study protocols, encouraging the assembly of multidisciplinary teams, involving experienced biostatisticians from the initial discovery phase, and fostering large international collaborations, such as the Ovarian Tumor Tissue Analysis (OTTA) consortium and the Ovarian Cancer Association Consortium (OCAC) [38, 39]. Such consortia could also help to achieve the targeted sample size for rare subtypes of ovarian cancer. Moreover, journal editors could demand better compliance to the reporting guidelines for primary studies, also as this may inform authors of how to better design a study for the individual clinical question.

Future commentaries and editorials in scientific journals about specific markers could additionally help to improve the practice of biomarker research, if they not only highlight the great potential of the putative biomarker, but also discuss the limitations in the research performed so far. These commentaries could, more consistently, highlight the need for real-world studies of the actual performance of biomarkers and the design of trials to document incremental effectiveness in improving patient outcomes, keeping the clinical context at the focus throughout biomarker development [8, 40]. As in intervention trials, involved stakeholders, such as companies that develop markers and funders, also need to facilitate such studies and trials.

We acknowledge a number of potential limitations of our own analysis. The data extraction form used to identify study features had not been used before. It was developed in close collaboration between two authors who also piloted it extensively, and most features were relatively easy to identify from the study reports, if reported at all. Reporting was often limited, hampering identification of some of the critical study features. Our set of design features evaluated in this review does not cover all aspects of methodological quality of the included studies; we focused primarily on recruitment and sampling, and selected features because they had been highlighted before in commentaries and methodological analyses of other areas of testing and biomarker research.

Conclusions

The search for new biomarkers, fueled by the impressive advances in omics-research, continues to hold great promise for clinical medicine. Yet, to fulfill this promise we need to increase the number of well-executed studies, with properly selected participants recruited in sufficient numbers. Although almost half of the studies were multicenter and more than half were single-group studies, we found only one study that was free of the selected shortcomings. Working in cooperation, in multidisciplinary groups and in larger consortia, could therefore be the way forward, starting fewer but higher-quality studies that can produce results that are at low risk of bias and more readily

interpretable. This may avoid premature claims of biomarker performance, prevent the unwarranted removal of promising markers, and eventually produce the new tools that ovarian cancer patients can benefit from.

Acknowledgments: We acknowledge the advice and help from Els Goetghebeur, Chris Hyde, Van Nguyen Thu and Rene Spijker.
Author contributions: All authors made a significant contribution to this study. MO: Study design, data collection, data analysis, writing the manuscript; MG: Study design, data collection, writing the manuscript; CL: Data analysis, writing the manuscript; PB: Study design, data collection, data analysis, writing the manuscript. All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.
Availability of data and material: The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.
Research funding: This research was funded by H2020 Marie Sklodowska-Curie Actions, Funder Id: http://dx.doi.org/10.13039/100010665, Grant Number: 676207.
Employment or leadership: None declared.
Honorarium: None declared.
Competing interests: The funding organization(s) played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication.
Ethics approval and consent to participate: Not applicable.
Conflict of interest: The authors declare no potential conflicts of interest.

References

1. Liao C-I, Chow S, Chen L, Kapp DS, Mann A, Chan JK. Trends in the incidence of serous fallopian tube, ovarian, and peritoneal cancer in the US. Gynecol Oncol 2018;149:318–23.
2. Timmermans M, Sonke GS, Van de Vijver KK, van der Aa MA, Kruitwagen RF. No improvement in long-term survival for epithelial ovarian cancer patients: a population-based study between 1989 and 2014 in the Netherlands. Eur J Cancer 2018;88:31–7.
3. Diamandis EP. The failure of protein cancer biomarkers to reach the clinic: why, and what can be done to address the problem? BMC Med 2012;10:87.
4. Pavlou MP, Diamandis EP, Blasutig IM. The long journey of cancer biomarkers from the bench to the clinic. Clin Chem 2013;59:147–57.
5. Duffy MJ, Sturgeon CM, Soletormos G, Barak V, Molina R, Hayes DF, et al. Validation of new cancer biomarkers: a position statement from the European Group on tumor markers. Clin Chem 2015;61:809–20.
6. Ioannidis JP, Bossuyt PM. Waste, leaks, and failures in the biomarker pipeline. Clin Chem 2017;63:963–72.
7. Rifai N, Gillette MA, Carr SA. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. Nat Biotechnol 2006;24:971–83.
8. Horvath AR, Lord SJ, St John A, Sandberg S, Cobbaert CM, Lorenz S, et al. From biomarkers to medical tests: the changing landscape of test evaluation. Clin Chim Acta 2014;427:49–57.
9. Pepe MS, Feng Z. Improving biomarker identification with better designs and reporting. Clin Chem 2011;57:1093–5.
10. Pepe MS, Feng Z, Janes H, Bossuyt PM, Potter JD. Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: standards for study design. J Natl Cancer Inst 2008;100:1432–8.
11. Rutjes AW, Reitsma JB, Di Nisio M, Smidt N, van Rijn JC, Bossuyt PM. Evidence of bias and variation in diagnostic accuracy studies. Can Med Assoc J 2006;174:469–76.
12. Ioannidis JP. Biomarker failures. Clin Chem 2013;59:202–4.
13. Diamandis EP. Cancer biomarkers: can we turn recent failures into success? J Natl Cancer Inst 2010;102:1462–7.
14. Ransohoff DF. Opinion: bias as a threat to the validity of cancer molecular-marker research. Nat Rev Cancer 2005;5:142–9.
15. Ransohoff DF. How to improve reliability and efficiency of research about molecular markers: roles of phases, guidelines, and study design. J Clin Epidemiol 2007;60:1205–19.
16. Leung F, Diamandis EP, Kulasingam V. Ovarian cancer biomarkers: current state and future implications from high-throughput technologies. Adv Clin Chem 2014;66:25–77.
17. Tajik P, Zwinderman AH, Mol BW, Bossuyt PM. Trial designs for personalizing cancer care: a systematic review and classification. Clin Cancer Res 2013;19:4578–88.
18. Strimbu K, Tavel JA. What are biomarkers? Curr Opin HIV AIDS 2010;5:463–6.
19. Knottnerus JA, Muris JW. Assessment of the accuracy of diagnostic tests: the cross-sectional study. J Clin Epidemiol 2003;56:1118–28.
20. Lijmer JG, Mol BW, Heisterkamp S, Bonsel GJ, Prins MH, van der Meulen JH, et al. Empirical evidence of design-related bias in studies of diagnostic tests. J Am Med Assoc 1999;282:1061.
21. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155:529.
22. Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, et al. The REporting of studies conducted using observational routinely-collected health Data (RECORD) statement. PLoS Med 2015;12.
23. Pearl ML, Dong H, Tulley S, Zhao Q, Golightly M, Zucker S, et al. Treatment monitoring of patients with epithelial ovarian cancer using invasive circulating tumor cells (iCTCs). Gynecol Oncol 2015;137:229–38.
24. Chen X, Paranjape T, Stahlhut C, McVeigh T, Keane F, Nallur S, et al. Targeted resequencing of the microRNAome and 3′UTRome reveals functional germline DNA variants with altered prevalence in epithelial ovarian cancer. Oncogene 2015;34:2125–37.
25. Ransohoff DF, Gourlay ML. Sources of bias in specimens for research about molecular markers for cancer. J Clin Oncol 2010;28:698–704.
26. Furukawa TA, Guyatt GH. Sources of bias in diagnostic accuracy studies and the diagnostic process. Can Med Assoc J 2006;174:481–2.

27. Cohen JF, Korevaar DA, Altman DG, Bruns DE, Gatsonis CA, Hooft L, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. Br Med J 2016;6.
28. Altman DG, McShane LM, Sauerbrei W, Taube SE. Reporting Recommendations for Tumor Marker Prognostic Studies (REMARK): explanation and elaboration. PLoS Med 2012;9.
29. Vandenbroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE). Epidemiology 2007;18:805–35.
30. Moore HM, Kelly AB, Jewell SD, McShane LM, Clark DP, Greenspan R, et al. Biospecimen reporting for improved study quality (BRISQ). Cancer Cytopathol 2011;119:92–102.
31. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med 2015;162:55.
32. Korevaar DA, Wang J, van Enst WA, Leeflang MM, Hooft L, Smidt N, et al. Reporting diagnostic accuracy studies: some improvements after 10 years of STARD. Radiology 2015;274:781–9.
33. Simera I, Moher D, Hirst A, Hoey J, Schulz KF, Altman DG. Transparent and accurate reporting increases reliability, utility, and impact of your research: reporting guidelines and the EQUATOR Network. BMC Med 2010;8:24.
34. Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, Julious S, et al. Reducing waste from incomplete or unusable reports of biomedical research. Lancet 2014;383:267–76.
35. Simon RM, Paik S, Hayes DF. Use of archived specimens in evaluation of prognostic and predictive biomarkers. J Natl Cancer Inst 2009;101:1446–52.
36. Henry NL, Hayes DF. Cancer biomarkers. Mol Oncol 2012;6:140–6.
37. Ferraro S, Panteghini M. Making new biomarkers a reality: the case of serum human epididymis protein 4. Clin Chem Lab Med 2018 Dec 4. doi:10.1515/cclm-2018-1111 [Epub ahead of print].
38. The Ovarian Tumor Tissue Analysis consortium (OTTA). OTTA [Internet]. [cited 2018 Apr 3]. Available from: https://ottaconsortium.org/.
39. Ovarian Cancer Association Consortium (OCAC). OCAC [Internet]. [cited 2018 Apr 3]. Available from: http://apps.ccge.medschl.cam.ac.uk/consortia/ocac//aims/aims.html.
40. Monaghan PJ, Lord SJ, St John A, Sandberg S, Cobbaert CM, Lennartz L, et al. Biomarker development targeting unmet clinical needs. Clin Chim Acta 2016;460:211–9.

Supplementary Material: The online version of this article offers supplementary material (https://doi.org/10.1515/cclm-2019-0038).