

EVIDENCE REQUIRED BY HEALTH TECHNOLOGY ASSESSMENT AND REIMBURSEMENT BODIES EVALUATING DIAGNOSTIC OR PROGNOSTIC ALGORITHMS THAT INCLUDE OMICS DATA

Alexandre Barna

Assistance Publique - Hôpitaux de Paris, URC-éco

alexandre.barna@gmail.com

Teresita M Cruz-Sanchez

Karen Berg Brigham

Assistance Publique - Hôpitaux de Paris, URC-éco

Cong-Tri Thuong

Assistance Publique - Hôpitaux de Paris

Finn Boerlum Kristensen

Syddansk Universitet Det Sundhedsvidenskabelige Fakultet

Isabelle Durand-Zaleski

Assistance Publique - Hôpitaux de Paris, URC-éco

Objectives: Multi-analyte assays with algorithmic analyses (MAAAs) use combinations of circulating and clinical markers including omics-based sources for diagnostic and/or prognostic purposes. Assessing MAAAs is challenging under existing health technology assessment (HTA) methods or practices. We undertook a scoping review to explore the HTA methods used for MAAAs to identify the criteria used for clinical research and reimbursement purposes.

Methods: This review included only non-companion (stand-alone) tests that are actionable and that have been evaluated by leading HTA or insurer/reimbursement bodies up to September 2017.

Results: Twenty-five reports and articles evaluating seventeen MAAAs were examined, most of which have been developed in oncology. The two main models used were the EUnetHTA Core model and the Evaluation of Genomic Applications in Practice and Prevention ACCE framework. Clinical validity and utility criteria were used, as were economic, ethical, legal, and social aspects. Economic evidence on MAAAs was scarce, and there is no consensus on whether the perspectives used are sufficiently broad to include all relevant stakeholders.

Conclusions: Clinical utility and efficiency were the most used criteria, with stronger evidence needed linking the use of the algorithm with the clinical outcomes in real-life practice. HTA bodies must as well consider questions related to the analytical validity of MAAAs or with organizational aspects. The two main models, the EUnetHTA Core model and the ACCE framework, could be adapted to the assessment of MAAAs.

Keywords: Biomarkers, Assessment and regulation, Reimbursement policy, Personalized medicine

“Omics” (genomics, transcriptomics, proteomics, and metabolomics) encompass multiple sets of biological molecules, including DNAs, RNAs, proteins, and metabolites. Advances in the identification and use of omics-based biomarkers have facilitated stratification of patients based on risk. They have allowed improved prognosis, treatment selection, and measurement of response to treatment and outcomes for different diseases, especially in oncology. Among the molecular diagnostic or prognostic tests that have been developed, multi-analyte assays with algorithmic analyses (MAAAs) explore

multiple biomarkers in conjunction with clinical data. MAAA applies multi-parametric statistical algorithms to individual biomarkers, thus improving the positive and negative predictive value of a diagnostic or prognostic test. Diagnostic MAAAs help to determine whether a patient has a disease at the time the test is performed, while prognostic MAAAs predict the aggressiveness and the likelihood of relapse to determine the best management for the patient (personalized medicine). Some MAAAs may be used for both diagnostic and prognostic purposes.

MAAAs and other molecular diagnostic algorithms are regulated as *in vitro* diagnostic (IVD) medical devices (1) in both Europe and the USA. Medical device regulation is not formally centralized in Europe, and the IVD rules, under review since 2012 (2), have only recently been promulgated (3). As a result, the legal context surrounding MAAAs is not yet well-defined. The situation is also unclear in the United States, where regulatory supervision by the Food and Drug

This work was supported by funding from the European Union’s Horizon 2020 research and innovation program under grant agreement No. 633983 ENSAT-HT project.

This work greatly benefited from the observations made by Dr. Iñaki Gutierrez Ibarluzea, who served as a reviewer of this study. With his consent, we thank him and acknowledge his involvement and the quality of his suggestions.

[†]The title has been updated since original publication. An erratum notice detailing this change was also published (DOI: 10.1017/S0266462318000661).

Administration (FDA) depends upon whether or not the test is developed by a laboratory (4).

While the field of molecular diagnostics has grown rapidly, the use of such algorithms in clinical practice remains relatively limited (5). This may be due to a lack of awareness on the part of physicians, unproven utility in real world clinical settings as well as limited availability due to coverage/reimbursement issues. Health technology assessment (HTA) is key to the latter as it evaluates the clinical, economic, organizational, social, ethical, and legal issues of a health intervention to inform policy decision making including reimbursement by health systems (6). However, developers of MAAAs may have difficulties in meeting the evidentiary standards used in HTA evaluations, such as demonstrating the existence of a relationship between test results (clinical validity) and improved patient outcomes (clinical utility).

We undertook a review of the literature to explore methods and challenges for the assessment of MAAAs to identify the criteria that could be used for reimbursement purposes. The objectives of this review were to identify existing evaluation frameworks used for MAAAs, identify MAAAs that have undergone HTA evaluations to date, undertake an analysis of the most widely evaluated MAAAs and synthesize the particular challenges that MAAAs present to HTA bodies.

METHODS

We undertook a scoping review of the scientific and gray literature available in English and French (reports in other languages were summarily explored), structured around a five-phase framework for scoping reviews: (i) identifying the research question; (ii) identifying relevant studies; (iii) study selection; (iv) collating, summarizing, and reporting the results; (v) consultation (7).

Three researchers independently searched through the literature to refine the research question and identify key concepts and terminology, key studies, adapted assessment tools, models, and published assessments of MAAA diagnostic/prognostic tests. Keywords such as MAAAs, omics, assays with algorithmic analyses, biomarkers, genetic tests, combined with evaluation, technology assessment were used in the scoping search. A fourth senior researcher then reviewed the outputs, which are summarized in this report.

Identification of MAAAs That Have Been Evaluated by Leading HTA or Insurer/Reimbursement Bodies

A snowball sampling approach was used to identify MAAAs and to expand to additional sources as outlined in Figure 1. Five different starting points were used consisting of tests identified in two review articles analyzing coverage policies (5;8), tests listed by the FDA (9) and by the Tufts Evidence-based Practice Center under contract to the USA Agency for Healthcare Research and Quality (10), as well as tests identified

by a scoping search performed from February to March 2016 using PubMed and Google.

Tests Inclusion/Exclusion Criteria. Only multi-analyte, non-companion (stand-alone) diagnostic or prognostic tests that (i) provide actionable results (i.e., can lead to changes in the clinical management of patients), (ii) concern a chronic disease, and (iii) have been evaluated by leading HTA bodies (up to September 2017) were included. We excluded companion diagnostics because they are developed in the context of Phase II or Phase III drug trials and are evaluated under the regulatory scheme for drugs rather than *in vitro* medical devices.

Identification of Evaluated MAAAs. Preliminary searches revealed that the most widely MAAA assessed was Oncotype DX (ODX). As a result, we initially identified institutions performing reimbursement and coverage decisions for ODX. The websites of these institutions were searched for reports on other MAAAs. To be as inclusive as possible, the websites of the HTA bodies listed in the directory of the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) were searched (11). The assumption was that, in this way, we would not only identify existing reports, but also we would be able to obtain more documentation on the assessments performed. Only reports in English and French were included, although reports in other languages were also summarily explored. This search was completed by examining reports generated by USA and Canadian healthcare providers' organizations and insurance companies.

RESULTS

Scoping Review

We identified seventy-two potentially relevant reports/articles in our scoping search. We excluded forty-seven items that did not meet our inclusion criteria. This left twenty-five HTA reports evaluating seventeen MAAAs (Figure 2).

Frameworks for Evaluating Molecular Diagnostic Tests

We identified two main frameworks used in HTA evaluations of diagnostic tests including MAAAs: the EUnetHTA Core Model[®] for diagnostics and the EGAPP's ACCE framework. Other organizations and initiatives developed guidelines and models, such as the PHGEN II European network (12). The Canadian Agency for Drugs and Technologies in Health (CADTH) also published a report in 2012 identifying the main evaluation frameworks that were used internationally and outlining some of the most commonly used criteria (13).

Among the concepts used, analytic validity corresponds to the capacity of a test to precisely and reliably measure a genotype in the laboratory (*in vitro*) to predict a clinical disorder or phenotype of interest, such as overall survival. It includes

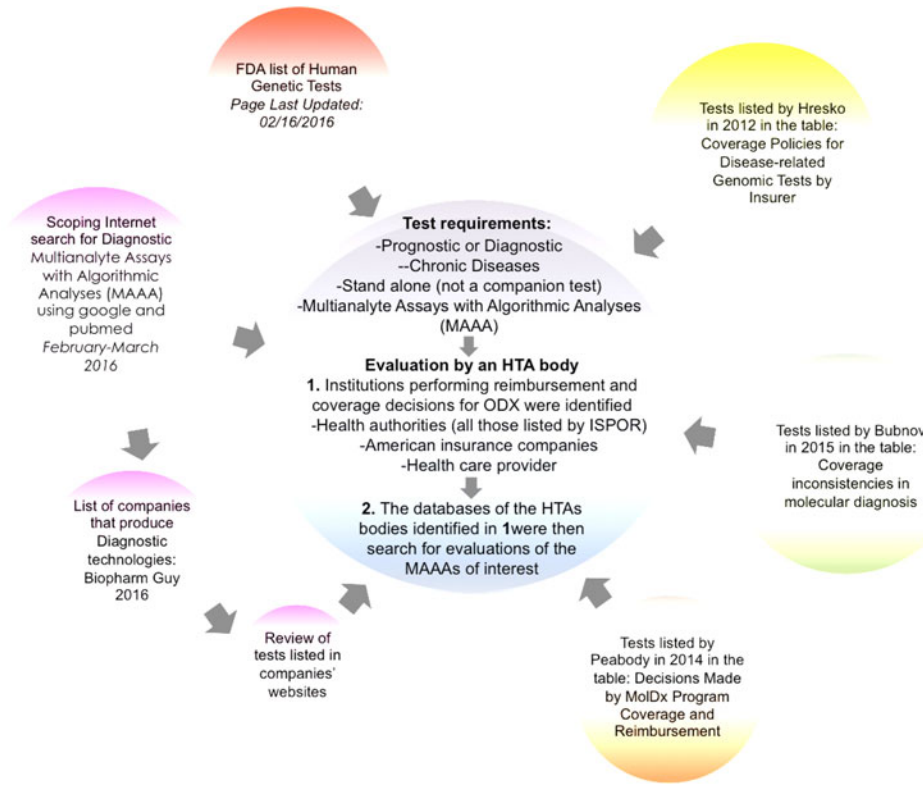


Figure 1. Identification of multi-analyte assays with algorithmic analyses (MAAAs) that have been evaluated by at least one health technology assessment (HTA) or insurer/reimbursement body. ODX, Oncotype DX.

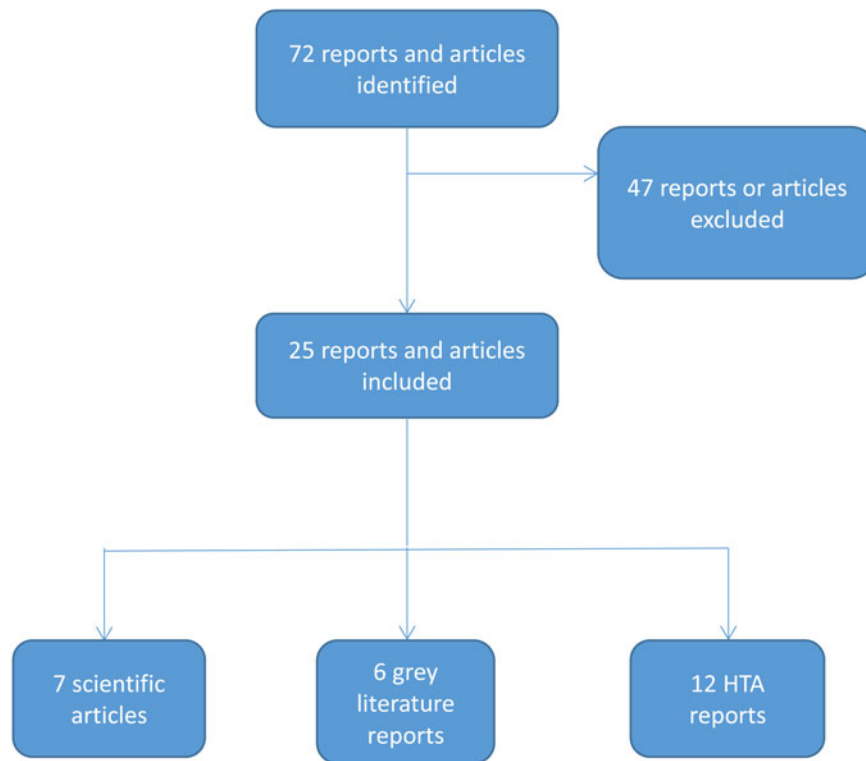


Figure 2. Flowchart of the scoping review. FDA, United States Food and Drug Administration; HTA, health technology assessment; ISPOR, International Society for Pharmacoeconomics and Outcomes Research; ODX, Oncotype DX.

sensibility, specificity, reliability, and reproducibility. It also encompasses positive and negative predictive values and disease prevalence. Clinical utility measures the ability of a test to improve patient care using clinical outcomes to assess the value for patient care. Finally, the ethical, legal, and social implications consider what impediments exist and what safeguards are required.

EUnetHTA Core Model[®]. The main aim of the EUnetHTA Core Model[®] is to facilitate collaboration between European and international HTA agencies by means of an HTA ontology containing an extensive list of generic questions, methodological guidance to help researchers find answers to these questions, and a common reporting structure providing a standardized format for the output of HTA projects. The third revision of the EUnetHTA Core Model[®] published in 2016 (14) is structured according to domains, topics, and issues.

Initially developed for medical and surgical therapeutic applications, it has been adapted for diagnostic and screening purposes. An evaluation of breast cancer MAAAs was undertaken by EUnetHTA using the Core Model framework (see section 3.3). By its own admission, the EUnetHTA Core Model[®] is not well adapted to prognostic tests, which include several MAAAs currently available. A study produced by EUnetHTA in 2015 discusses the study design that could be applied to personalized medicine technologies to ensure that the evidence available is at the same level as that of other health technologies (15).

The ACCE Framework. In 2000, the ACCE framework—Analytical validity; Clinical validity; Clinical utility; and Ethical, legal, and social aspects—was developed by the USA Office of Public Health Genomics (OHGP). Between 2005 and 2013, the Evaluation of Genomic Applications in Practice and Prevention (EGAPP[™]) working group undertook evidence reviews and published eight recommendations based on the ACCE framework (16).

Since 2013, the USA Centers for Medicare and Medicaid Services (CMS) has outsourced the assessment of molecular diagnostic tests, including MAAAs. Assessments under the MolDx program developed in 2011 by Palmetto GBA, a CMS Medicare Administrative Contractor, are based on the ACCE framework (17). The technology assessment framework summarizes all evidence of clinical validity and clinical utility when seeking CMS coverage and reimbursement (5).

The EGAPP working group based the ACCE framework on forty-four questions related to five main criteria: analytical validity, clinical validity and clinical utility, as well as ethical, legal, and social aspects (18).

In 2013, the working group initiative undertook a review of its work (16), examining the quality of evidence and the particular challenges in evaluating molecular tests. They found that evidence on analytic validity and clinical validity was

sparse, that reviews were time-consuming and expensive compared with the rapid evolution of genomic technologies, that modeling could be useful in estimating the benefits and harms of genomic technologies, and that greater stakeholder involvement was necessary in developing evaluation recommendations. The group, therefore, proposed that rapid ACCE reviews taking less than 8 weeks and costing less than \$20,000, using panels of experts and independent reviewers, may be suitable for topics lacking a large evidence base (19).

The ACCE model has been adopted and adapted by several European bodies, including the UK Genetic Testing Network (GTN) (20), the European Commission project EuroGentest (21), and the British PHG Foundation (22;23).

MAAA Tests That Have Been Evaluated by HTA or Insurer/Reimbursement Bodies

Seventeen MAAAs found in twenty-five reports met our criteria (Table 1). Thirteen of the seventeen tests were for cancer indications, most frequently breast cancer: ODX (HTA assessments found in fourteen countries) and MammaPrint (assessed in six countries). MAAAs targeting four noncancer indications, diabetes, obstructive coronary artery disease (CAD), heart transplant, and rheumatoid arthritis (RA), were also identified. Assessments for these noncancer MAAAs were only found in the United States. Overall, ten of the seventeen tests have been assessed only by HTA bodies in the United States.

To further explore the critical elements considered by HTA bodies in assessing MAAAs, we undertook a closer analysis of the most widely evaluated MAAA: ODX for breast cancer.

Evaluations of ODX/MammaPrint for Breast Cancer Carried on by HTA or Insurer/Reimbursement Bodies. The ODX algorithm generates a recurrence score (RS) that identifies the likelihood of distant recurrence of breast cancer as low risk (RS <18), intermediate risk (RS 18–30), and high risk (RS ≥31). It is designed to better target chemotherapy to higher risk patients by estimating baseline risk and response to chemotherapy.

We examined HTA evaluations of ODX by nine HTA bodies in five countries, identifying the methods and criteria applied, with a particular focus on clinical utility and economic analysis (Table 2).

In terms of clinical utility, the greatest challenge identified was the lack of direct evidence of improved health outcomes. While the impact on clinical decision making was emphasized, with moderate evidence showing that ODX led to changes in treatment decisions, none of the assessments examined both physician decision-making and downstream health outcomes for patients. Indirect evidence of clinical utility was based on the correlation between the risk of disease recurrence and thus the likelihood of chemotherapy benefit.

Cost-effectiveness of this MAAA test was assessed by five HTA bodies, each of which relied upon Markov models adopting the payer's perspective.

Table 1. MAAAs Evaluated by at Least One HTA or Insurer/Reimbursement Body

	MAAA	Disease	Description	Country
Non-cancer	Corus CAD	Obstructive coronary artery disease	Expression of 23 genes (RNA; RT-PCR)	USA
	Allomap	Heart transplant	RNA from each gene in a 20-gene panel (RT-PCR)	USA
	HeartsBreath	Heart transplant	Measures the abundance of C4-C20 alkanes and monomethylalkanes in the breath	USA
	VectraDA	Rheumatoid arthritis	Measures serum concentrations of 12 proteins	USA
	PreDX	Diabetes	7 different biomarkers (glucose, insulin, proteins)	USA
Cancer	Oncotype DX	Breast cancer	Expression of 21 genes (RNA; RT-PCR)	USA, UK, IE, FR, BE, CH, IT, ES, DE, GR, NL, CA, IL, AU, NZ
	MammaPrint	Breast cancer	mRNA level of 70 genes (micro-array)	USA, UK, FR, BE, DE, AU, NZ
	MammoStrat	Breast cancer	Expression of 5 proteins (IHC)	USA, UK, BE, AU
	Endopredict	Breast cancer	Expression of 12 genes (RNA; RT-PCR)	USA, UK, BE, AU
	Prosigna	Breast cancer	RNA expression profile of 58 genes (DNA probes)	USA, UK, AU
	UroVysion FISH test	Bladder cancer	Targets aneuploidy of 3 chromosomes or the loss of a genetic locus (DNA; FISH)	USA, UK, AU
	Oncotype DX	Colon cancer	12 gene assay (RNA; RT-PCR)	USA, UK, IS
	Cologuard (multi-target stool DNA testing)	Colon cancer	11 molecular markers; 10 DNA markers (gene expression and methylation status (PCR)) and hemoglobin (IHC)	USA
	Lung RS	Lung cancer	Expression of 14 genes (RNA; RT-PCR)	USA
	MyPRS Plus™ GEP70	Myeloma	Expression of 70 genes (RNA; RT-PCR)	USA
	Oncotype DX	Prostate cancer	Expression of 17 genes (RNA; RT-PCR)	USA
	ResponseDX : Tissue of Origin Test™	Cancers of unknown primary	Expression of 1550 genes (RNA; cDNA microarray)	USA

AU, Australia; BE, Belgium; CA, Canada; CH, Switzerland; DE, Denmark; ES, Spain; FISH, fluorescence in situ hybridization; FR, France; GR, Germany; IE, Ireland; IHC, immunohistochemistry; IL, Israel; IS, Iceland; IT, Italy; NL, The Netherlands; NZ, New Zealand; RT-PCR, reverse transcriptase polymerase chain reaction; UK, United Kingdom; USA, United States.

Three HTA bodies used Budget Impact Analysis (BIA) as part of their economic analyses. Overall, the budget impact depended on whether the test was made available to all breast cancer patients or to a risk-specific subgroup.

The studies consistently showed an overall shift to less-intensive treatment recommendations as a result of ODX. While lower use of chemotherapy would result not only in reduced costs but also presumably in lower exposure to harms, none of the studies followed patients to assess the overall balance of clinical benefits and harms. However, some evaluations considered the effect of ODX on decisional conflict (i.e., whether use of the test increased patients' confidence in the treatment decision).

In 2012, EUnetHTA assessed prognostic tests for breast cancer recurrence and stressed the fundamental differences between diagnostic and prognostic tests. The first version of the EUnetHTA Core Model was considered insufficient and not suited to prognostic technologies (24).

In January 2018, the National Institute for Health and Care Excellence (NICE) (25) released for public consultation an

evidence overview regarding the tumor profiling tests for breast cancer (EndoPredict, MammaPrint, Oncotype DX Breast Recurrence Score, Prosigna, and IHC4+C). The outcomes used for the assessment of clinical and cost-effectiveness were: (i) Intermediate measures: time to test results, analytical validity, prognostic ability, ability to predict benefit from chemotherapy, impact of test results on decision making. (ii) Clinical outcomes: disease-free survival, overall survival, distant recurrence, disease-related morbidity and mortality, chemotherapy-related morbidity and mortality. (iii) Patient-reported outcomes: health-related quality of life, anxiety. (iv) Costs were considered from an NHS and Personal Social Services perspective and included: costs of treating breast cancer, including drug cost, administration cost, outpatient appointments, and treatment of adverse events, costs of the tests, including equipment costs and reagents when relevant, costs of staff and associated training. (v) The cost-effectiveness measure by incremental cost per quality-adjusted life year.

In February 2018, EUnetHTA released an assessment report regarding the added value of using MammaPrint for

Table 2. Criteria Used by HTA or Insurer/Reimbursement Bodies in Evaluating Oncotype DX (Breast Cancer)

HTA body	Country	Method			Modeling			Meta-analysis			Systematic review			
		Information	Cost-effectiveness	Economic Impact	Overall decrease in chemotherapy recommendations	Prognostic ability	Clinical validity	General	Impact on clinical decision-making	Impact on patient anxiety	Chemotherapy-associated benefits	Patient outcomes	Clinical utility	
													Impact on clinical decision-making	Impact on patient anxiety
NCPe	Ireland	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
NICE	UK	✓				✓	✓	✓	✓	✓	✓	✓	✓	✓
EGAPP	USA				✓			✓	✓	✓	✓	✓	✓	✓
Palmetto	USA				✓			✓	✓	✓	✓	✓	✓	✓
MolDX	USA				✓			✓	✓	✓	✓	✓	✓	✓
Medicare	USA				✓			✓	✓	✓	✓	✓	✓	✓
Blue Cross	USA				✓			✓	✓	✓	✓	✓	✓	✓
Clalit Health	Israel	✓		✓				✓	✓	✓	✓	✓	✓	✓
Alberta HTA	Canada	✓		✓				✓	✓	✓	✓	✓	✓	✓
Ontario HTA	Canada	✓		✓				✓	✓	✓	✓	✓	✓	✓

EGAPP, Evaluation of Genomic Applications in Practice and Prevention; HTA, health technology assessment; NCPe, National Centre for Pharmacoeconomics; NICE, National Institute for Health and Care Excellence; UK, United Kingdom; USA, United States.

adjuvant chemotherapy in early breast cancer (26). The authors determined that MammaPrint testing has not yet demonstrated improved patient outcomes due to withholding adjuvant chemotherapy and thus its clinical utility has not been proven.

Analysis of Assessments of Non-cancer MAAAs Conducted by HTA or Insurer/Reimbursement Bodies. The evaluations of non-cancer MAAAs (ncMAAAs) focused on clinical utility, clinical validity, or both (Table 3). The ncMAAAs related to very different diseases and had disparate objectives: Corus CAD was used in the diagnosis of CAD, Allomap in the diagnosis and prognosis of acute cellular heart transplant rejection, Vectra DA as a measurement for RA disease progression, and PreDX to stratify patients into low and high risk for developing diabetes.

HTA bodies evaluated the clinical utility of the ncMAAAs by the impact on clinical decision making/management and on patient outcomes, as well as the relationship between changes in management and patient outcomes, such as net reclassification improvement or stratification. All HTA bodies considered the ability of ncMAAAs to diagnose or predict future outcomes.

Clinical validity was most often evaluated by comparing the performance of ncMAAAs with alternative tests in current clinical practice. Comparators of Corus CAD included clinical predictors and imaging technologies. Allomap and Heartsbreath were compared with the current diagnostic gold standard; the disease activity measure Vectra DA was compared with patient-reported and activity measures of disease progression, while PreDX was compared with current diabetes risk scores.

The HTA bodies identified several limitations regarding the studies supporting the ncMAAAs. They were critical of industry sponsorship of studies, small sample sizes, lack of controls and the nature of the control groups used (e.g., in the case of Corus CAD, lack of paired anatomic studies or the use of historical controls was criticized), limited follow-up times and thus evaluation of short-term outcomes, as well as the inclusion/exclusion criteria that resulted in study populations potentially having biased disease prevalence or pretest probability (e.g., in the case of Allomap, only patients who had received a cardiac transplant more than 6 months previously, and who were at a lower risk of rejection compared with the first 6 months following transplantation, were eligible).

Upon comparison with alternatives, other limitations were identified: the reported sensitivity of a comparator of Corus CAD in some studies was considerably lower than generally reported in the literature, and in the case of VectraDA, only moderate level of agreement above chance with alternative scores of disease progression was cited. In terms of the impact on clinical decision making, the lack of standardized use of PreDx for clinical decision-making and the fact that the choices of interventions were left up to physician and/or

Table 3. Summary of the General Characteristics of the Studies Cited by HTAs or Insurer/Reimbursement Bodies in Their Evaluation of MAAAs

		Systematic review																	
		Clinical validity			Clinical utility														
		Diagnostic/prognostic capacity	General	Comparison to alternative	Correlation with response to treatment	Stratification	General	Impact on clinical decision making	Optimizes individual diagnostic evaluation	Effect on clinical outcome	Effect on patient outcome	Decrease number of invasive procedures	Relationship between changes in management and patient outcomes	Performance in men vs women	Integration of the test into patient management	Economic modeling evaluation	Overcomes problems associated with alternatives	Effect of targeted preventive interventions on patient outcomes	
MAAA	Body		✓		✓														
	Corus CAD																		
Allomap	BC/BS	✓		✓			✓	✓	✓	✓	✓	✓	✓						
	Humana	✓		✓			✓	✓	✓	✓	✓	✓		✓					
	MolDx	✓		✓			✓	✓	✓	✓	✓	✓							
	Aetna	✓		✓			✓	✓	✓	✓	✓	✓				✓			
	America health Caritas	✓		✓			✓	✓	✓	✓	✓	✓							
Heatsbreath	Anthem	✓		✓			✓	✓	✓	✓	✓	✓							
	United Health	✓		✓			✓	✓	✓	✓	✓	✓		✓					
	Anthem	✓		✓			✓	✓	✓	✓	✓	✓							
VectraDA	Healthnet	✓		✓			✓	✓	✓	✓	✓	✓	✓						
	Aetna	✓		✓			✓	✓	✓	✓	✓	✓							
	Capital blue	✓		✓			✓	✓	✓	✓	✓	✓							
PreDX	Anthem	✓		✓			✓	✓	✓	✓	✓	✓		✓					
	Capital blue	✓		✓	✓		✓	✓	✓	✓	✓	✓							
	MolDX	✓	✓	✓			✓	✓	✓	✓	✓	✓							
PreDX	Capital blue	✓		✓	✓		✓	✓	✓	✓	✓	✓							
	MolDX	✓		✓			✓	✓	✓	✓	✓	✓							
	Aetna	✓	✓	✓			✓	✓	✓	✓	✓	✓							
	CGS	✓		✓			✓	✓	✓	✓	✓	✓							
PreDX	Anthem	✓		✓			✓	✓	✓	✓	✓	✓							
	BC/BS	✓		✓		✓	✓	✓	✓	✓	✓	✓							✓

Investigational Medical necessary

BC/BS, Blue Cross/Blue Shield; CGS, ; HTA, health technology assessment; MAAA, multi-analyte assays with algorithmic analyses.

patient discretion, meant that it could not be determined whether observed changes would have occurred in the absence of any intervention or with only routine medical care.

The limitations identified by the HTA bodies regarding the different ncMAAAs did not predict whether a particular test was deemed investigational (and thus not reimbursable) or medically necessary. The cost-effectiveness of the ncMAAAs and their effect on preventive interventions and patient outcomes was given little attention. None of the HTA evaluations of the ncMAAAs mentioned any economic evaluation or meta-analysis.

DISCUSSION

We described the assessment processes for MAAAs mostly in Europe and North America. To our knowledge, it is the first review contributing to the discussion on the pertinent criteria for MAAA assessment and providing insights for clinical research and evidence development.

Among the difficulties encountered in identifying which HTA bodies had evaluated the MAAAs was the fact that assessments were not always undertaken by the national HTA body. For example, in France, the assessment of three breast cancer MAAAs was carried out initially by the national cancer institute (INCa), rather than the national HTA authority the *Haute Autorité de Santé* (HAS) (27). The evaluation of one breast cancer MAAA was included in the 2016 working program of HAS (28).

Despite multiple search methods, the data may simply not have been available. In a recent study examining the U.S. payer coverage policies (29), authors found that payers less often cited clinical studies, systematic reviews, technology assessments, and cost-effectiveness analyses in their coverage policies for multi-gene panels than in their policies for other intervention types.

Another reason few HTA assessments of non-companion MAAAs were identified in Europe may be the fact that HTA bodies have not adapted their processes to accommodate such evaluations, although this appears to be changing. In the United Kingdom, the NICE has developed a Diagnostics Assessment Programme (DAP) that mainly focuses on stand-alone molecular diagnostics (30). DAP essentially follows the same approach as applied to drugs and measures patient benefit in quality-adjusted life-years. With the public consultation published in early 2018 (25), the NICE broadened its scope to include MAAAs.

In Germany, the German Institute for Quality and Efficiency in Health Care (IQWiG), expressly referenced MAAAs and recommended that the same evaluation principles as for diagnostic tests be applied (31). In Spain, a guideline for the evaluation of genetic based technologies was elaborated by the Andalusian Agency for Health Technology Assessment (AETSA) (32). Indeed, these countries concurred in applying their assessment framework to MAAAs, with minor

adaptations. Analytic validity is normally established as part of the market access process. However, in the case of stand-alone MAAAs, this may not take place before HTA evaluations, and HTA bodies must, therefore, consider questions related to the analytical validity of MAAAs as well.

Finally, the two frameworks identified in our review, EUnetHTA Core Model[®] and the ACCE framework, have not been until now adopted by key European national HTA bodies (e.g., NICE, HAS, IQWiG), although these bodies are evaluating clinical utility which “plays a key role in the vast majority of reimbursement decisions” (33).

The concept of clinical utility for diagnostics has evolved in recent years, and the MoIDX program used by Medicare in the United States is considered by some to be the *de facto* measure (34). In the assessment process of MoIDX, organizational aspects must play an increasing role, especially linked with the external validity of algorithms. This aspect is taken into account in the EUnetHTA model but not in the ACCE framework developed by EGAPP. Other issues that have to be developed by the models are ethical ones, not least because of the genetic data that are communicated to manufacturers.

MAAAs are better-developed in cancer than in other disease areas. Nonetheless, even cancer test developers face challenges in providing evidence of clinical utility. In 2014, the USA Agency for Healthcare Research and Quality (AHRQ) undertook a systematic review of molecular pathology tests that estimate the prognosis (recurrence) of common cancers to determine whether there was evidence that test results changed physician decision making and improved clinical outcomes for patients (35). In terms of clinical utility, no studies directly assessing the impact of a test on both physician decision making and downstream health outcomes were found. In terms of impact of tests in changing treatment decisions, moderate strength of evidence was found only for ODX Breast. This important point must be addressed in future research and must be a requirement of HTA agencies, namely providing strong evidence that links the use of the algorithm with the clinical outcomes for the patient in real life practice. Difficulty in producing final outcomes/endpoints could also be due to the existence of a confounding factor, namely the clinical effectiveness of the treatment proposed in the case of a positive test.

Not surprisingly, HTA bodies use different criteria in evaluating cancer MAAAs versus non-cancer MAAAs. Targeted use of chemotherapy resulting in reduced costs and side effects to patients was found to be a crucial element in the assessment process for ODX. No such single strong driving criteria was found in any of the evaluations of ncMAAAs. In addition, impact on clinical outcomes was a criterion sought by HTA bodies in their evaluations of ncMAAAs, whereas impact on patient anxiety was a parameter of interest only in evaluations of ODX. Common to evaluations of both cancer and non-cancer MAAAs was their impact on clinical decision making and on patient outcomes; of interest, the relationship between

changes in patient management and improvement of patient outcomes was a parameter of interest mainly in the HTA evaluations of ncMAAAs.

Differences between algorithms for diagnostic and prognostic purposes were expected but not found. Indeed, the prognosis of a disease is determined by final results, and the best way to provide evidence is related to the final clinical outcomes. One explanation could be that, at this point in time, clinical research has provided few clinical endpoints to HTA bodies. Another could be that frameworks used are quite wide and differences in the precise criteria used for evaluating clinical utility were not accurately described in the HTA reports (29).

Neglecting the external validity of these algorithms could in part explain the limitations of their use in clinical practice toward the goal of personalized medicine. Regardless of the nature of the test being evaluated, the main problem encountered by HTA agencies for the assessment of clinical utility of diagnostic and prognostic tests was the lack of available data, which may be due to trial length and proprietary interests.

Economic evidence on MAAAs was scarce, and we found no consensus on whether existing economic evaluation methods were sufficient (36). Some of the methodological challenges pertaining to any diagnostic workup include the choice of comparator, perspective, and timeframe. Costing MAAAs may be difficult because of the need to collect a broad range of costs, often in a data-limited environment. Measuring the impact of MAAAs on clinical outcomes requires information on patients' and clinicians' behavior. Alternative metrics (e.g., personal utility) are underdeveloped and alternative approaches (patients reported outcomes and the measure of health and non-health outcomes) are underused (37).

In developing trial protocols for MAAAs in Europe, the potential challenges and considerations for future HTA evaluations must be considered (5). They include the ability to measure clinical utility and to estimate long-term costs as well as benefit from the inputs of new categories of stakeholders concerned by precision medicine. The HTA Core Model recently became accessible free of charge to allow development of the model and broaden its future applications (37;38). MAAA assessments will be reinforced by new European regulations (3) designed to ensure that by May 26, 2022, increased requirements for clinical evidence (including an EU-wide coordinated procedure for authorization of multi-center clinical investigations), general safety, performance, and postmarketing surveillance are met.

The regulation also reinforces the criteria for designation and processes for oversight of Notified Bodies. This action was supported by PACITA and PerMed projects. The PerMed 2015 report (39) estimates that a Europe-wide process to evaluate and validate biomarkers, the development of new clinical trial designs adapted to these new approaches, and the integration of preclinical testing with innovative clinical trials may further improve the effectiveness of interventions.

In conclusion, the assessment of MAAAs that include “omics” is possible with current health technology assessment methods but requires some adjustments.

Stronger evidence is needed, especially on clinical utility, to link these algorithms with the clinical outcomes in real life practice. Existing economic evaluation methods appear to be insufficiently developed to evaluate MAAAs, although they continue to be applied in major systems.

CONFLICTS OF INTEREST

Dr. Barna has nothing to disclose. Ms. Cruz-Sanchez has nothing to disclose. Ms. Berg Brigham reports grants from European Commission, during the conduct of the study. C.T. Thuong has nothing to disclose. Dr. Kristensen has nothing to disclose. Dr. Durand-Zaleski reports personal fees from Roche, outside the submitted work.

REFERENCES

- Henschke C, Panteli D, Perleth M, Busse R. Taxonomy of medical devices in the logic of health technology assessment. *Int J Technol Assess Health Care*. 2015;31:324–330.
- Garrido MV, Kristensen FB, Nielsen CP, Busse R. Health technology assessment and health policy-making in Europe: Current status, challenges and potential [Internet]. WHO Regional Office Europe; 2008. 198 p. http://www.euro.who.int/_data/assets/pdf_file/0003/90426/E91922.pdf (accessed July 12, 2018).
- European Commission. Regulatory framework for medical devices [Internet]. Growth. [cited February 12, 2018]. https://ec.europa.eu/growth/sectors/medical-devices/regulatory-framework_en (accessed July 12, 2018).
- Food and Drug Administration. IVD Regulatory Assistance - Overview of IVD Regulation [Internet]. [cited October 11, 2017]. <https://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/IVDRegulatoryAssistance/ucm123682.htm> (accessed July 12, 2018).
- Peabody JW, Shimkhada R, Tong KB, Zubiller MB. New thinking on clinical utility: Hard lessons for molecular diagnostics. *Am J Manag Care*. 2014;20:750–756.
- Rosenkötter N, Vondeling H, Blancquaert I, et al. The contribution of health technology assessment, health needs assessment, and health impact assessment to the assessment and translation of technologies in the field of public health genomics. *Public Health Genomics*. 2011;14: 43–52.
- Armstrong R, Hall BJ, Doyle J, Waters E. Cochrane update. “Scoping the scope” of a Cochrane review. *J Public Health Oxf Engl*. 2011;33:147–150.
- Hresko A, Haga SB. Insurance coverage policies for personalized medicine. *J Pers Med*. 2012;2:201–216.
- Food and Drug Administration. In vitro diagnostics - Nucleic acid based tests [Internet]. [cited October 11, 2017]. <https://www.fda.gov/medical-devices/productsandmedicalprocedures/invitrodiagnostics/ucm330711.htm> (accessed July 12, 2018).
- Raman G, Avendano EE, Chen M. Update on emerging genetic tests currently available for clinical use in common cancers [Internet]. Rockville, MD: Agency for Healthcare Research and Quality (US); 2013. (AHRQ Technology Assessments). <http://www.ncbi.nlm.nih.gov/books/NBK285327/> (accessed July 12, 2018).
- International Society for Pharmacoeconomics and Outcomes Research. Directory of HTA Organizations Worldwide [Internet]. [cited October

- 20, 2017]. <https://www.ispor.org/HTADirectory/Index.aspx> (accessed July 12, 2018).
12. Brand A, Lal JA. European best practice guidelines for quality assurance, provision and use of genome-based information and technologies: the 2012 declaration of Rome. 2012. <https://www.degruyter.com/dg/view/article/jS002fdmdi.2012.27.issue-3S002fdmdi-2012-0026S002fdmdi-2012-0026.xml> (accessed July 12, 2018).
 13. Morrison A, Boudreau R. Evaluation frameworks for genetic tests (Environmental Scan: issue 36). Ott Can Agency Drugs Technol Health [Internet]. 2012. <http://scholar.google.com/scholar?cluster=1450821541151138323&hl=en&oi=scholar> (accessed July 12, 2018).
 14. European network for Health Technology Assessment. HTA core model version 3.0 [Internet]. [2016 December 1, 2016]. <http://www.eunetha.eu/wp-content/uploads/2018/03/HTACoreModel3.0-1.pdf> (accessed July 12, 2018).
 15. European network for Health Technology Assessment. Reflection paper. Personalised medicine and co-dependent technologies, with a special focus on issues of study design [Internet]. 2015 [cited February 7, 2015]. http://www.eunetha.eu/sites/default/files/sites/5026.fedimbo.belgium.be/files/news-attachments/2016-03-07_reflection_paper_pm_2nd_draft.pdf (accessed July 12, 2018).
 16. Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Working Group. The EGAPP initiative: lessons learned. *Genet Med*. 2014;16:217–224.
 17. Palmetto GBA. Palmetto GBA - MolDx - Technical Assessment (TA) Process (M00095, V9) [Internet]. [cited 2016 Nov 30]. <https://www.palmettogba.com/palmetto/MolDX.nsf/docsCat/MolDx%20Website~MolDx~Browse%20By%20Topic~Technical%20Assessment>
 18. Teutsch SM, Bradley LA, Palomaki GE, et al. The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Initiative: Methods of the EGAPP Working Group. *Genet Med*. 2009;11:3–14.
 19. Kroese M, Elles R, Zimmern R. The evaluation of clinical validity and clinical utility of genetic tests: Summary of an Expert Workshop, 26 and 27 June 2006. PHG Foundation; 2007.
 20. National Health Service. UK Genetic Testing Network [Internet]. [cited 2016 November 30, 2016]. <http://ukgtn.nhs.uk/> (accessed July 12, 2018).
 21. EuroGentest. [cited 2016 November 30, 2016]. <http://www.eurogentest.org/> (accessed July 12, 2018).
 22. PHG Foundation. [cited November 30, 2016]. <http://www.phgfoundation.org/> (accessed July 12, 2018).
 23. Burke W, Zimmern RL. Moving beyond ACCE: an expanded framework for genetic test evaluation. 2007. <http://www.phgfoundation.org/file/16270/> (accessed July 12, 2018).
 24. European network for Health Technology Assessment. Prognostic tests for breast cancer recurrence (PTBCR) [Internet]. [cited December 21, 2016]. <https://www.eunetha.eu/prognostic-tests-for-breast-cancer-recurrence-ptbcr/> (accessed July 12, 2018).
 25. National Institute For Health and Clinical Excellence. Tumour profiling tests to guide adjuvant chemotherapy decisions in people with breast cancer (update of DG10) [Internet]. [cited February 7, 2018]. <https://www.nice.org.uk/guidance/indevelopment/gid-dg10015/documents> (accessed July 12, 2018).
 26. European network for Health Technology Assessment. Added value of using the gene expression signature test MammaPrint[®] for adjuvant chemotherapy decision-making in early breast cancer [Internet]. [cited May 23, 2018]. https://www.eunetha.eu/wp-content/uploads/2018/01/EUnetha_assessment_mammaprint_final.pdf (accessed July 12, 2018).
 27. Luporsi E, Bellocq JP, Barrière J, et al. uPA/PAI-1, Oncotype DX[™], MammaPrint[®] Valeurs pronostique et prédictive pour une utilité clinique dans la prise en charge du cancer du sein. *Oncologie*. 2014;16:196–206.
 28. Haute Autorité de Santé. Working program [Internet]. [cited December 1, 2016]. https://www.has-sante.fr/portail/jcms/c_1267303/fr/programme-de-travail-de-la-has (accessed July 12, 2018).
 29. Chambers JD, Saret CJ, Anderson JE, et al. Examining evidence in U.S. payer coverage policies for multi-gene panels and sequencing tests. *Int J Technol Assess Health Care*. 2017;33:534–540.
 30. National Institute For Health and Clinical Excellence. Diagnostics assessment programme manual [Internet]. [cited December 1, 2016]. <https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/nice-diagnostics-guidance> (accessed July 12, 2018).
 31. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. General Method Version 4.2 [Internet]. [cited December 1, 2016]. <https://www.iqwig.de/en/methods/methods-paper.3020.html> (accessed July 12, 2018).
 32. Márquez Calderón S, Briones Pérez de la Blanca E. Framework for the assessment of genetic testing in the Andalusian Public Health System [Internet]. [cited 2017 July 21, 2017]. <https://pdfs.semanticscholar.org/8ae8/a5263d9775c3442bba2d9b71a3266ee06711.pdf> (accessed July 12, 2018).
 33. Akhmetov I, Bubnov RV. Assessing value of innovative molecular diagnostic tests in the concept of predictive, preventive, and personalized medicine. *EPMA J* [Internet]. [cited November 30, 2016];6. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4588236/> (accessed July 12, 2018).
 34. Palmetto GBA. Molecular Diagnostic Program (MolDX[®]) [Internet]. [cited December 1, 2016]. <http://www.palmettogba.com/moldx> (accessed July 12, 2018).
 35. Meleth S, Reeder-Hayes K, Ashok M, et al. Technology assessment of molecular pathology testing for the estimation of prognosis for common cancers [Internet]. Rockville, MD: Agency for Healthcare Research and Quality (US); 2014 [cited November 30, 2016]. (AHRQ Technology Assessments). <http://www.ncbi.nlm.nih.gov/books/NBK285410/> (accessed July 12, 2018).
 36. Garau M, Towse A, Garrison L, Housman L, Ossa D. Can and should value-based pricing be applied to molecular diagnostics? *Pers Med*. 2012;10:61–72.
 37. Buchanan J, Wordsworth S, Schuh A. Issues surrounding the health economic evaluation of genomic technologies. *Pharmacogenomics*. 2013;14:1833–1847.
 38. Kristensen FB, Lampe K, Wild C, et al. The HTA Core Model[®]—10 years of developing an international framework to share multidimensional value assessment. *Value Health*. 2017;20:244–250.
 39. PerMed. Shaping Europe's vision for personalised medicine [Internet]. [cited May 23, 2018]. http://www.permed2020.eu/_media/PerMed_SRIA.pdf (accessed July 12, 2018).