How is genetic testing evaluated? A systematic review of the literature

Erica Pitini1 · Corrado De Vito1 · Carolina Marzuillo1 · Elvira D’Andrea1,2 · Annalisa Rosso1 · Antonio Federici3 · Emilio Di Maria4 · Paolo Villari1

Received: 22 May 2017 / Revised: 7 November 2017 / Accepted: 19 December 2017
© The Author(s) 2018. This article is published with open access

Abstract
Given the rapid development of genetic tests, an assessment of their benefits, risks, and limitations is crucial for public health practice. We performed a systematic review aimed at identifying and comparing the existing evaluation frameworks for genetic tests. We searched PUBMED, SCOPUS, ISI Web of Knowledge, Google Scholar, Google, and gray literature sources for any documents describing such frameworks. We identified 29 evaluation frameworks published between 2000 and 2017, mostly based on the ACCE Framework (n = 13 models), or on the HTA process (n = 6), or both (n = 2). Others refer to the Wilson and Jungner screening criteria (n = 3) or to a mixture of different criteria (n = 5). Due to the widespread use of the ACCE Framework, the most frequently used evaluation criteria are analytic and clinical validity, clinical utility and ethical, legal and social implications. Less attention is given to the context of implementation. An economic dimension is always considered, but not in great detail. Consideration of delivery models, organizational aspects, and consumer viewpoint is often lacking. A deeper analysis of such context-related evaluation dimensions may strengthen a comprehensive evaluation of genetic tests and support the decision-making process.

Introduction
The increased availability of genetic tests has made the assessment of their performance crucial for clinical and public health practice. However, the evaluation of genetic tests, especially predictive ones, is not straightforward. The main challenge is the lack of scientific evidence on which to base such evaluations [1]. Generating scientific evidence on genetic tests is made difficult by different factors including their complexity, their rapid development and marketing, their widespread impact on families and society, and the lack of standardized outcomes for their evaluation [1]. For predictive genetic tests, perhaps the greatest challenge is to perform high quality, randomized control trials to demonstrate that the test confers an improvement in survival or quality of life [2]. Moreover, the lack of evidence on effectiveness affects the evaluation of cost-effectiveness.

Despite this, several frameworks have been proposed for the evaluation of genetic tests, but it is unclear how and in what respect they differ. The importance of a transparent and well-planned evaluation strategy is twofold. On the one hand, it would avoid the uncontrolled implementation of technologies without proven benefits, which can lead to inappropriate management of patients and detrimental effects on patient health, as well as a waste of resources and loss of public confidence in the medical profession. On the other hand, in line with the requirement for public health programs to maximize population health benefits, a reliable evaluation strategy would support the implementation of those currently available tests that have proven effectiveness and cost-effectiveness [3].

To guide the appropriate translation of genomics into clinical practice, Italy developed a National Plan for Public Health Genomics, which, to our knowledge, is the first specific policy example of its kind in Europe. It has various

Electronic supplementary material The online version of this article (https://doi.org/10.1038/s41431-018-0095-5) contains supplementary material, which is available to authorized users.
strategic objectives including the development of a well-planned evaluation strategy for genetic tests [4]. Our systematic review was conducted as part of a project financed by the Italian Ministry of Health to implement this plan and aims to identify and compare the existing evaluation frameworks for genetic tests, taking into account their methodology and evaluation criteria.

Materials and methods

This review was performed according to the Cochrane Handbook for systematic reviews of intervention and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [5, 6].

Selection criteria

We included any document that describes an original evaluation framework for genetic tests, defined as a structured process for the collection of the scientific evidence needed to assess the performance of a genetic test, from the laboratory to clinical practice. We excluded partial evaluation frameworks, defined as those embracing less than three evaluation components (analytic validity, clinical effectiveness, etc.). We limited our search to frameworks specifically created for the evaluation of genetic tests.

Search methods

Two reviewers (EP and CD) searched the bibliographic databases Pubmed, Scopus, ISI Web of Knowledge, Google Scholar, and the world wide web through Google for all English language articles between January 1990 and April 2017. The search terms were grouped as two strings: String A, “genetic testing OR genetic test OR genomic test OR genomic technology OR pharmacogenetic test” AND “evaluation OR assessment OR evaluating OR assessing OR evaluate OR assess” AND “framework OR criteria OR tool OR model OR process OR methods OR evidence based” OR “analytic validity OR clinical validity OR clinical utility”; and String B, “genetic testing OR genetic test OR genomic test OR genomic technologies OR pharmacogenetic test OR public health genomics OR pharmacogenetics OR pharmacogenomics” AND “health technology assessment” (see Supplementary Information for the full electronic search for each database). This search was supplemented by exploring the websites of government agencies and research organizations involved in the evaluation of genetic tests (see Supplementary Information for a list of the respective websites) and by scanning the reference lists of all the relevant articles retrieved. Moreover, experts of the Italian Network of Public Health Genomics were asked to share the evaluation frameworks they were aware of through a Delphi procedure.

Study selection

The two reviewers (EP and CD) removed duplicates and screened the title and abstract of all retrieved records. Studies that clearly did not meet the eligibility criteria were excluded. Full texts of potentially relevant studies were examined for inclusion in the systematic review and reasons for exclusion recorded. Disagreements were resolved by discussion.

Data collection and analysis

Two reviewers (EP and EDA) extracted the following information about the retrieved frameworks: authors; country; year of publication; reference institution; framework name; type of target test; reference frameworks; methodology (format, sources of evidence, quality of the evidence, grading of recommendations, research priorities); practical application; purpose; primary audience; evaluation components (see Supplementary Information for a definition of each category of information extracted). A narrative synthesis of the evaluation frameworks identified was performed, comparing their general features, evaluation components, and methodological aspects.

Results

Study selection

After removal of duplicates, 6027 records resulted from the initial search (Fig. 1). Screening by title and abstract selected 289 records for full text analysis, from which 30 records were selected. Reasons for exclusion were: documents not describing a framework for the evaluation of genetic tests; appraisals of individual genetic tests using an original framework described elsewhere; partial evaluation frameworks; documents focusing on only one evaluation component; broader frameworks of implementation research not proposing an original evaluation process; guidelines on the evaluation of genetic tests; reviews and commentaries of evaluation frameworks and evaluation criteria for genetic tests; full text not available. Six records were added to the previous 30 from the reference lists of relevant articles retrieved. A total of 36 records were included in the systematic review, describing 29 frameworks for the evaluation of genetic tests (some records describe the same framework) [7–42]. The Delphi procedure did not add new frameworks to those already retrieved.
Frameworks retrieved

The systematic search identified 29 frameworks from various countries (USA, \( n = 12 \); Canada, \( n = 4 \); Europe, \( n = 9 \); Australia, \( n = 2 \); international, \( n = 2 \)) published between 2000 and 2017 (Table 1).

The majority are based on the ACCE Framework (whose name derives from the evaluation components used: analytic validity, clinical validity, clinical utility, ethical, legal and social implications) \( (n = 13) \) [7–24], on the Health Technology Assessment (HTA) process \( (n = 6) \) [25–31], or both \( (n = 2) \) [32, 33]. The remaining frameworks refer to the Wilson and Jungner screening criteria \( (n = 3) \) [34–37], or to a mixture of preexisting frameworks, which are not necessarily specific for genetic tests, even if the ACCE framework is often included \( (n = 5) \); Table 1) [38–42].

Seventeen frameworks deal with genetic tests in general [7–19, 22, 26, 27, 32, 36–42]; five refer to genetic susceptibility tests [20, 24, 33–35]; three to pharmacogenetic tests [23, 30, 31]; two to predictive genetic tests (including susceptibility and presymptomatic tests) [25, 28]; one to the new technologies of personalized medicine [29] and another to newborn screening [21] (see Supplementary Information for definitions of types of test) (Table 1). Most of the frameworks pursue a wider aim than simply summarize evidence, for example, support provision and coverage.
Table 1: Frameworks used to evaluate genetic tests retrieved by the systematic review process

<table>
<thead>
<tr>
<th>Framework name</th>
<th>Institution/organization&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Year&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Country&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Reference frameworks</th>
<th>Target test</th>
<th>Purpose</th>
<th>Primary audience</th>
<th>Practical application</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACCE</td>
<td>CDC</td>
<td>2000</td>
<td>USA</td>
<td>ACCE</td>
<td>Genetic tests</td>
<td>Clinical guidance and policy making</td>
<td>Policy makers, health professionals, general public</td>
<td>Website CDC [7, 43]</td>
<td></td>
</tr>
<tr>
<td>NHS UKGTN Gene Dossier</td>
<td>UKGTN</td>
<td>2004</td>
<td>UK</td>
<td>ACCE</td>
<td>Genetic tests</td>
<td>Provision and coverage decisions</td>
<td>Decision makers</td>
<td>Website UKGTN [8–10, 44]</td>
<td></td>
</tr>
<tr>
<td>EGAPP</td>
<td>EGAPP</td>
<td>2005</td>
<td>USA</td>
<td>ACCE</td>
<td>Genetic tests</td>
<td>Clinical guidance and policy making</td>
<td>Clinicians and other key stakeholders</td>
<td>Website EGAPP [11, 12, 45]</td>
<td></td>
</tr>
<tr>
<td>Andalusian</td>
<td>AETSA</td>
<td>2006</td>
<td>Spain</td>
<td>ACCE</td>
<td>Genetic tests</td>
<td>Provision and coverage decisions</td>
<td>Decision makers</td>
<td>Not available [13]</td>
<td></td>
</tr>
<tr>
<td>Rapid ACCE</td>
<td>None</td>
<td>2007</td>
<td>USA</td>
<td>ACCE</td>
<td>Genetic tests</td>
<td>Clinical guidance and policy making</td>
<td>Policy makers, physicians, patients</td>
<td>Genetics in Medicine Journal [14, 46]</td>
<td></td>
</tr>
<tr>
<td>Expanded ACCE</td>
<td>PHGf</td>
<td>2007</td>
<td>UK</td>
<td>ACCE</td>
<td>Genetic tests</td>
<td>Clinical guidance and policy making</td>
<td>Policy makers</td>
<td>Not available [15]</td>
<td></td>
</tr>
<tr>
<td>GFH Indication Criteria</td>
<td>GFH</td>
<td>2008</td>
<td>Germany</td>
<td>ACCE</td>
<td>Genetic tests</td>
<td>Clinical guidance</td>
<td>Referrers, service providers, payers</td>
<td>GFH website (German)/ESHG website (English) [16,47]</td>
<td></td>
</tr>
<tr>
<td>Clinical Utility Gene Card</td>
<td>EuroGentest</td>
<td>2010</td>
<td>Europe</td>
<td>ACCE</td>
<td>Genetic tests</td>
<td>Clinical guidance</td>
<td>Clinicians, geneticists, referrers, service providers, payers</td>
<td>Website EuroGentest/Website European Journal of Human Genetics [17–19, 48, 49]</td>
<td></td>
</tr>
<tr>
<td>Complex Diseases</td>
<td>None</td>
<td>2010</td>
<td>UK</td>
<td>ACCE</td>
<td>Susceptibility tests</td>
<td>Clinical guidance and policy making</td>
<td>Policy makers, physicians, general public</td>
<td>Not available [20]</td>
<td></td>
</tr>
<tr>
<td>ACHDNC</td>
<td>ACHDNC</td>
<td>2010</td>
<td>USA</td>
<td>ACCE</td>
<td>Newborn screening</td>
<td>Decision on inclusion in the recommended newborn screening panel</td>
<td>Decision makers</td>
<td>Website ACHDNC [21, 50]</td>
<td></td>
</tr>
<tr>
<td>ECRI</td>
<td>ECRI</td>
<td>2011</td>
<td>USA</td>
<td>ACCE</td>
<td>Genetic tests</td>
<td>Provision decisions</td>
<td>Patients, providers, payers, regulators, test developers</td>
<td>AHRQ website [22, 51]</td>
<td></td>
</tr>
<tr>
<td>CAT</td>
<td>None</td>
<td>2015</td>
<td>USA</td>
<td>ACCE</td>
<td>Pharmacogenetic tests</td>
<td>Evidence summary</td>
<td>Decision makers</td>
<td>Not available [23]</td>
<td></td>
</tr>
<tr>
<td>Clinical Utility Card</td>
<td>MSAC</td>
<td>2016</td>
<td>Australia</td>
<td>ACCE</td>
<td>Susceptibility test</td>
<td>Provision and coverage decisions</td>
<td>Decision makers</td>
<td>MSAC website [24, 52]</td>
<td></td>
</tr>
<tr>
<td>PACNPBT</td>
<td>PACNPBT</td>
<td>2001</td>
<td>Canada</td>
<td>HTA</td>
<td>Predictive tests</td>
<td>Provision and coverage decisions</td>
<td>Decision makers</td>
<td>Not available [25]</td>
<td></td>
</tr>
<tr>
<td>AETMIS HTA</td>
<td>AETMIS</td>
<td>2001</td>
<td>Canada</td>
<td>HTA</td>
<td>Genetic tests</td>
<td>Policy making</td>
<td>Policy makers</td>
<td>Website AETMIS [26, 27, 53]</td>
<td></td>
</tr>
<tr>
<td>MC Master</td>
<td>MC Master University</td>
<td>2003</td>
<td>Canada</td>
<td>HTA</td>
<td>Predictive tests</td>
<td>Provision and coverage decisions</td>
<td>Decision makers</td>
<td>Not available [28]</td>
<td></td>
</tr>
<tr>
<td>HTA Personalized Health Care</td>
<td>None</td>
<td>2012</td>
<td>Spain</td>
<td>HTA</td>
<td>Personalized Health Care</td>
<td>Provision and coverage decisions</td>
<td>Decision makers</td>
<td>Not available [29]</td>
<td></td>
</tr>
<tr>
<td>Codependent Technologies</td>
<td>None</td>
<td>2013</td>
<td>Australia</td>
<td>HTA</td>
<td>Pharmacogenetic tests</td>
<td>Provision and coverage decisions</td>
<td>Decision makers</td>
<td>Not available [30]</td>
<td></td>
</tr>
<tr>
<td>HTA Pharmacogenetics</td>
<td>None</td>
<td>2013</td>
<td>UK</td>
<td>HTA</td>
<td>Pharmacogenetic tests</td>
<td>Clinical guidance and policy making</td>
<td>Policy makers, clinicians</td>
<td>Appraisal itself [31]</td>
<td></td>
</tr>
<tr>
<td>Framework name</td>
<td>Institution/organization</td>
<td>Year&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Country&lt;sup&gt;c&lt;/sup&gt;</td>
<td>Reference frameworks</td>
<td>Target test</td>
<td>Purpose</td>
<td>Primary audience</td>
<td>Practical application</td>
<td>Ref.</td>
</tr>
<tr>
<td>--------------------------</td>
<td>--------------------------</td>
<td>------------------</td>
<td>----------------------</td>
<td>----------------------</td>
<td>-----------------------------------------------</td>
<td>---------------------------------</td>
<td>-------------------------------------------------------</td>
<td>-----------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Hayes GTE</td>
<td>Hayes Inc</td>
<td>2008</td>
<td>USA</td>
<td>HTA-ACCE</td>
<td>Genetic tests</td>
<td>Provision decisions</td>
<td>Policy makers, insurers, hospitals</td>
<td>Website Hayes Inc (with fee)</td>
<td>[32, 54]</td>
</tr>
<tr>
<td>HTA Susceptibility Test</td>
<td>Italian Ministry for Education and University</td>
<td>2012</td>
<td>Italy</td>
<td>HTA-ACCE</td>
<td>Susceptibility tests</td>
<td>Clinical guidance and policy making</td>
<td>Policy makers, physicians, citizens, other stakeholders</td>
<td>Appraisal itself</td>
<td>[33]</td>
</tr>
<tr>
<td>Cancer Screening</td>
<td>Crossroads 99 group</td>
<td>2001</td>
<td>Canada</td>
<td>Wilson and Jungner principles</td>
<td>Susceptibility tests in screening programs</td>
<td>Policy making</td>
<td>Policy makers</td>
<td>Not available</td>
<td>[34]</td>
</tr>
<tr>
<td>Chronic Disease Screening</td>
<td>None</td>
<td>2001</td>
<td>USA</td>
<td>Wilson and Jungner principles</td>
<td>Susceptibility tests in screening programs</td>
<td>Policy making</td>
<td>Policy makers</td>
<td>Not available</td>
<td>[35]</td>
</tr>
<tr>
<td>Genetic Screening</td>
<td>None</td>
<td>2011</td>
<td>International</td>
<td>Wilson and Jungner principles</td>
<td>Genetic tests in screening programs</td>
<td>Policy making</td>
<td>Policy makers, other stakeholders</td>
<td>Not available</td>
<td>[36, 37]</td>
</tr>
<tr>
<td>GETT</td>
<td>IFCC</td>
<td>2010</td>
<td>International</td>
<td>Various</td>
<td>Genetic tests</td>
<td>Provision decisions</td>
<td>Policy makers, researchers, clinical laboratory scientists and administrators, health care system managers</td>
<td>Website CanGeneTest&lt;sup&gt;d&lt;/sup&gt;</td>
<td>[38]</td>
</tr>
<tr>
<td>SynFRAME</td>
<td>None</td>
<td>2012</td>
<td>USA</td>
<td>Various</td>
<td>Genetic tests</td>
<td>Clinical guidance and policy making</td>
<td>Policy makers, innovators, health technology appraisers, physicians, patients, regulators</td>
<td>Not available</td>
<td>[39]</td>
</tr>
<tr>
<td>Practical Framework</td>
<td>None</td>
<td>2012</td>
<td>USA</td>
<td>Various</td>
<td>Genetic tests</td>
<td>Clinical guidance and policy making</td>
<td>Policy makers, clinicians, laboratorians, researchers, test developers, patients</td>
<td>Not available</td>
<td>[40]</td>
</tr>
<tr>
<td>Six-part framework</td>
<td>None</td>
<td>2014</td>
<td>USA</td>
<td>Various</td>
<td>Genetic tests</td>
<td>Provision and coverage decisions</td>
<td>Payers</td>
<td>Not available</td>
<td>[41]</td>
</tr>
<tr>
<td>NASEM Framework</td>
<td>NASEM</td>
<td>2017</td>
<td>USA</td>
<td>Various</td>
<td>Genetic tests</td>
<td>Provision decisions</td>
<td>Decision makers</td>
<td>Not available</td>
<td>[42]</td>
</tr>
</tbody>
</table>

<sup>ACCE</sup> Analytic validity, Clinical validity, Clinical utility, Ethical, legal and Social Implications, CDC Centers for Disease Control and Prevention, NHS UKGTN National Health Service UK Genetic Testing Network, EGAPP Evaluation of Genomic Applications in Practice and Prevention, AETSA Agencia de Evaluación de Tecnologías Sanitarias (Andalusian Agency for Health Technology Assessment), PHG/PHG foundation, GFH Deutsche Gesellschaft Für Humangenetik (German Society of Human Genetics), ESHG European Society of Human genetics, ACHDNC Advisory Committee on Heritable Disorders in Newborns and Children, MSAC Medical Service Advisory Committee (Australian Government, Department of Health), AHRQ Agency for Healthcare Research and Quality, PACNPGT Provincial Advisory Committee on New Predictive Genetic Technologies, AETMIS Agence d’évaluation des technologies et des modes d’intervention en santé (Quebec Agency for Health Services Research and Technology Assessment), Hayes GTE Hayes Genetic Test Evaluation, GETT Genetic testing Evidence Tracking Tool, IFCC International Federation of Clinical Chemistry and Laboratory Medicine, NASEM National Academies of Sciences, Engineering, and Medicine

<sup>a</sup>Institution/organization: name of the institution/organization supporting the development of the framework

<sup>b</sup>Year: year of development or year of publication

<sup>c</sup>Country: country of the reference institution/organization or of the first author

<sup>d</sup>The authors of the GETT framework declared that it has been used to evaluate five genetic tests but we were not able to find these appraisals
decisions or guide clinical practice; the intended primary audience is mainly represented by decision/policy makers (Table 1).

In addition to the two frameworks created as appraisal tools for individual genetic tests (HTA Pharmacogenetics and HTA Susceptibility Test) [31, 33], 11 frameworks were used to generate reports that are available on the web (all open access, except Hayes GTE reports) (Table 1) [43–54]. The most productive frameworks are the Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Process (11 evidence reports and 10 recommendations), the GFH Card (46 cards), the Clinical Utility Gene Card (about 136 cards), and the NHS UKGTN Gene Dossier (476 dossiers) [44, 45, 47–49].

### Evaluation components

The most-represented evaluation components in the retrieved frameworks are analytic validity (included in 93% of retrieved frameworks), clinical validity (96%), clinical utility (100%), economic validity (100%), and ethical, legal, and social implications (ELSI) (76%). The analysis of these components is usually introduced by an overview of the disease and the test under study (86%). Evaluation components frequently missing from the evaluation frameworks are organizational aspects (lacking in 48% of retrieved frameworks), delivery models (73%), and the patient/citizen’s point of view (93%) (Table 2).

Analytic validity is the ability of the test to accurately and reliably measure the genotype of interest [55]. It is considered in markedly different ways in 27 retrieved frameworks (Table 2) [7–28, 30–34, 36–42]. It is most frequently addressed in terms of sensitivity and specificity, but assay robustness and quality assurance, including internal and external control programs, are often considered (e.g., ACCE, EGAPP) [7, 11, 12]; some frameworks (e.g., Expanded ACCE, SynFRAME) extend the concept using more criteria [15, 39].

Clinical validity is the ability of the test to accurately and reliably detect or predict a clinical condition [55]. It is considered in 28 retrieved frameworks (Table 2) [7–28, 30–42]. The majority identify clinical validity as test performance and measure it in terms of sensitivity, specificity, and positive and negative predictive value. Other frameworks (e.g., Expanded ACCE, Complex Diseases) preface the evaluation of the performance of the test with explicit evaluation of the scientific validity, that is, the evidence of gene–disease association, which is usually expressed as an odds ratio or relative risk [15, 20].

Clinical utility, in its narrowest sense, compares the risks and benefits of testing and provides evidence of clinical usefulness for the integrated package of care in terms of measurable health outcomes [56]. It is considered in all 29 frameworks retrieved, with a certain heterogeneity (Table 2) [7–42]. Some frameworks (e.g., HTA Personalized Health Care, HTA Pharmacogenetics) embrace the narrow definition of clinical utility and consider only efficacy, effectiveness, and safety [29, 31]. Others extend the concept and

<table>
<thead>
<tr>
<th>Table 2 Evaluation components and methodological aspects considered in the retrieved evaluation frameworks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evaluation components</td>
</tr>
<tr>
<td>Overview disease/test under study</td>
</tr>
<tr>
<td>Analytic validity</td>
</tr>
<tr>
<td>Clinical validity</td>
</tr>
<tr>
<td>Clinical utility</td>
</tr>
<tr>
<td>Ethical, legal, and social implications</td>
</tr>
<tr>
<td>Delivery models</td>
</tr>
<tr>
<td>Organizational aspects</td>
</tr>
<tr>
<td>Economic evaluationa</td>
</tr>
<tr>
<td>Patient/citizen’s point of viewb</td>
</tr>
<tr>
<td>Methodological aspects</td>
</tr>
<tr>
<td>Format</td>
</tr>
<tr>
<td>Key questions</td>
</tr>
<tr>
<td>Card</td>
</tr>
<tr>
<td>Checklist</td>
</tr>
<tr>
<td>Set of principle/methodological guidance</td>
</tr>
<tr>
<td>Evidence collection and evaluation</td>
</tr>
<tr>
<td>Source of evidencec</td>
</tr>
<tr>
<td>Quality of the evidenced</td>
</tr>
<tr>
<td>Evidence gaps/research priorities</td>
</tr>
<tr>
<td>Recommendations</td>
</tr>
<tr>
<td>Note: where it was unclear whether an evaluation component was present, it was considered missing and the methodological aspect unfulfilled</td>
</tr>
</tbody>
</table>

aEconomic evaluation: the evaluation component was considered as present in the evaluation framework even if only a cost-analysis, and not a complete economic evaluation, was mentioned

bPatient/citizen’s point of view: the evaluation component was counted only if a direct reference to patient consultation was made

cSource of evidence: the methodological aspect was considered fulfilled even if the topic of the source of evidence was given only a brief mention without further details

dEvaluation of quality of the evidence: the methodological aspect was considered fulfilled even if the topic of the quality of evidence was only mentioned briefly without details of the criteria used to assess the quality
How is genetic testing evaluated?

include considerations of aspects otherwise evaluated independently from clinical utility, such as organizational aspects, cost-effectiveness analysis, and ELSI (e.g., ACCE, Expanded ACCE, ACHDNC) [7, 15, 21]. The broadening of the perception of benefits reaches its greatest extent in the concept of personal utility, adopted in some frameworks (e.g., Complex Diseases, ACHDNC), that is, the full range of personal effects that the test may have on patients, such as improved understanding of the disease, enabling reproductive choices or risk-reducing behaviors [20, 21, 57].

The economic evaluation of genetic tests involves the comparative analysis of both the costs and consequences of the various tests under study [58]. All 29 frameworks retrieved consider the economic dimension, but not always in great detail (Table 2) [7–42]. Many models address the cost-effectiveness of the test under study only in the most general terms (e.g., ACCE), and only a few involve precise quantitative and qualitative evaluations of cost-effectiveness and cost-utility evidence (e.g., Codependent Technologies, SynFrame) [7, 30, 39]; other frameworks consider only the financial aspects, either in terms of the cost of the intervention or related savings (e.g., NHS UKGTN Gene Dossier, PACNPGT) [8–10, 25].

The ELSI evaluation component is concerned with the moral value that society confers on the proposed interventions, the specific related legal norms and the impact on the social life of the patient and his or her family [55]. They are considered in 22 retrieved frameworks and analyzed independently (e.g., ACCE, Andalusian) or integrated into other components of evaluation, such as clinical utility, as psychosocial outcomes of testing (e.g., EGAPP, AETMIS HTA) (Table 2) [7–15, 20–22, 25–29, 32–40, 42].

We defined a delivery model for the provision of genetic tests as the broad context in which genetic tests are offered to individuals and families with or at risk of genetic disorders. It includes the health care programs (any type of health intervention preceding and following a genetic test), the clinical pathways (the patient flow through different professionals during the testing process), and the level of care (e.g., primary or specialist care level) in which the test is delivered [59]. Although a complete description of the delivery models is lacking in all frameworks retrieved, we ascribe it to eight retrieved frameworks (Table 2): the three screening frameworks, as they include the concept of health care program [34–37], and five other frameworks, which mentioned some of the elements, albeit not in detail (Table 2) [13, 24, 29, 30, 38].

Organizational aspects include the human, material, and economic resources needed to implement the genetic program as well as the consequences of the implementation on the organizations involved and the whole health care system. Although they do not include a thorough feasibility analysis, 15 retrieved frameworks attempt to estimate the resources required to start up and maintain a particular genetic testing service (Table 2) [7–10, 13, 14, 24–30, 32, 34–38].

The perspective of patients provides experiential evidence that can be used in the evaluation process [60]. Only two of the retrieved frameworks evaluate the direct experience of the patient and other affected individuals, for example, by their participation in surveys (Table 2) [15, 33]. One of these (Expanded ACCE) includes the patient perspective as part of the clinical utility component, whereas the other (HTA Susceptibility Test) assigns it its own dedicated section.

Methodological aspects

The most frequently used formats are the key questions format (12 frameworks) [7, 11–15, 20–23, 32, 41, 42], the card format (five frameworks) [8–10, 16–19, 24, 25], and the checklist format (two frameworks) [30, 38]. The other frameworks have a less structured format and resemble general manuals (Table 2) [26–29, 31, 33–37, 39, 40].

With respect to the process of evidence review, 13 frameworks provide some indication, albeit in scant detail, of the sources referred to [7, 11–15, 21, 22, 25, 30, 31, 33, 39, 42]; 12 frameworks provide an evaluation of the quality of evidence, but the criteria adopted are not always stated clearly [11–14, 21, 22, 25, 30, 31, 33, 35 39, 42]; and finally 12 frameworks attempt to deal with evidence gaps through the formulation of research priorities (Table 2) [7, 11, 12, 14, 21, 23, 25, 28, 31, 34, 38, 40, 42].

Only five of the retrieved frameworks provide, or at least suggest, criteria for making recommendations based on the evidence collected [11, 12, 21, 28, 32, 42]. The most frequently used criteria are the magnitude of the net benefit and the level of certainty of the evidence.

Discussion

Our review identified three main approaches to the evaluation of genetic testing: the ACCE model, the HTA process, and the Wilson and Jungner screening criteria. The most popular is the ACCE model, developed in 2000 by the US Centers for Disease Control and Prevention [7]. In 2004, it was further developed to become EGAPP initiative, to make recommendations for clinical and public health practice [11, 12]. The UK Genetic Testing Network and the Andalusian Agency for Health Technology Assessment elaborated the ACCE model to guide the introduction of new genetic tests into their public health system, creating the 2004 NHS UKGTN Gene Dossier and the 2006 Andalusian Framework, respectively [8–10, 13]. In 2007,
the ACCE model was reworked again: an expanded version of ACCE, supported by the PHG Foundation, added health quality measures to the evaluation process, whereas a more streamlined version shortened the systematic review process for emerging genetic tests [14, 15]. In 2010, the ACCE model was applied to specific types of genetic test through the Complex Disease Framework and the ACHDNC Newborn Screening Framework of the Advisory Committee on Heritable Disorders in Newborns and Children [20, 21]. The ACCE model also inspired two related European frameworks, the 2008 GFH Indication Criteria of the German Society of Human Genetics and the 2010 Clinical Utility Gene Card of EuroGentest [16–19], the latter of which in turn inspired the 2017 Australian Clinical Utility Card [24]. In 2011, the ECRI Institute used the EGAPP process to develop a set of analytical frameworks for different testing scenarios and stakeholder perspectives [22]. Finally, the 2015 Companion tests Assessment Tool (CAT) used the ACCE model as a filter mechanism to determine which tests, in which specific areas, required evaluation [23]. Our results show that the ACCE framework is the main conceptual frame for the evaluation of genetic tests and the most used in practice, as it inspired very productive frameworks in terms of evidence reports such as the EGAPP, the Clinical Utility Gene Card, and the NHS UKGTN Gene Dossier. Some attempts have been made to merge the ACCE and the HTA model, for example, the 2009 framework for genetic tests used by the private American company Hayes and the 2012 framework for susceptibility tests financed by the Italian Ministry of Education, Universities and Research [32, 33].

Due to the widespread use of the ACCE framework, the most frequently employed evaluation components are analytic validity, clinical validity, clinical utility, and ELSI. Although these evaluation components clearly address the technical value of a genetic test, less attention is given to the wider context. Thus, although the clinical context is given sufficient consideration in some cases, in particular the NHS UKGTN evaluation process, whose testing criteria define the appropriate clinical situations for use of a given test [60], the broader context for implementation of a genetic test is often disregarded. In fact, even where an economic evaluation is performed, it is usually rather superficial; similarly, the analysis of delivery models and organizational aspects, when presented, is usually not well structured. These context-related evaluation components are more often considered by the HTA-based evaluation frameworks. The direct experience of patients is almost totally neglected. Nevertheless, being patients the direct beneficiaries of a genetic technology, their perspective could help understanding its value [61]. Finally, the criteria for making recommendations on the clinical implementation of tests are rarely explored.

Since decision makers are the main audience of the evaluation process, the lack of attention to the context-related evaluation components (delivery models, economic evaluation, and organizational aspects) and to the recommendation-making process are arguably the main limitations of the retrieved frameworks. The analysis of the context of implementation, peculiar to the HTA, is critical for securing an efficient and equitable allocation of health care resources and services. An EU-funded research project named HIScreenDiag [62], which closed in 2011, aimed to assess genetic tests using the HTA methodology of the European Network for Health Technology Assessment, which includes a detailed analysis of the economic, organizational, and delivery aspects [63]. Moreover, the adoption of an evidence grading system such as Grading of Recommendations Assessment, Development and Evaluation (GRADE), which scores the strength of recommendations after taking into consideration aspects like patient values and use of resources, would help move the evaluation process from evidence to implementation, and would make the process more comprehensive [64]. Finally, because these frameworks were mainly developed to address single-gene testing, we might consider how appropriate they are for tests based on next-generation sequencing (NGS). However, frameworks that have been adapted for NGS, such as the NHS UKGTN Gene Dossier and the Clinical Utility Gene Card, have proved effective [65, 66].

In contrast to our systematic review, the majority of reviews in the literature on evaluation frameworks for genetic tests have a narrative structure. The only three systematic reviews we retrieved are described as methods for the construction of an evaluation framework (ECRI, SynFRAME, Practical Framework), so their methodology is not strictly reported [22, 39, 40].

One limitation of our work might be a failure to retrieve some of the studies published in the gray literature. To maximize the sensitivity of the search, we used broad search terms, but these yielded results with low specificity; however, we corrected this during the selection process. Moreover, the comparison between the retrieved frameworks could have been affected by the fact that not all frameworks clearly defined their evaluation components, especially with respect to delivery models and organizational aspects.

In conclusion, the ACCE model proves to be a base for the technical appraisal of genetic tests. However, this model is not completely satisfying. We suggest the adoption of a broader HTA approach, including the assessment of the context-related evaluation dimensions (delivery models, economic evaluation, and organizational aspects). This approach would maximize population health benefits, facilitate decision-making and address the main challenges.
of the implementation of genetic tests, particularly in universal health care systems, where economic sustainability is a major issue.

Acknowledgements This work is part of the Italian project "Definizione e promozione di programmi per il sostegno all’attuazione del Piano d’Intesa del 13/3/13 recante Linee di indirizzo su La Genomica in Sanità Pubblica (Definition and promotion of programs to support the implementation of the Guidelines on Genomics in Public Health)”, funded by the Italian Ministry of Health. This work is also partially supported by the project "Personalized pREvention of Chronic Diseases consortium (PRECoD)" funded by the European Union’s Horizon 2020 research and innovation programme MSCA-RISE-2014 (Marie Skłodowska-Curie Research and Innovation Staff Exchange), under the grant agreement N°645740.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license. You do not have permission under this license to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

References


34. Goel V. Appraising organised screening programmes for testing for genetic susceptibility to cancer. BMJ. 2001;7295:1174–8.


How is genetic testing evaluated?


