Journal of Clinical Epidemiology 181 (2025) 111729

Journal of Clinical Epidemiology

ORIGINAL RESEARCH

A scoping review of the assessment reports of genetic or genomic tests reveals inconsistent consideration of key dimensions of clinical utility

Angelo Maria Pezzullo[a,*,1], Angelica Valz Gris[a,1], Nicolò Scarsi[a], Diego Maria Tona[a], Martina Porcelli[a], Matteo Di Pumpo[a], Peter Piko[b], Roza Adany[b], Pragathy Kannan[c], Markus Perola[c], Maria Luis Cardoso[d,e], Alexandra Costa[d,f], Astrid M. Vicente[d,e], Anu Reigo[g], Mariliis Vaht[g], Andres Metspalu[g], Mark Kroese[h], Roberta Pastorino[a,i], Stefania Boccia[a,i]

[a]*Section of Hygiene, Department of Life Sciences and Public Health, Università Cattolica del Sacro Cuore, Rome, Italy*
[b]*HUN-REN-UD Public Health Research Group, Department of Public Health and Epidemiology, University of Debrecen, Debrecen, Hungary*
[c]*Finnish Institute for Health and Welfare (THL), Helsinki, Finland*
[d]*Department of Health Promotion and NCDs Prevention, Doctor Ricardo Jorge National Health Institute, Lisbon, Portugal*
[e]*Faculty of Sciences, BioISI—Biosystems and Integrative Sciences Institute, University of Lisbon, Lisbon, Portugal*
[f]*Institute of Social and Political Sciences, University of Lisbon, Lisbon, Portugal*
[g]*Estonian Genome Centre, Institute of Genomics, University of Tartu, Tartu, Estonia*
[h]*PHG Foundation, University of Cambridge, Cambridge, UK*
[i]*Department of Woman and Child Health and Public Health, Fondazione Policlinico Universitario A. Gemelli IRCCS, Rome, Italy*

Accepted 17 February 2025; Published online 20 February 2025

Abstract

Objectives: Genetic and genomic tests are the cornerstone of personalized preventive approaches. Inconsistency in evaluating their clinical utility is often cited as a reason for their limited implementation in clinical practice. Previous reviews have primarily focused on theoretical frameworks used for clinical utility evaluations of genetic tests, rather than actual assessments and examined dimensions, rather than specific indicators within these dimensions. We aimed to review the dimensions and the specific indicators measured in published assessment reports of genetic or genomic tests.

Study Design and Setting: We conducted a scoping review of assessment reports of genetic and genomic tests used for prevention, searching through PubMed, Web of Science, Scopus, the websites of 20 different organizations, Google, and Google Scholar. From the included assessments, we extracted the reported indicators of clinical utility, compiling a list of disease-specific indicators that detailed their numerator, denominator, and calculation methods. We analyzed the extracted indicators by stratifying them according to ten comprehensive dimensions of clinical utility, the assessment framework used, and the type of indicator (categorized as quantitative, qualitative, reference, or no evidence reported). From these indicators, we then distilled a list of general indicators.

Results: We reviewed 3054 unique references and 12,000 results from gray literature searches, ultimately selecting 57 assessment reports. The reference frameworks used were health technology assessment (HTA) (42%), Evaluation of Genomic Applications in Practice and Prevention (EGAPP) (25%), ACCE (21%), and others (12%). We identified 951 disease-specific indicators. The dimensions most frequently evaluated (ie, had at least one indicator) were analytic validity (60%), clinical validity (79%), clinical efficacy (79%), and economic impact (58%). Only 12 assessments compared health outcomes between tested and untested groups, and fewer than 15% of the assessments addressed equity, acceptability, legitimacy, and personal value.

Funding: This research is supported by the ''PROPHET − a PeRsOnalised Prevention roadmap for the future HEalThcare'' project funded by the European Union (HORIZON Research and Innovation Actions no. 101057721). The UK participant in Horizon Europe Project PROPHET is supported by UKRI grant number 10040946 (Foundation for Genomics & Population Health). The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

[1] Equal first author contribution.

* Corresponding author. Section of Hygiene, Department of Life Sciences and Public Health, Università Cattolica del Sacro Cuore, Largo Francesco Vito, 1, Rome 00168, Italy.

E-mail address: angelo.pezzullo@unicatt.it (A.M. Pezzullo).

https://doi.org/10.1016/j.jclinepi.2025.111729

Conclusion: Our study illustrates that, although dimensions such as equity and acceptability, are significantly emphasized in traditional evaluation frameworks, these are often not considered in the assessments. Additionally, our study has underscored a significant dearth of reported primary evidence concerning the clinical efficacy of these tests.

Keywords: Genetic testing; Genomics; Precision medicine; Clinical utility; Health technology assessment; Personalized prevention

Plain Language Summary

Genetic and genomic tests analyze a person's genes to predict health risks and guide healthcare decisions, potentially identifying who might benefit from certain treatments or check-ups. However, determining whether these tests are genuinely useful for wide use in health services is complex, because there is no standard way to define ''clinical utility'' of a genetic test. To understand how these tests are evaluated, we reviewed 57 evaluation reports from high-income countries, most of which focused on cancer-related genetic tests. We found that many evaluations looked mainly at how well a test predicted a condition (validity) and considered some form of effectiveness, yet often failed to measure whether the test truly improved patient health outcomes, such as lowering death rates or enhancing the quality of life. Moreover, factors like patient acceptance, equity, and personal relevance (eg, reducing anxiety) were frequently overlooked. Without including these broader considerations, evaluations risk missing critical evidence that would indicate whether a test is helpful, fair, and worth using. From over 900 unique indicators used to measure clinical utility, we created a simpler list of about 150 general indicators that can guide future evaluations. This consolidated list can help test developers decide which factors to investigate, evaluators determine what to measure, and policymakers identify what might be missing before deciding if a test should be adopted in healthcare. By highlighting the gaps—areas that should be assessed but currently are not—our study encourages a more comprehensive approach to evaluating genetic tests. If we fail to consider issues like equity, patient preferences, and proven health benefits, we risk investing in tests that may do little good or even harm patients. Ultimately, recognizing these shortcomings can lead to better-informed decisions, ensuring that genetic testing is used in ways that truly benefit patients and deliver safer, more personalized, and fairer healthcare for everyone.

1. Introduction

Genetic risk factors can substantially contribute to the loss of healthy life years for individuals and populations [1]. As chronic diseases become a major burden on health systems and communities worldwide, genomic data can potentially help identify patient subgroups to prioritize preventive actions, including tailored screening programs, thus enabling a personalized prevention approach [2]. This approach underscores the potential role of genomics in driving patient-centered, proactive health interventions [3,4].

Several barriers are often highlighted when the slow translation of genomic medicine into health services is discussed, including challenges in establishing the necessary infrastructure for the -omics sciences, difficulties in engaging stakeholders, and a lack of evidence regarding clinical utility [5,6]. In general, the term ''clinical utility'' refers to the usefulness or value of a health-related practice or test. It does not have a singular or universally agreed-upon definition and, given its broad nature, it can be interpreted and applied in various ways [7,8]. Since the value of pursuing a test implementation in practice may be influenced by many factors, including organizational, ethical, legal, and societal issues, a judgment of clinical utility may be viewed as a comprehensive assessment of all these factors.

Previous studies identified several existing assessment frameworks for the evaluation of clinical utility, detailing their methodology and the dimensions evaluated [9,10], examining the characteristics of those adopted by health technology assessment (HTA) bodies [11], and reviewing the indicators used in primary research studies [12]. Meanwhile, a number of formal assessment reports of genetic or genomic tests that actually used such frameworks were published, with heterogeneity in the dimensions included and related indicators [11]. However, none of these previous studies have attempted to comprehensively extract all the details of the indicators used in the assessment reports of genetic or genomic tests. Therefore, our study aims to review the dimensions and the specific indicators measured in the published assessment reports of genetic or genomic tests used for primary, secondary, or tertiary prevention, regardless of their application context.

What is new?

Key Findings

- This scoping review identified 951 specific indicators used in the assessment of genetic and genomic tests, from which 156 general indicators were distilled.
- The study revealed significant variability in assessment methodologies and a lack of consideration of various dimensions of clinical utility, such as equity, acceptability, and clinical efficacy, in many of the assessment reports.

What this adds to what was known?

- Previous studies have reviewed frameworks for evaluating clinical utility but have not comprehensively extracted detailed indicators used in genetic and genomic test assessments. This study fills that gap by showing how a broad range of dimensions, such as context, acceptability, legitimacy, and personal value, emphasized in theoretical frameworks, are practically measured and operationalized.

What is the implication and what should change now?

- The study underscores the need for more comprehensive and evidence-based assessments of genetic and genomic tests, incorporating underrepresented dimensions like equity and acceptability. To achieve this, there is a call for innovative research methodologies and real-world data collection to generate robust evidence on clinical utility. The catalog of indicators provided by this study can inform evaluators' and researchers' efforts across the translational spectrum, ultimately leading to more informed decisions regarding the implementation of personalized preventive interventions.

2. Methods

This scoping review of the literature is reported according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) extension for scoping review guidelines [13]. The research protocol was registered on the Open Science Framework platform (https://osf.io/h3cxn). Subsequent amendments and clarifications are available in Supplementary Table 1.

2.1. Preliminary research

To inform the search of clinical utility indicators, we performed preliminary research to collect the available definitions and dimensions of clinical utility related to genetic and genomic testing. This involved an in-depth analysis of 30 frameworks, evaluating genetic tests from the systematic review published by Pitini et al and its update [9,14]. Detailed methodology is in the Supplementary Materials.

Preliminary research identified the following ten unique dimensions of clinical utility used in genetic test assessment frameworks: (i) acceptability, (ii) analytical validity, (iii) clinical validity, (iv) context, (v) economic impact, (vi) equity, (vii) feasibility, (viii) clinical efficacy, (ix) legitimacy, and (x) personal value (Table 1). Detailed results of the preliminary research are available as Supplementary Materials (Supplementary Tables 2−4). For two dimensions, we established subcategories for a more accurate evaluation:

Clinical efficacy:

- Decision impact
- Direct impact
- Indirect impact

Context:

- Disease
- Treatment
- Guidelines.

Definitions for these subcategories are also provided within Table 1.

2.2. Search strategy

We searched PubMed, Web of Science, and Scopus databases till 1 March 2023 to identify studies assessing the clinical utility of genetic or genomic tests. The complete search string is in Supplementary Table 5. We also used Google and Google Scholar with ten specific search strings, analyzing the first 100 results for each and applying them to 22 institutional websites (eg, HTA agencies, international organizations) (Supplementary Tables 6 and 7). Additionally, we conducted forward citation screening of assessment frameworks from the preliminary research and backward citation screening of literature reviews identified during selection.

2.3. Identification and selection of relevant studies

We included published articles, reports, and documents describing formal assessments of any genetic or genomic test and/or personalized preventive program based on genetic or genomic test information. Assessments lacking an evaluation of clinical utility or at least one of its dimensions, as defined in the preliminary research, were excluded. Only English documents were included. After training on a random sample of 50 references, 10 reviewers screened the scientific and gray literature. Each reference was independently screened by two reviewers by title, abstract, and full text, with discrepancies resolved by discussion.

Table 1. Dimension and subdimension of clinical utility

Dimension	Definition	Subdimension	Subdimension definition
Acceptability	The test's conformity to the wishes, desires, and expectations of patients and their families.		
Analytic validity	How accurately and reliably the test measures the genotype of interest.		
Clinical validity	The ability to detect or predict the associated disorder.		
Context	Description of the test use, including the genetic variability, target condition, availability of treatment options, and recommendations	Disease	Description of the target condition
		Treatment	Availability and efficacy (regardless of the genetic status) of treatment options
		Guidelines	Presence of recommendations and guidelines
Economic impact	Assessment of the cost and economic benefits of genetic testing		
Equity	The test's conformity to the principle of just and fair distribution of health		
Feasibility	Sustainability of the intervention and the capability to address potential barriers to using it		
Clinical efficacy	A measurable change in symptoms, overall health, ability to function, quality of life, or survival outcomes of patients.	Decision-making	Impact of the test on clinical decision-making
		Indirect	Changes in health outcomes that are not directly confirmed by comparing the health outcomes of similar patients who have been tested and those who have not. Instead, they are inferred from methods like modeling, retrospective analysis, and other reasonable approaches
		Direct	Changes in health outcomes that are directly confirmed by comparing the health outcomes of similar patients who have been tested and those who have not
Legitimacy	The conformity of a test to social preferences expressed in ethical, principles, values, norms, mores, laws, and regulations.		
Personal value	The indirect health-related and other nonmedical benefits to the individual of having the information.		

A formal assessment was defined as a systematic and standardized process using health impact assessment (HIA), HTA, Grading of Recommendations Assessment, Development and Evaluation (GRADE) Evidence to Decision framework (EtD), or any other framework developed to evaluate genetic or genomic tests (eg, ACCE model, and the Evaluation of Genomic Applications in Practice and Prevention (EGAPP)). The discontinued Clinical Utility Gene Cards [15,16] and the similar National Center for Biotechnology Information GeneReviews were excluded, as their focus is on guiding individual clinician decision-making [16,17]. Omics tests encompass genomics, transcriptomics, proteomics, and metabolomics. We included assessments of tests whose provided information can be applied to primary, secondary, or tertiary prevention settings. Documents without information on assessment indicators were excluded.

2.4. Data extraction

Data were extracted by six reviewers using a predesigned form and revised by a single reviewer. The first sheet included the following document details: reference, first author, organization, journal, year of publication, geographical scope, country, prevention level, disease addressed,

disease categories, evaluated test, gene or variants, assessment methodology and study quality assessment, and evidence level and recommendation strength evaluation. The second sheet reported clinical utility indicators, defined as any quantitative or qualitative measure used to evaluate the test. Four types of indicators were extracted as follows: (i) quantitatively measured, (ii) qualitatively measured, (iii) quantitative without reported measurement but only referenced, and (iv) indicators considered but not measured due to insufficient data. For example, ''clinical sensitivity'' is a quantitative indicator of clinical validity, expressed as the proportion of positive tests in patients with the disease or ''potential stigmatization of carriers,'' is a qualitative indicator of legitimacy. Indicators were identified through preliminary reading of recommendations to determine the dimension assessed, followed by an extensive search of the report to collect the indicators used to assess each dimension. For each indicator, we extracted the name as reported in the assessment, the type of indicator, calculation methods, numerator, denominator, associated diseases, and linked genetic test or variant.

2.5. *Syntheses of results*

The main characteristics of included documents were tabulated, frequencies were calculated, and graphs were created with R 4.2.2. We listed all identified indicators, categorizing them based on assessment methodology and dimension, as defined in the preliminary research. To evaluate dimensions assessed by reports, we calculated the proportion of assessments with at least one indicator in each dimension, stratified by methodology. Additionally, we calculated frequencies of assessments with at least one indicator in each of the clinical efficacy subdimensions and counted those relying on the following: (i) direct quantitative evidence; (ii) indirect quantitative evidence or modeling; (iii) quantitative evidence on clinical decision-making impact; and (iv) no quantitative evidence. Then, we compared reports containing a direct quantitative indicator of clinical efficacy with those that did not. Finally, to explore the types of indicators used for assessing each dimension, we calculated the proportion of indicators within each dimension, stratified by type.

After analyzing extracted indicators, duplicates were removed to generate a list of general indicators. An indicator was considered a duplicate if it shared the same numerator and denominator and the calculation method was either identical or conceptually similar (eg, risk difference, hazard ratio, risk ratio, odds ratio) to another indicator. The list of general indicators was tabulated, detailing the dimension, subdimension, name, calculation method, numerator, denominator, and the assessment methodologies in which each was used. We calculated the percentage of documents using each general indicator, stratified by dimension and assessment methodology.

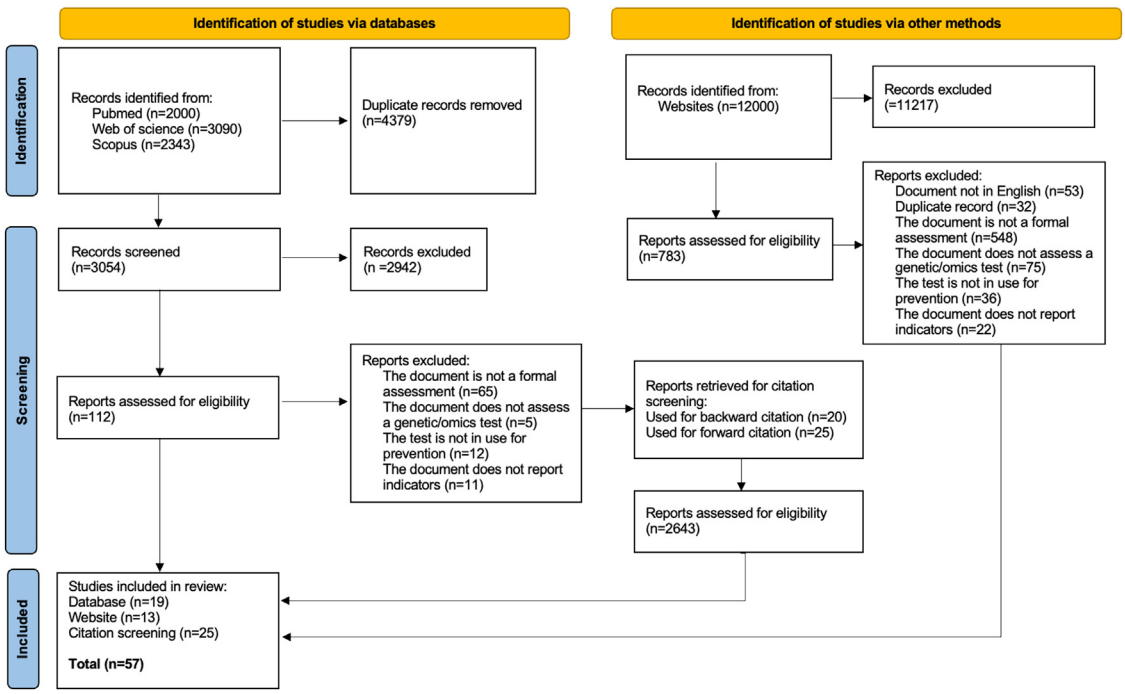

Figure 1. Flowchart of the search and selection process.

Table 2. Characteristics of the included assessments

Characteristics	Number of documents (%) (n = 57)
Geographical scope	
Local	4 (7%)
National	35 (61%)
Regional	4 (7%)
Global	14 (25%)
Disease categories	
Breast cancer	18 (32%)
Colorectal cancer	6 (11%)
Other cancers	7 (12%)
Hereditary conditions	7 (12%)
Cardiovascular diseases	10 (18%)
Other	9 (16%)
Test categories	
Germline mutation	28 (49%)
Monogenic or multigenic somatic mutation	26 (46%)
Polygenic risk score	2 (3%)
mRNA gene expression levels	1 (2%)
Assessment methodology	
ACCE Model	12 (21%)
EGAPP	14 (25%)
GRADE EtD	4 (7%)
HTA	24 (42%)
Other	3 (5%)
Quality assessment	
Yes	13 (23%)
Level of evidence	
Yes	31 (54%)
Strength of recommendation	
Yes	20 (35%)

EGAPP Initiative, Evaluation of Genomic Applications in Practice and Prevention Initiative; EtD, Evidence to Decision; HTA, health technology assessment of medical devices.

3. Results

3.1. Characteristics of the included assessments

Our search of assessment reports yielded a total of 2694 unique records retrieved from scientific databases, from which we selected 19 documents. 13 additional documents were retrieved through the gray literature, and 25 were identified through the citation screening, resulting in a set of 57 included reports (Fig 1). All documents were developed by entities located in high-income countries, with a nearly equal distribution between North American and European regions. Tests used in cancer prevention were most prevalent (31, 54%), followed by cardiovascular health (10, 18%) and hereditary conditions (7, 12%). HTA was the most common assessment method (24, 42%), followed by EGAPP (14, 25%) and ACCE reports (12, 21%). Tests targeting the identification of germline mutations were the most prevalent (28, 49%), followed by test targeting monogenic or multigenic disease-related somatic mutations (26, 46%) (Table 2). Detailed characteristics of each assessment are reported in Supplementary Table 8.

3.2. Clinical utility indicators

We identified a total of 951 indicators, with a median of 14 indicators per report (ranging from 4 to 56). Out of the 57 reports reviewed, 60% contained at least one indicator that related to analytical validity, 79% to clinical validity, 79% to clinical efficacy, 12% to acceptability, 58% to context, 14% to equity, 16% to feasibility, 12% to legitimacy, 14% to personal utility, and 58% to economic impact. Figure 2 illustrates the distribution of indicators across each dimension, stratified by assessment methodology used. Figure 3 displays how indicators found within each dimension distribute in terms of different types of indicators (quantitative, qualitative, reference, or no evidence reported).

Out of the 45 assessments that included clinical efficacy indicators, 12 documents (27%) reported primary direct quantitative evidence of clinical efficacy, 12 (27%) relied on primary quantitative indirect evidence or modeling, 4 (9%) reported only primary quantitative evidence on the impact on clinical decision-making, and 17 (38%) did not report any primary quantitative evidence. Assessments with the consideration of primary direct evidence of clinical efficacy were more recent, more often conducted at the national level, used HTA as a reference framework, and focused on germline mutations (Supplementary Table 9).

From the 951 extracted indicators, we distilled 156 general indicators. The absolute frequencies of the indicators stratified by dimension and assessment methodology are presented in Table 3. A list of the indicators grouped by dimension and ranked by relative frequency is presented in Table 4 (first tertile for each dimension), Supplementary Table 10 (complete list) and 11 (stratified by assessment method). For example, within the clinical validity dimension, frequently reported indicators were clinical sensitivity and specificity of the test (ie, the ability of the test to correctly classify patients with and without the disease).

4. Discussion

In a scoping review of 57 reports, we observed context-dependent variability in assessment approaches, reflecting diverse methodologies tailored to specific circumstances. HTA was the most frequently employed assessment type and seems to be used increasingly. We identified 951 specific indicators and distilled a list of 156 general indicators to aid future evaluations of similar technologies. Most assessments focused on cancer tests and tertiary prevention, highlighting a potential gap in evaluating genetic or genomic tests for other diseases and for primary or

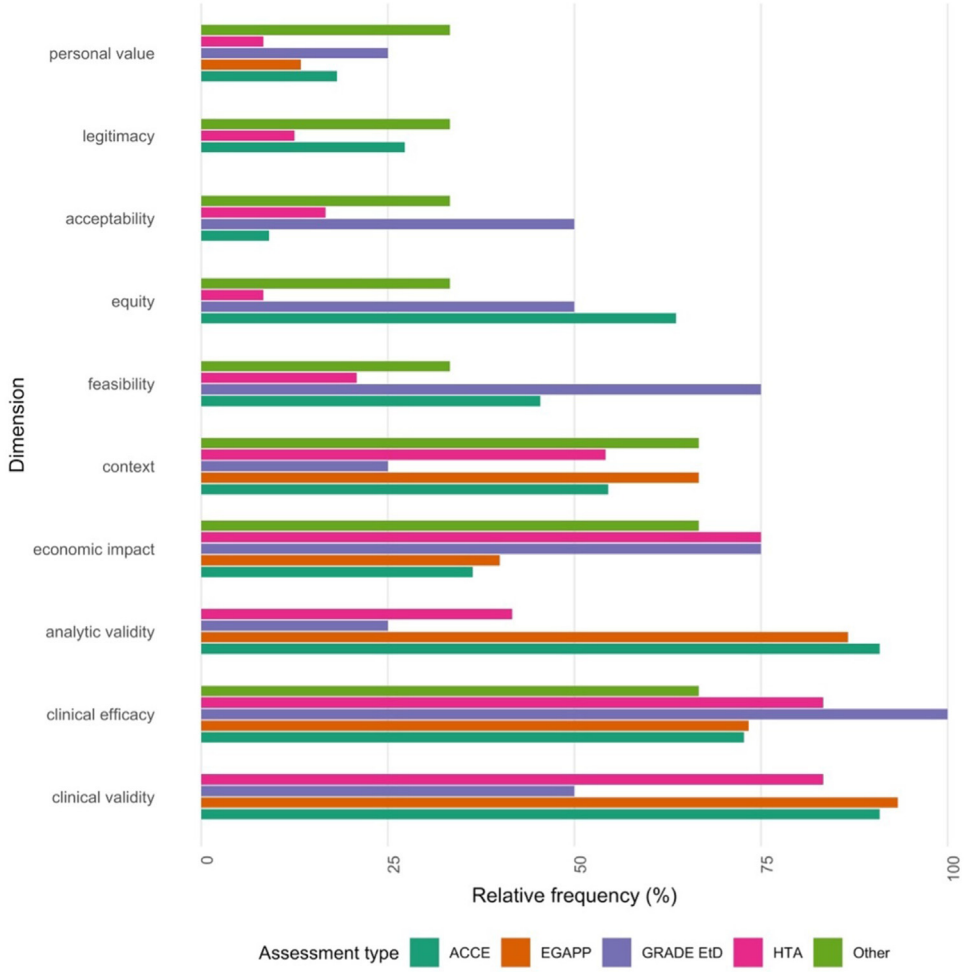

Figure 2. Proportions of the assessments that had at least one indicator in each of the dimensions by assessment methodology.

secondary prevention. Additionally, there was limited primary direct evidence of clinical efficacy, underscoring the challenges of collecting robust data for these assessments.

Our findings indicate that many genetic and genomic tests are evaluated for clinical use based on insufficient evaluation of various dimensions, notably including clinical efficacy. Most evidence of clinical efficacy relates to how they influence decision-making or stems from indirect evidence [12,18,19]. As seen in other areas of clinical medicine, the anticipated benefits of new technologies may not materialize as expected, and unintended consequences can arise [20]. Reliance on surrogate outcomes, such as changes in clinical decision-making rather than direct patient health outcomes, may overestimate the true clinical utility of these tests [21]. Moreover, assessments underrepresenting equity, acceptability, legitimacy, and personal utility can lead to the implementation of tests that are not aligned with patient values or that exacerbate existing health disparities [22–25] and hinder the equitable and effective integration of genomic medicine into healthcare systems. Generating evidence on clinical efficacy is particularly challenging for genetic and genomic tests due to the low prevalence of certain genetic conditions, rapid technological advancements, extensive variety of gene panel/intervention combinations, and issues related to data sharing [26–28]. In primary and secondary prevention, this becomes even more intricate and resource-intensive, given the long interval between interventions and expected outcomes. Innovative trial designs like adaptive trials can expedite evidence generation but may not overcome the inherent time lag in preventive interventions. Real-world data and modeling can supplement evidence but also have limitations, including potential biases and assumptions affecting validity. An ongoing debate persists regarding the level of evidence required for the clinical implementation of personalized preventive technologies [29–35], as it is not always feasible to require direct empirical evidence of clinical outcomes. In specific scenarios, a more refined approach may be warranted, where reliance on surrogate outcomes or proof-of-principle studies could be appropriate. Such situations call for careful, case-by-case judgment, as determining if these alternatives are sufficiently robust is

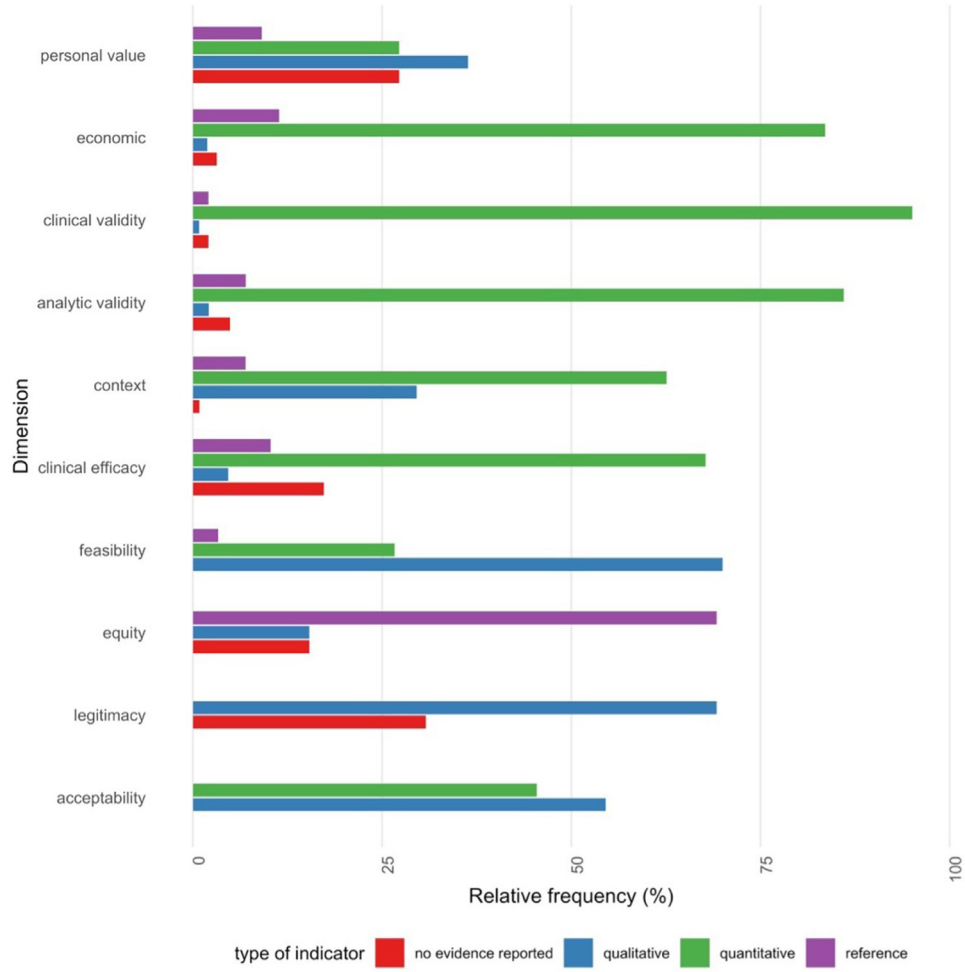

Figure 3. Distribution of indicator types within each dimension*. * The sum of the four proportions (no evidence reported, qualitative, quantitative, and reference) for each dimension equals 100%. Reference: quantitative indicators lacking a reported measurement in the assessment but with a bibliographic reference provided. No evidence reported: indicators that were reported by the authors but were not measured due to insufficient evidence in the literature.

inherently complex. For this reason, it is crucial to maintain rigorous and transparent evaluation processes, ensuring that assumptions made in the absence of direct empirical evidence are clearly articulated to provide accountable guidance for technology adoption [36].

In the early 2000s, with the development of the ACCE model, the evaluation of genetic tests used specific frameworks tied to classic dimensions like analytic validity, clinical validity, clinical utility, and ethical, legal, and social implications (ELSI). However, from 2010 onward, the evaluation of genomic tests has become more standardized, aligning with common assessment methodologies like HTA, following the European Union's efforts to establish it as the primary methodology for assessing healthcare technologies [37]. Our study highlights significant challenges in adapting HTA to comprehensively address the complex dimensions of genetic and genomic test assessments. While HTA assesses key ACCE dimensions, such as clinical validity and utility, it tends to overlook analytic validity, assuming that the technical characteristics of the test are preassessed. Moreover, it prioritizes economic evaluations and modeling analyses, while essential aspects like equity, acceptability, and personal value of information often receive insufficient attention. This aligns with previous reviews emphasizing that, although HTA adoption has led to a more thorough consideration of contextual factors and economic implications, it tends to overlook vital issues like privacy, informed consent, data sharing, equitable access, and other aspects traditionally categorized under the ELSI domain in ACCE and EGAPP methodologies [11,14,38]. Our search also aimed to identify formal assessments of policies and programs but found none, similar to previous findings [39]. Formal prospective evaluations, focusing on the impact of personalized prevention policies or programs in specific geographical contexts rather than solely on test efficacy, could provide decision-makers with

Table 3. Number of indicators for each dimension in total from all assessments reviewed by the framework used

	Reported indicators[a] (n = 951)					
Dimension	**ACCE**	**EGAPP**	**GRADE EtD**	**HTA**	**Other**	**Total**
Acceptability	1	-	2	6	2	11
Analytical validity	51	47	8	37	-	143
Clinical validity	44	69	5	125	-	243
Context	17	30	6	54	8	115
Economic impac	18	12	5	112	11	158
Equity	8	-	2	2	1	13
Feasibility	7	-	5	11	7	30
Clinical efficacy	22	53	23	114	2	214
Legitimacy	6	-	-	6	1	13
Personal value	2	2	2	4	1	11
Total	176	216	58	471	33	951
	General indicators[b] (n = 156)					
Dimension	**ACCE**	**EGAPP**	**GRADE EtD**	**HTA**	**Other**	**Total[c]**
Acceptability	1	-	1	2	2	3
Analytical validity	16	10	4	16	-	27
Clinical validity	11	17	2	13	-	21
Context	12	14	4	17	6	27
Economic impact	12	6	3	16	5	17
Equity	5	-	1	2	1	6
Feasibility	3	-	4	9	4	10
Clinical efficacy	13	14	9	27	2	34
Legitimacy	6	-	-	5	1	7
Personal value	1	2	1	2	1	4
Total	80	63	29	109	22	156

EGAPP Initiative, Evaluation of Genomic Applications in Practice and Prevention Initiative; EtD, Evidence to Decision; HTA, health technology assessment of medical devices.

[a] Reported indicators refer to the specific indicators extracted from the assessments, including duplicates and context-specific variations.

[b] General indicators are unique indicators obtained after consolidating similar indicators and removing duplicates.

[c] Row totals do not sum up because general indicators can be categorized into multiple framework bins simultaneously.

crucial insights into various implementation aspects, including organizational, economic, and ethical considerations.

Our database includes 951 disease- and test-specific indicators and 156 general indicators derived from them, which can guide future assessments. It contributes to the existing literature, which focused more on general guidance on questions to be answered and dimensions to be considered when evaluating genetic tests [9–11]. Comparable work identified outcomes used to operationalize and measure the concept of clinical utility of genetic testing, but its focus was on primary studies rather than assessment reports [12]. The authors employed the Fryback and Thornbury hierarchical model of efficacy (FT model) as a conceptual framework to select articles and cluster indicators. Indicators in the dimensions of context, acceptability, legitimacy, equity, and personal value are less considered in Walcott's review compared to ours, probably due to the use of the FT model and the scarcity of primary evidence for these dimensions.

This review has limitations. We did not systematically investigate whether certain dimensions were excluded from assessments, because they were deemed irrelevant to the specific comparison being made. For instance, if a report compares two tests that produce similar organizational impacts, feasibility might be considered noninfluential and thus omitted. Likewise, if both tests involve the same germline mutation but differ only in their somatic variants, their ethical profiles may be effectively identical, making that domain unnecessary for the decision. While we cannot exclude that some assessments may have intentionally omitted dimensions for these kinds of reasons, these instances were not commonly reported in the documents we analyzed. Our analysis did not investigate the reasons

Table 4. General indicators by dimension (only the first tertile of the most used indicators for each dimension is displayed, for the complete list of indicators see Supplementary Tables 10 and 11)

Dimension (number of documents that assessed the dimension[a])	Name of the indicator (percentage of documents that used this indicator among all the documents that assessed the dimension[b])
Acceptability (n = 8)	Personal attitude toward genomic testing (38%); proportion of adequately informed patients who refuse to undergo genetic testing (38%); patient satisfaction with the information provided by the genetic test (38%).
Analytical validity (n = 34)	Analytical sensitivity of the genetic test (59%); analytical specificity of the genetic test (59%); reproducibility of test results (56%); robustness of test results (38%); failure rate of the test (35%); overall accuracy of the test in detecting the variant (26%); range of specimens tested (15%); overall error rate of the test (15%); limit of blank of the assay (9%).
Clinical validity (n = 46)	Clinical sensitivity of the test (39%); clinical specificity of the test (39%); association between genetic risk score and disease-specific health outcome (33%); negative predictive value of the test (22%); positive predictive value of the test (22%): association between positive test results and disease-specific health outcome (17%) association between genetic risk score and overall mortality or survival rate (15%).
Context (n = 32)	*Disease:* cumulative incidence of the disease (31%); prevalence of the disease (25%); mortality rate of the disease (25%); risk of a disease in patients with a genetic disorder (13%); prevalence of a mutation in population subgroups (13%). *Treatment:* the existence of preventive treatment (25%); potentially avoidable adverse drug reaction due to treatment (19%); the proportion of patients undergoing a treatment that could be personalized through the use of a test (13%); the proportion of patients that may be a candidate for a precision treatment (13%).
Economic impact (n = 33)	Incremental cost-effectiveness ratio of the test compared with standard practice (58%); mean incremental cost comparing personalized approach with standard practice (36%); estimate quality of life difference of personalized approach compared with standard of care (30%); total personalized prevention program costs for the healthcare system (27%); overall cost (direct and indirect) related to a specific disease (21%): direct costs of providing the test or the screening (15%); per-patient savings to hospital (15%).
Equity (n = 12)	Difference in analytical performance of the test by race or ethnicity (33%); ability of a genetic test in leading to a decrease in existing health disparities (33%).
Feasibility (n = 14)	Requirements in terms of qualification and quality assurance processes needed for the use or maintenance of the technology (50%); availability of tests in a national laboratory (29%); number of laboratories needed to manage the screening volume (21%); resources required to implement the test in clinical practice (21%).
Clinical efficacy (n = 45)	*Decision-making impact* of the test on treatment recommendation (42%); the proportion of patients following the treatment suggested by test results (16%); uptake rate of preventive intervention in positive patients (13%).
	Indirect effect difference in disease-specific health outcomes between patients whose treatment has been modified based on the information provided by the test and patients treated as usual (36%); comparison of risk of a specific health outcome in treated patients between genetic score risk groups (20%); comparison of estimated mortality in patients tested and in patients not tested (9%); difference in risk of specific health outcome between patients treated with a therapy and patients not treated among patients positive to a test (9%).
	Direct effect adverse events due to the personalized preventive approach (13%); the ability of a test to reduce the rate of adverse events of treatment (13%); improvement in Quality Adjusted Life Years due to the use of the test (9%); the psychological impact of receiving a diagnosis of a newly described condition for which there is limited prognostic data (9%).
Legitimacy (n = 7)	Presence and effectiveness of safeguards in place (43%); potential stigmatization of carrier individuals and couples (29%); impact of the test on ethical implications (29%); laboratory faced risks associated with the failure to perform screening tests according to current standards and guidelines (29%); and quality of the informed consent decision (29%).
Personal value (n = 8)	Likelihood that genomic data will be of relevance to family members (50%).

[a] The number of documents refers to all assessments that included at least one indicator within the specified dimension.
[b] The percentages in parentheses represent the proportion of documents with the indicator out of the total number of documents that assessed the dimension.

why certain assessments did not include some of the dimensions we identified. Understanding the rationale behind the inclusion or exclusion of specific domains would require a more in-depth analysis of each assessment's methodology, which was beyond the scope of our study. There is potential for publication bias, given the reasonable assumption that a

significant number of evaluations might not be publicly available. Furthermore, our review included only evaluations written in English, possibly overlooking national reports published exclusively in their original language. There was also a degree of subjectivity involved in both the identification of indicators used for the assessment, which was often not clearly defined in the documents and in the merging of reported indicators into general ones. To mitigate these limitations, we conducted a comprehensive search that spanned three scientific databases, 20 institutional websites, and extensive searches on Google and Google Scholar. Each stage of the review was conducted by at least two researchers independently except for data extraction, which underwent cross-checks by a second author to enhance accuracy. We documented every stage of our review process and made the research protocol, its subsequent amendments, and the complete database publicly available to ensure transparency and accessibility, in compliance with open science practices. Lastly, our aim was not to recommend the use of specific indicators among the 156 distilled ones, as this should be better addressed in the context of an expert panel review.

5. Conclusion

Our study has illustrated that, although dimensions, such as equity, acceptability, and legitimacy, are significantly emphasized in traditional evaluation frameworks, these are often not considered in assessments. Additionally, our study has underscored a significant dearth of reported primary evidence concerning various dimensions of clinical utility of these tests, including clinical efficacy. Addressing these gaps is essential to prevent potential harm, avoid unnecessary healthcare costs, and ensure that the benefits of genomic medicine are realized equitably across populations. The catalog of indicators presented in this study stands as a valuable resource, poised to guide researchers, technology developers, and decision-makers in steering research efforts toward generating the necessary evidence for the implementation of these tests with the potential to enable a more coordinated and effective personalized prevention.

Data statement

The protocol, data, and code used for this analysis are available (or will be made upon publication) at the Open Science Framework: https://osf.io/h3cxn.

CRediT authorship contribution statement

Angelo Maria Pezzullo: Writing − review & editing, Writing − original draft, Project administration, Investigation, Formal analysis, Data curation. **Angelica Valz Gris:** Writing − review & editing, Writing − original draft, Project administration, Investigation, Formal analysis, Data curation, Conceptualization. **Nicolò Scarsi:** Writing − review & editing, Investigation, Conceptualization. **Diego Maria Tona:** Writing − review & editing, Investigation, Conceptualization. **Martina Porcelli:** Writing − review & editing, Investigation, Conceptualization. **Matteo Di Pumpo:** Writing − review & editing, Investigation, Conceptualization. **Peter Piko:** Writing − review & editing, Investigation, Conceptualization. **Roza Adany:** Writing − review & editing, Funding acquisition, Conceptualization. **Pragathy Kannan:** Writing − review & editing, Conceptualization. **Markus Perola:** Writing − review & editing, Funding acquisition, Conceptualization. **Maria Luis Cardoso:** Writing − review & editing, Investigation, Conceptualization. **Alexandra Costa:** Writing − review & editing, Investigation, Conceptualization. **Astrid M. Vicente:** Writing − review & editing, Funding acquisition, Conceptualization. **Anu Reigo:** Writing − review & editing, Conceptualization. **Mariliis Vaht:** Writing − review & editing, Investigation, Conceptualization. **Andres Metspalu:** Writing − review & editing, Funding acquisition, Conceptualization. **Mark Kroese:** Writing − review & editing, Conceptualization. **Roberta Pastorino:** Writing − review & editing, Funding acquisition. **Stefania Boccia:** Writing − review & editing, Writing − original draft, Supervision, Investigation, Funding acquisition, Conceptualization.

Ethical approval

Not applicable.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.jclinepi.2025.111729.

References

[1] Jukarainen S, Kiiskinen T, Kuitunen S, Havulinna AS, Karjalainen J, Cordioli M, et al. Genetic risk factors have a substantial impact on healthy life years. Nat Med 2022;28(9):1893−901.

[2] Kasztura M, Richard A, Bempong NE, Loncar D, Flahault A. Cost-effectiveness of precision medicine: a scoping review. Int J Public Health 2019;64(9):1261−71.

[3] PROPHET − a PeRsOnalised Prevention roadmap for the future HEalThcare. Defining personalised prevention. Available at: https://prophetproject.eu/latest_news/defining-personalised-prevention/. Accessed March 7, 2025.

[4] Concept paper on strategic research and innovation agenda (SRIA) - PROPHET D1.3. 2024. Available at: https://prophetproject.eu/wp-content/uploads/2023/04/PROPHET-Concept-paper.pdf.pagespeed.ce.BARvqEpt2e.pdf. Accessed May 25, 2024.
[5] Burke W, Laberge AM, Press N. Debating clinical utility. Public Health Genom 2010;13(4):215−23.
[6] Hyde C. Will better evidence on clinical utility bring about greater use of (genetic) tests? NPJ Genom Med 2021;6(1):22.
[7] PHG Foundation. Polygenic risk score and clinical utility. Cambridge, England: PHG Foundation; 2021. Available at: https://www.phgfoundation.org/wp-content/uploads/2023/10/polygenic-scores-and-clinical-utility.pdf. Accessed March 7, 2025.
[8] Moorthie S, Hall A, Babb de Villiers C, Janus J, Brigden T, Blackburn L, et al. How can we address the uncertainties regarding the potential clinical utility of polygenic score-based tests? Pers Med 2022;19(3):263−70.
[9] Pitini E, De Vito C, Marzuillo C, D'Andrea E, Rosso A, Federici A, et al. How is genetic testing evaluated? A systematic review of the literature. Eur J Hum Genet 2018;26(5):605−15.
[10] Sanderson S, Zimmern R, Kroese M, Higgins J, Patch C, Emery J. How can the evaluation of genetic tests be enhanced? Lessons learned from the ACCE framework and evaluating genetic tests in the United Kingdom. Genet Med 2005;7(7):495−500.
[11] Hoxhaj I, Govaerts L, Simoens S, Van Dyck W, Huys I, Gutiérrez-Ibarluzea I, et al. A systematic review of the value assessment frameworks used within health technology assessment of omics technologies and their actual adoption from HTA agencies. Int J Environ Res Public Health 2020;17(21):8001.
[12] Walcott SE, Miller FA, Dunsmore K, Lazor T, Feldman BM, Hayeems RZ. Measuring clinical utility in the context of genetic testing: a scoping review. Eur J Hum Genet 2021;29(3):378−86.
[13] Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018;169(7):467−73.
[14] Pitini E, Baccolini V, Migliara G, Isonne C, Sindoni A, Mazzalai E, et al. Time to align: a call for consensus on the assessment of genetic testing. Front Public Health 2021;9:807695.
[15] McNeill A. Fond farewell to clinical utility gene cards. Eur J Hum Genet 2021;29(10):1473−4.
[16] Dierking A, Schmidtke J, Matthijs G, Cassiman JJ. The EuroGentest clinical utility gene cards continued. Eur J Hum Genet 2013;21(1):1.
[17] Adam MP. GeneReviews®. Available at: https://www.ncbi.nlm.nih.gov/books/NBK1116/. Accessed March 7, 2025.
[18] Wallerstedt SM, Nilsson Ek A, Olofsson Bagge R, Kovács A, Strandell A, Linderholm B. Personalised medicine and the decision to withhold chemotherapy in early breast cancer with intermediate risk of recurrence - a systematic review and meta-analysis. Eur J Clin Pharmacol 2020;76(9):1199−211.
[19] Shickh S, Mighton C, Uleryk E, Pechlivanoglou P, Bombard Y. The clinical utility of exome and genome sequencing across clinical indications: a systematic review. Hum Genet 2021;140(10):1403−16.
[20] Prasad V, Vandross A, Toomey C, Cheung M, Rho J, Quinn S, et al. A decade of reversal: an analysis of 146 contradicted medical practices. Mayo Clin Proc 2013;88(8):790−8. https://doi.org/10.1016/j.mayocp.2013.05.012.
[21] Siontis KC, Siontis GC, Contopoulos-Ioannidis DG, Ioannidis JP. Diagnostic tests often fail to lead to changes in patient outcomes. J Clin Epidemiol 2014;67(6):612−21.
[22] Martin AR, Kanai M, Kamatani Y, et al. Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet 2019;51(4):584−91. https://doi.org/10.1038/s41588-019-0379-x.
[23] Burke W. Utility and diversity: challenges for genomic medicine. Annu Rev Genom Hum Genet 2021;22:1−24. https://doi.org/10.1146/annurev-genom-120220-082640.
[24] Halbert CH. Equity in genomic medicine. Annu Rev Genom Hum Genet 2022;23:613−25. https://doi.org/10.1146/annurev-genom-112921-022635.
[25] Khoury MJ, Bowen S, Dotson WD, et al. Health equity in the implementation of genomics and precision medicine: a public health imperative. Genet Med 2022;24(8):1630−9. https://doi.org/10.1016/j.gim.2022.04.009.
[26] Cohen JP, Felix AE. Personalized medicine's bottleneck: diagnostic test evidence and reimbursement. J Personalized Med 2014;4(2):163−75.
[27] Rogowski WH, Grosse SD, John J, Kääriäinen H, Kent A, Kristofferson U, et al. Points to consider in assessing and appraising predictive genetic tests. J Community Genet 2010;1(4):185−94.
[28] Rogowski WH, Grosse SD, Khoury MJ. Challenges of translating genetic tests into clinical and public health practice. Nat Rev Genet 2009;10(7):489−95.
[29] Wouters RHP, van der Graaf R, Rigter T, Bunnik EM, Ploem MC, de Wert G, et al. Towards a responsible transition to learning healthcare systems in precision medicine: ethical points to consider. J Pers Med 2021;11(6):593.
[30] Goddard KAB, Knaus WA, Whitlock E, Lyman GH, Feigelson HS, Schully SD, et al. Building the evidence base for decision making in cancer genomic medicine using comparative effectiveness research. Genet Med 2012;14(7):633−42.
[31] Burke W, Clayton EW, Wolf SM, Berry SA, Evans BJ, Evans JP, et al. Improving recommendations for genomic medicine: building an evolutionary process from clinical practice advisory documents to guidelines. Genet Med 2019;21(11):2431−8.
[32] Khoury MJ. Dealing with the evidence dilemma in genomics and personalized medicine. Clin Pharmacol Ther 2010;87(6):635−8.
[33] Cavallari LH, Pratt VM. Building evidence for clinical use of pharmacogenomics and reimbursement for testing. Clin Lab Med 2022;42(4):533−46.
[34] Luzum JA, Petry N, Taylor AK, Van Driest SL, Dunnenberger HM, Cavallari LH. Moving pharmacogenetics into practice: it's all about the evidence. Clin Pharmacol Ther 2021;110(3):649−61.
[35] Wilson BJ, Miller FA, Rousseau F. Controversy and debate on clinical genomics sequencing-paper 1: genomics is not exceptional: rigorous evaluations are necessary for clinical applications of genomic sequencing. J Clin Epidemiol 2017;92:4−6.
[36] Eddy DM, Hollingworth W, Caro JJ, Tsevat J, McDonald KM, Wong JB. Model transparency and validation: a report of the ISPOR-SMDM modeling good research practices task force-7. Med Decis Mak 2012;32(5):733−43.
[37] Regulation (EU) 2021/2282 of the European Parliament and of the Council of 15 December 2021 on health technology assessment and amending Directive 2011/24/EU. Off J Eur Union 2021;L 458:1−32.
[38] Love-Koh J, Peel A, Rejon-Parrilla JC, Ennis K, Lovett R, Manca A, et al. The future of precision medicine: potential impacts for health technology assessment. Pharmacoeconomics 2018;36(12):1439−51.
[39] Ádám B, Lovas S, Ádány R. Use of genomic information in health impact assessment is yet to come: a systematic review. Int J Environ Res Public Health 2020;17(24):9417.