**Clinical Chemistry** 64:11
1657–1667 (2018)



# Diagnostic Biomarkers: Are We Moving from Discovery to Clinical Application?

Lucy A. Parker,[1,2] Elisa Chilet-Rosell,[1,2] Ildefonso Hernández-Aguado,[1,2] María Pastor-Valero,[1,2] Sonia Gea,[1] and Blanca Lumbreras[1,2*]

**BACKGROUND:** Despite considerable research investment, moving from biomarker discovery to clinical application has presented unique challenges. We aimed to evaluate progress toward clinical application of a sample of molecular- and "omics"-based diagnostic tests over a 10-year period.

**METHODS:** We used Scopus to locate studies, published before the December 31, 2016, citing 107 original-research articles published in 2006 that assessed the diagnostic value of a molecular- or "omics"-based test. We identified diagnostic studies of the same test and disease and determined whether the article represented progress in the validation of the molecular test. We classified the types of progress: (*a*) clinical validation (measuring diagnostic accuracy in a series of patients similar to the population in which the test will be used in practice), (*b*) technical improvement, (*c*) extended diagnostic application (modification of the diagnostic question attended initially by the test), (*d*) economic evaluation, or (*e*) clinical use or implementation.

**RESULTS:** In the 10-year period analyzed, 4257 articles cited the 107 diagnostic studies; 118 (2.8%) were diagnostic studies of the same test, and of these papers, 25 (21.2%) did not constitute progress toward validation of the test for use in clinical practice (potential research waste). Of the 107 molecular- or "omics"-based tests described in 2006, only 28 (26.2%) appeared to have made progress toward clinical application. Only 4 (9.1%) of 44 proteomics-based tests had made progress toward clinical application.

**CONCLUSIONS:** Articles evaluating molecular- or "omics"-based diagnostic tests are numerous in biomedical journals. Few tests have made progress toward clinical application in the 10 years following their discovery.


Identifying biomarkers that allow earlier or more accurate detection of disease is a major endeavor of biomedical research. Despite billions of dollars spent on research to develop new diagnostic tests based in biomarkers, most are still only classified as "promising" *(1, 2)*. The question is particularly relevant for technologies based in "-omics" techniques, in which economic investment and researchers' and patients' expectations have not translated into recognizable benefits in patient care *(3)*. Some may suggest that the failure to translate these technologies into clinically useful tools is due to their recent discovery, but in fact, the technologies have existed for more than 15 years. The distance between bench-top research and clinical research may hinder the path from discovery to implementation. Some primary research fails to answer questions relevant for clinicians and patients, and there are a great number of biomedical discoveries without effective translation in healthcare *(4, 5)*. Substantial levels of knowledge "waste" are reflected in "-omics"-based technologies research, in which previous data have shown an inverse relationship between publications and patenting of biomarkers *(4)*. To avoid (or at least reduce) knowledge waste, some authors have called for the prompt identification of scientific discoveries that have the ability to affect health *(6)* and have emphasized the importance of adding value to existing evidence to prioritize research gaps before starting a new line of research *(7, 8)*.

Analytical and clinical factors *(9)* can make test validation challenging. Sample collection, storage, and handling can influence test results and introduce bias to

[1] Department of Public Health, University Miguel Hernández, Alicante, Spain; [2] CIBER Epidemiology and Public Health (CIBERESP), Madrid, Spain.
* Address correspondence to this author at: Departamento Salud Pública, Historia de la Ciencia y Ginecología, Universidad Miguel Hernández de Elche, Campus de San Juan, Crta. Alicante-Valencia Km 9˙8, Sant Joan d'Alacant, Alicante, 03550, Spain. E-mail blumbreras@umh.es.

Received June 4, 2018; accepted August 7, 2018.
Previously published online at DOI: 10.1373/clinchem.2018.292854

diagnostic studies, yet most discovery research pays little attention to these details *(10)*. Analytical and postanalytical phases (data interpretation dealing with aspects such as small sample size or lack of clinical and sociodemographic patient information) can also influence test implementation *(11)*. Furthermore, diagnostic tests based on "-omics" technologies are especially susceptible to biases linked to the inappropriate initial selection of the biomarkers or an inaccurate strategy of biomarker validation or analysis *(12)*. To minimize error and stimulate progress from biomarker discovery to clinical application, a formal validation strategy based on available evidence is essential *(13)*.

We previously analyzed whether articles on molecular diagnostic tests interpreted the clinical applicability of their results appropriately *(2)*. We showed that most published research involved preclinical phases to assess the diagnostic accuracy of the test, by comparing test results between sick individuals and healthy controls (or those with an alternate diagnosis). We showed that authors' frequently overinterpreted the clinical applicability of their findings. This phenomenon, referred to as "spin," has since been described in other settings, including systematic reviews of molecular diagnostic tests *(14, 15)*. Spin can be misleading to readers and will tend to underestimate the continued research required to ensure translation of the discovery into a clinically useful tool.

Ten years later, and taking the same sample of studies as reference, we aimed to evaluate if there has been progress toward clinical application of these tests. We traced articles that cited the 107 molecular diagnostic studies published in 2006 *(2)* in the 10-year period from their publication and determined which articles described research that constituted progress in the validation of the test for use in clinical practice.

## Materials and Methods

### DATA SOURCES AND STUDY SELECTION

We used the Scopus (Elsevier) database to identify all articles that cited the 107 studies evaluating the diagnostic value of a molecular- or "-omics"-based test published in 2006 *(2)* (the references from the 107 studies are included in Supplemental Material in the Data Supplement that accompanies the online version of this article at http://www.clinchem.org/content/vol64/issue11), from the time of their publication until the December 31, 2016. The original sample of molecular- or "-omics"-based tests was determined through a thorough search of Medline by use of PubMed and included tests based in technologies that provided a comprehensive analysis of cellular-specific constituents, such as RNA, DNA, proteins, and intermediary metabolites, as well as techniques such as in situ hybridization of chromosomes for cytogenetic analysis, identification of pathogenic organisms via analysis of species-specific DNA sequences, and detection of mutations with PCR. Fig. 1 describes the details of the current search and selection process. We excluded reviews, editorials, letters, and case reports electronically before screening. To identify articles that could potentially constitute progress toward clinical application, we screened the titles and abstracts of the articles to identify continued research on the diagnostic value of the same molecular test for the same disease. We considered "-omics" studies to be the same diagnostic test only if they used the same protein peaks or gene, lipid, or metabolome patterns identified in the referenced study from 2006. Studies that used the same technique but repeated the process of biomarker discovery were therefore not the same test. To assess the reliability of the selection process, 2 researchers (EC and one of the other authors) screened all abstracts independently. Overall agreement in selection of articles that could potentially represent progress (diagnostic study of same test and same disease) was 93.6%. Two researchers then independently evaluated the full text of the articles to confirm they were indeed diagnostic studies of the same test for the same disease. Agreement in this final step was 96.1%. A third researcher resolved discrepancies through discussion and arbitration.

### DATA EXTRACTION AND DEFINITIONS

The following data had already been extracted from the 107 diagnostic studies of molecular tests *(2)*: study design, journal categories selected by Thomson Reuters' Web of Science (Journal Citation Reports 2006), the disease studied, the molecular methodology used, authors' statements regarding the clinical applicability of the test, authors' statement regarding the need for further clinical evaluation, and whether or not the authors overinterpreted the results.

For each of the new diagnostic studies on the same test for the same disease, 2 researchers independently extracted the following data: year of publication, journal, study design, and any variations in the molecular test and/or in the study approach (target population, diagnostic question) in comparison with the original report from 2006. We then classified each study in terms of whether or not it described research that constitutes progress toward clinical application of the test. The type of progress was categorized according to definitions specified in Box 1, which are loosely based on recommendations from other research and initiatives to improve clinical translation of omics research *(16)*.

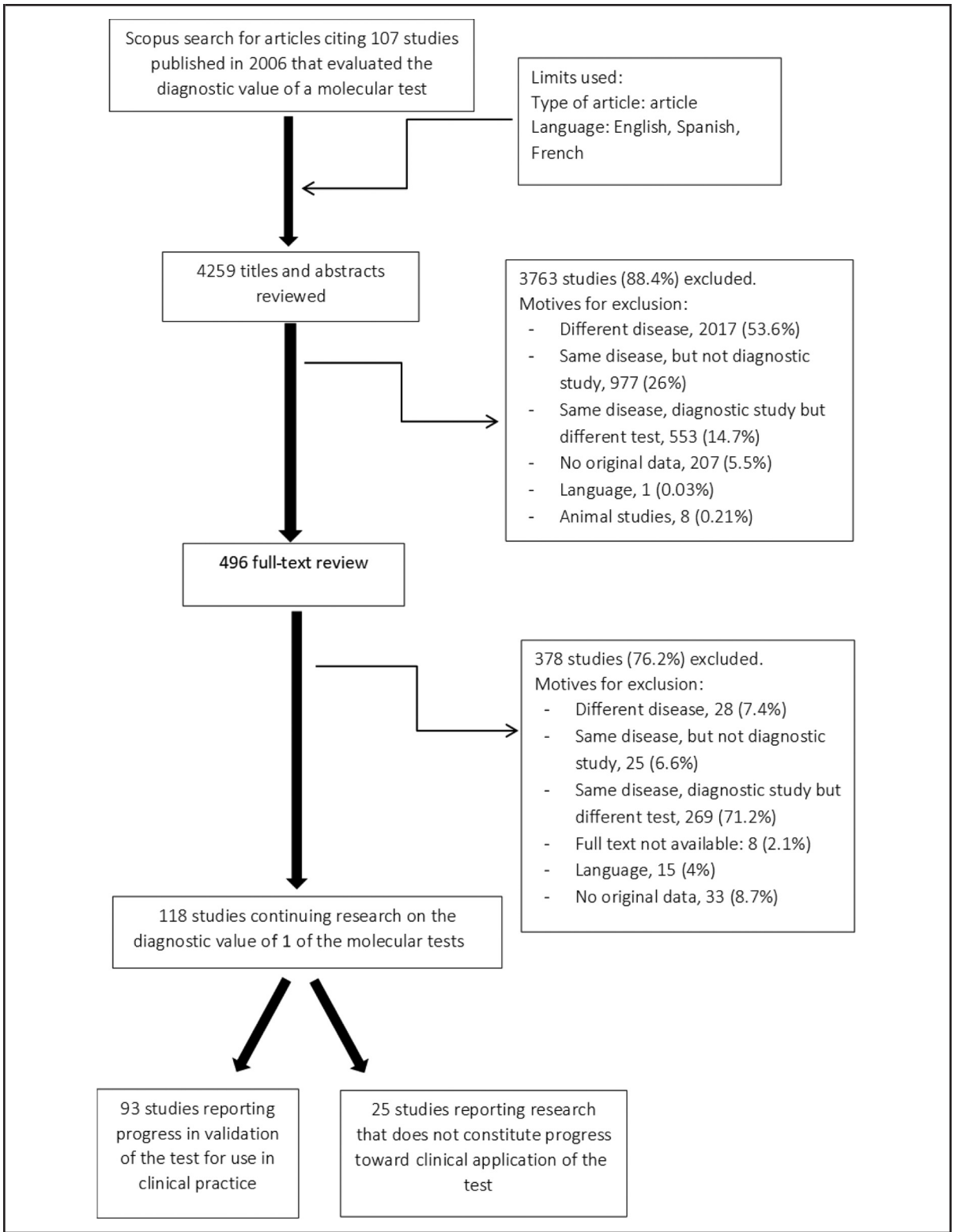


**Fig. 1.** Flow chart detailing search and selection process.

BOX 1.

Types of progress in validation of molecular- or "-omics"-based diagnostic tests for use in clinical practice.

- Advance in the clinical validation: Further study reporting the diagnostic accuracy of the test in an independent patient series comparable to the population on whom the test would be used in practice.
- Technical improvement: Further study reporting modification of the assay or computational procedures to improve diagnostic accuracy.
- Extended diagnostic application: Further study reporting application of the test to a different diagnostic question in the same disease, independently of the study design used.
- Economic evaluation[a]: Further study performed specifically to estimate the cost of using the test in clinical practice.
- Clinical use or implementation[a]: Further study evaluating the effect of using the test in practice or addressing questions relevant to implementation of the test in practice (e.g., resources needed, training, turnover time).

[a] *Only applicable if the clinical validity has previously been established in an independent patient series comparable to the population on whom the test would be used in practice*

The definition of which studies constituted an advance in clinical validation was heavily dependent on the design of the initial diagnostic study from 2006. For example, if the initial study used a healthy control or an alternate-diagnosis control design, it automatically constituted progress when a new study used an independent patient series comparable to the population on whom the test would be used in practice. Studies comparing sick individuals with controls could also be classified as an advance in clinical validation, if the authors correctly justified the patient selection, for example, the evaluation of the likeliness of false-positive or false-negative results in certain patient groups in which we may suspect the diagnostic accuracy to be modified (comorbidities, pregnancy, and such like). If both the initial diagnostic study and the new study used independent patient series comparable to those on whom the test would be used in practice, the new study was only considered to constitute progress if the authors justified why it was important to carry out the new evaluation and indicated the additional information relevant to the clinical application of the test that is provided by the study.

Two authors independently classified each study according to whether or not it constituted progress and the type of progress. Agreement in the classification of progress was 91.5%; agreement in the classification of the type of progress was 91%.

### ANALYSIS

Using medians and interquartile ranges, we summarized the number of citations received by the 107 molecular diagnostic studies published in 2006 in a 10-year period. Statistical comparisons between different subgroups were made with a Kruskal–Wallis test. We considered that the molecular diagnostic tests published in 2006 had made a step toward clinical application if there was at least 1 citation in the subsequent 10 years that we classified as "progress" according to the criteria established in Box 1. Citations from studies assessing the diagnostic value of the 107 tests could represent more than 1 type of progress as relevant. We calculated the proportion of studies that had made progress and assessed the relationship between these proportions and the characteristics of the initial study, using the Pearson $\chi^2$ test, when possible, and the Fisher exact test when the expected count in 1 of the subgroups of analysis was less than 5. We performed the statistical analysis using Stata SE version 12.

## Results

We retrieved 4259 articles that cited the 107 molecular- or "-omics"-based diagnostic studies from 2006 through December 2016. After screening abstracts and examining the full texts, we selected 118 (2.8%) for further analysis (Fig. 1) because they continued to evaluate the diagnostic value of 1 of the initial molecular- or "-omics"-based tests proposed. The total number of citations per study ranged from 3 to 282, with a median of 25 (Table 1). The number of citations received was significantly associated with the impact factor of the journal that published the initial study, with high-impact journals tending to receive more citations (Table 1, $P = 0.004$). There were also statistical differences in the citations received according to the molecular technology used (Table 1). Of the 118 diagnostic studies analyzed in detail, 93 (78.8%) reflected progress toward clinical application of the test, and the remaining 25 (21.2%) did not constitute progress toward validation of the test for use in clinical practice and hence were classified as potential research waste. Although research waste was more frequent in lower impact journals (or journals not indexed in Web of Science Journal Citation Reports), the differences observed were not statistically significant. Similarly, there were no significant differences in the proportion of papers classified as waste over the 10-year period studied (see Table 1 in the online Data Supplement).

Only 33 (30.8%) of the 107 molecular tests published in 2006 were the object of continued diagnostic research in the following 10 years, and 28 (26.2%) of

**Table 1. Citations over a 10-year period of 107 studies on molecular diagnostic tests, according to the characteristics of the initial study.**

| Main characteristics of 107 diagnostic studies of molecular tests published in 2006 | Total citations, 2006–2016 N, % | Median number of citations per study (interquartile range) | *P* value |
|---|---|---|---|
| Study design | | | 0.800 |
| Healthy control or alternative-diagnosis control (*n* = 82) | 3452 | 27 (15–51) | |
| Consecutive series or series of clinically relevant patients (*n* = 14) | 435 | 21.5 (9–43) | |
| Other (*n* = 11) | 372 | 22 (15–40) | |
| Journal category | | | 0.387 |
| Medical (*n* = 35) | 1663 | 28 (21–58) | |
| Oncology (*n* = 32) | 1148 | 26.5 (13.5–46) | |
| Biomedical or general science (*n* = 19) | 591 | 20 (10–40) | |
| Lab and methodology (*n* = 21) | 857 | 20 (16–49) | |
| Ranking of journal citation report | | | 0.004 |
| Q1 (*n* = 61) | 2906 | 58 (26.3–92.3) | |
| Q2 (*n* = 20) | 735 | 39 (26.5–54.8) | |
| Q3 (*n* = 14) | 385 | 15 (11.5–31.5) | |
| Q4 (*n* = 3) | 28 | 14 (11–17) | |
| Non-journal citation report (*n* = 9) | 205 | 20 (10–23) | |
| Disease type | | | 0.834 |
| Autoimmune disease and transplant rejection (*n* = 8) | 255 | 26 (18–46.5) | |
| Cancer (*n* = 61) | 2338 | 25 (16–44) | |
| Congenital disorder (*n* = 9) | 231 | 13 (6–37) | |
| Infectious disease (*n* = 19) | 783 | 30 (12–48) | |
| Neurological (*n* = 6) | 424 | 54.5 (16–138) | |
| Other (*n* = 4) | 228 | 35.5 (17–97) | |
| Type of molecular technology | | | 0.029 |
| *Gene-targeting tests* | *1926* | | |
| PCR based (*n* = 33) | 858 | 20 (9–30) | |
| Microarray (*n* = 20) | 1068 | 34.5 (18.5–57.5) | |
| *Protein-targeting tests* | *2320* | | |
| Mass spectrometry or 2-dimensional gel electrophoresis (*n* = 44) | 1946 | 26 (16.5–54) | |
| Antibody array or protein microarray (*n* = 9) | 374 | 43 (27–58) | |
| *Lipidomics (n = 1)* | 13 | 13 | |
| Authors' conclusion on clinical applicability | | | 0.366 |
| Definitively favorable (*n* = 54) | 2060 | 21.5 (13–48) | |
| Promising (*n* = 49) | 2094 | 27 (19–53) | |
| Unfavorable (*n* = 4) | 105 | 26.5 (23.5–29) | |
| Authors' conclusion regarding need for further validation | | | 0.090 |
| Mention further validation (*n* = 56) | 2488 | 29 (17.5–56.5) | |
| Do not mention further validation (*n* = 51) | 1771 | 23 (13–43) | |
| Overinterpretation | | | 0.356 |
| Yes (*n* = 61) | 2449 | 23 (14–48) | |
| No (*n* = 46) | 1810 | 28.5 (16–51) | |
| Total (*n* = 107) | 4259 | 25 (14–49) | |

these 33 made some progress toward clinical application of the initial test (Table 2). Progress was most common among tests that had initially evaluated diagnostic accuracy in a patient series comparable to the population on whom the test would be used in practice (Table 2). Conversely, <20% of the tests that were

**Table 2. Progress in the validation of 107 molecular diagnostic tests for use in clinical practice over a 10-year period and as detected in published research about the test.**

| Main characteristics of 107 diagnostic studies of molecular tests published in 2006 | Continued research on diagnostic value of test (2006–2016), N, % | *P* value | Progress in validation of the test for use in clinical practice, N, % | *P* value |
|---|---|---|---|---|
| Study design | | 0.001 | | 0.003 |
| Healthy control or alternative-diagnosis control (*n* = 82) | 18 (22.0) | | 15 (18.3) | |
| Consecutive series or series of clinically relevant patients (*n* = 14) | 10 (71.4) | | 8 (57.1) | |
| Other (*n* = 11) | 5 (45.5) | | 5 (45.5) | |
| Journal category | | 0.374 | | 0.705 |
| Medical (*n* = 35) | 10 (28.6) | | 8 (22.9) | |
| Oncology (*n* = 32) | 9 (28.1) | | 8 (25) | |
| Biomedical or general science (*n* = 19) | 9 (47.4) | | 7 (36.8) | |
| Lab and methodology (*n* = 21) | 5 (23.8) | | 5 (4.7) | |
| Ranking of journal citation report | | 0.161 | | 0.286 |
| Q1 (*n* = 61) | 21 (34.4) | | 17 (27.9) | |
| Q2 (*n* = 20) | 8 (40) | | 7 (35) | |
| Q3 (*n* = 14) | 4 (28.6) | | 4 (28.6) | |
| Q4 (*n* = 3) | 0 | | 0 | |
| Non–journal citation report (*n* = 9) | 0 | | 0 | |
| Disease type | | 0.136 | | 0.089 |
| Autoimmune disease and transplant rejection (*n* = 8) | 3 (37.5) | | 1 (23.8) | |
| Cancer (*n* = 61) | 14 (23.00) | | 12 (19.7) | |
| Congenital disorder (*n* = 9) | 4 (44.4)) | | 4 (44.4) | |
| Infectious disease (*n* = 19) | 9 (47.4) | | 8 (42.1) | |
| Neurological (*n* = 6) | 3 (50.0) | | 3 (50) | |
| Other (*n* = 4) | 0 | | 0 | |
| Type of molecular technology | | <0.001 | | <0.001 |
| *Gene-targeting tests* | 24 (45.2) | | 22 (41.5) | |
| PCR based (*n* = 33) | 19 (57.6) | | 17 (51.5) | |
| Microarray (*n* = 20) | 6 (30.0) | | 5 (25.0) | |
| *Protein-targeting tests* | 8 (15.1) | | 6 (11.3) | |
| Mass spectrometry or 2-dimensional gel electrophoresis (*n* = 44) | 5 (11.4) | | 4 (9.1) | |
| Antibody array or protein microarray (*n* = 9) | 2 (22.2) | | 1 (11.1) | |
| *Lipidomics (n = 1)* | 1 (100) | | 1 (100) | |
| Authors' conclusion on clinical applicability | | 0.028 | | 0.035 |
| Definitively favorable (*n* = 54) | 20 (37.0) | | 16 (29.6) | |
| Promising (*n* = 49) | 10 (20.4) | | 9 (18.4) | |
| Unfavorable (*n* = 4) | 3 (75.0) | | 3 (75.0) | |
| Authors' conclusion regarding need for further validation | | 0.341 | | 0.446 |
| Mention further validation (*n* = 56) | 15 (26.8) | | 13 (23.2) | |
| Do not mention further validation (*n* = 51) | 18 (35.3) | | 15 (29.4) | |
| Overinterpretation | | 0.731 | | 0.669 |
| Yes (*n* = 61) | 18 (29.5) | | 13 (24.6) | |
| No (*n* = 46) | 15 (32.6) | | 13 (28.3) | |
| Total (*n* = 107) | 33 (30.8) | | 28 (26.2) | |

evaluated in 2006 with patients with established disease and a control group had made a step toward clinical application in the 10 years from discovery. Progress was more common for tests that were based in PCR, in which approximately half had made progress toward clinical application. Proteomics tests using mass spectrometry or 2-dimensional gel electrophoresis were the least likely to have made an advance; only 4 of 44 (9.1%) tests had made a step toward clinical application in the 10-year period studied.

Overall, the most frequent type of progress was an advance in clinical validation of the same diagnostic question (of the 28 tests that had made progress, 17 (60.7%) made an advance of this type), followed by technical improvement of test (13 tests, 46.4%, Table 3). Ten tests (9.3%) were cited by diagnostic studies that had changed the diagnostic question, thereby expanding the potential use of the test. Of the 82 studies that had used a healthy control or alternate-diagnosis control design in 2006, only 7 (8.5%) had advanced in the clinical validation of the test in a more relevant patient series. Eight (9.8%) had made technical improvements (Table 3), and 67 (81%) repeated the diagnostic evaluation with a similar design that did not reflect the patients in whom the test is likely to be used in practice. Whether or not the authors of the initial study had stated that the test needed further validation before clinical application did not appear to influence the likeliness of further clinical validation studies being carried out (16.1% and 15.6%, respectively; Table 3). Only 3 (2.8%) of the tests had made progress in terms of clinical implementation studies and/or studies of clinical use (Table 3).

## Discussion

The Committee on the Review of Omics-Based Tests for Predicting Patient Outcomes in Clinical Trials of the National Academy of Medicine published several recommendations to strengthen the development and evaluation of omics-based test. The report pursued the strengthening of omics-based test development and evaluation to avoid unnecessary use of research resources and prevent application of premature and useless tests. The findings of this study suggest that adherence to the National Academy of Medicine recommendations by research in "-omics" diagnostic technologies has been low in the last 10 years. Of more than 4000 research papers that cited molecular diagnostic studies from 2006, very few were even diagnostic studies of the same disease and molecular test. A substantial proportion was not empirical research, including narrative reviews, which is a notable finding in itself as it shows a high level of scientific acceptance of the proposed biomarkers, even without proper validation. Although this was not analyzed in detail in this paper, the potential for "spin" here is significant. Furthermore, only 1 in 5 of the papers that were diagnostic studies described research that constituted progress toward clinical application of the test. It should be noted that our definition of "progress toward clinical application" was rather generous (we classified the test as having made progress if there was just 1 study over the course of 10 years that was considered a positive step toward validating the test for clinical use). A stricter definition of progress would find an even larger gap between discovery and application. It appears that the current challenge confronting molecular-based diagnostic research is not the development of equipment but rather the interpretation and analysis of data and the movement toward implementation of clinically useful tools, which ultimately produce improved health outcomes. Only 3 of the 107 tests had evaluations of clinical implementation and/or use *(17–19)*.

Approximately 20% of the tests that had been evaluated with a case control in 2006 made progress in the following 10 years, and of those that did make progress, many focused on technical aspects of the test or extended diagnostic application rather than diagnostic validation in a more relevant patient series. Critical analysis of research production and funding has observed a bias toward preclinical studies and the persistent gap between the type of research needed and the type of research that is ultimately produced *(4)*.

Despite being the subject of studies that received fewer citations, PCR-based diagnostic tools were more likely to make progress toward clinical application, which was perhaps a result of the relative ease of their reproduction compared with "-omics" technologies. Little over 1 in 10 of the proteomics tests from 2006 had made progress toward clinical application. This is remarkable if we consider that replication is essential in the omics fields given the analytical and computational complexities of this type of research, which can lead to errors. Proteomics has the potential to greatly effect clinical diagnosis and drug discovery, but progress in this area has been slower than in genomics. The higher cost of proteomic techniques, the lack of reproducibility often ascribed to small samples, and the characteristics of proteins in comparison with genes (the proteome, unlike the fixed genome of the cell, possesses an intrinsic complexity and is unstable *(20, 21)*) could explain the limited progress in proteomics research. Moreover, the availability of antibodies is considerably lower than the great number of potential protein candidates *(22)*.

The sources of error introduced during preanalytical and analytical phases of test development are not limited to proteomic biomarkers. There are several preanalytical factors affecting the analysis of mRNA production, such as cold ischemia time, specimen size, and block storage *(10)*. A major concern in "-omics"-based research is the translation of "signatures" from biological assays into

**Table 3. Types of progress in validation of 107 molecular diagnostic tests for use in clinical practice over a 10-year period.**

| Main characteristics of 107 diagnostic studies of molecular tests published in 2006 | No progress | Advance in Clinical validation | Technical improvement | Extended diagnostic application | Economic evaluation | Clinical utility or implementation |
|---|---|---|---|---|---|---|
| Study design | | | | | | |
| Healthy control or alternative-diagnosis control (*n* = 82) | 67 (81.7) | 7 (8.5) | 8 (9.8) | 6 (7.3) | 0 | 0 |
| Consecutive series or series of clinically relevant patients (*n* = 14) | 6 (42.9) | 6 (48) | 4 (28.6) | 3 (21.4) | 1 (7.1) | 3 (21.4) |
| Other (*n* = 11) | 6 (54.5) | 4 (36.4) | 1 (9.1) | 1 (9.1) | 0 | 0 |
| Journal category | | | | | | |
| Medical (*n* = 35) | 27 (77.1) | 5 (14.3) | 4 (11.4) | 3 (8.6) | 0 | 2 (5.7) |
| Oncology (*n* = 32) | 24 (75) | 5 (15.6) | 1 (2.9) | 3 (9.3) | 0 | 0 |
| Biomedical or general science (*n* = 19) | 12 (63.2) | 5 (26.3) | 4 (21.1) | 2 (10.5) | 1 (5.3) | 1 (5.3) |
| Lab and methodology (*n* = 21) | 16 (76.1) | 2 (9.1) | 4 (19.0) | 2 (9.5) | 0 | 0 |
| Ranking of journal citation report | | | | | | |
| Q1 (*n* = 61) | 44 (72.1) | 11 (18) | 7 (11.5) | 8 (13.1) | 0 | 2 (3.3) |
| Q2 (*n* = 20) | 13 (65) | 5 (25) | 3 (15) | 2 (10.0) | 2 (10) | 1 (5) |
| Q3 (*n* = 14) | 10 (71.4) | 1 (7.1) | 3 (21.4) | 0 | 0 | 0 |
| Q4 (*n* = 3) | 3 (100) | 0 | 0 | 0 | 0 | 0 |
| Non–journal citation report (*n* = 9) | 9 (100) | 0 | 0 | 0 | 0 | 0 |
| Disease type | | | | | | |
| Autoimmune disease and transplant rejection (*n* = 8) | 7 (87.5) | 0 | 0 | 1 (12.5) | 0 | 0 |
| Cancer (*n* = 61) | 49 (80.3) | 8 (13.1) | 4 (6.6) | 5 (8.2) | 0 | 1 (1.6) |
| Congenital disorder (*n* = 9) | 5 (55.6) | 1 (11.1) | 4 (44.4) | 1 (11.1) | 0 | 0 |
| Infectious disease (*n* = 19) | 11 (57.9) | 7 (36.8) | 4 (21.1) | 2 (10.5) | 1 (5.3) | 0 |
| Neurological (*n* = 6) | 3 (50) | 1 (16.7) | 1 (16.7) | 1 (16.7) | 0 | 2 (33.3) |
| Other (*n* = 4) | 4 (100) | 0 | 0 | 0 | 0 | 0 |
| Type of molecular technology | | | | | | |
| *Gene-targeting tests* | 31 (58.5) | 14 (26.4) | 10 (18.9) | 7 (13.2) | 0 | 2 (3.8) |
| PCR based (*n* = 33) | 16 (48.5) | 10 (30.3) | 7 (21.1) | 5 (15.2) | 0 | 2 (6.1) |
| Microarray (*n* = 20) | 15 (75) | 4 (20) | 3 (15.0) | 2 (10.0) | 0 | 0 |
| *Protein-targeting tests* | 48 (90) | 3 (5.7) | 2 (3.8) | 3 (5.7) | 1 (1.9) | 1 (1.9) |
| Mass spectrometry or 2-dimensional gel electrophoresis (*n* = 44) | 40 (90.9) | 2 (4.5) | 1 (2.3) | 3 (6.8) | 1 (2.3) | 1 (2.3) |
| Antibody array or protein microarray (*n* = 9) | 8 (88.9) | 1 (11.1) | 1 (11.1) | 0 | 0 | 0 |
| *Lipidomics (n = 1)* | 0 | 0 | 1 (100.0) | 0 | 0 | 0 |

*Continued on page 1665*

**Table 3. Types of progress in validation of 107 molecular diagnostic tests for use in clinical practice over a 10-year period.** *(Continued from page 1664)*

| Main characteristics of 107 diagnostic studies of molecular tests published in 2006 | No progress | Advance in Clinical validation | Technical improvement | Extended diagnostic application | Economic evaluation | Clinical utility or implementation |
|---|---|---|---|---|---|---|
| Authors' conclusion on clinical applicability | | | | | | |
| Definitively favorable ($n$ = 54) | 38 (70) | 10 (18.5) | 10 (18.5) | 6 (11.1) | 1 (1.8) | 2 (3.7) |
| Promising ($n$ = 49) | 40 (81.6) | 5 (10.5) | 2 (4.1) | 3 (6.1) | 0 | 0 |
| Unfavorable ($n$ = 4) | 1 (25) | 2 (50) | 1 (25) | 1 (25) | 0 | 1 (25) |
| Authors' conclusion regarding need for further validation | | | | | | |
| Mention further validation ($n$ = 56) | 43 (76.8) | 9 (16.1) | 8 (14.3) | 5 (8.9) | 1 (1.8) | 2 (3.6) |
| Do not mention further validation ($n$ = 51) | 36 (70.6) | 8 (15.6) | 5 (9.8) | 5 (9.8) | 0 | 1 (2) |
| Overinterpretation | | | | | | |
| Yes ($n$ = 61) | 46 (75.4) | 10 (16.4) | 8 (13.1) | 5 (8.2) | 1 (1.6) | 1 (1.6) |
| No ($n$ = 46) | 33 (71.7) | 7 (15.2) | 5 (10.9) | 5 (10.9) | 0 | 2 (4.3) |
| Total ($n$ = 107) | 79 (73.8) | 17 (15.9) | 13 (12.1) | 10 (9.3) | 1 (0.9) | 3 (2.8) |

clinical results relevant for patient management. The lack of reproducibility in previous studies that implemented results from microarray profiles to better predict the cytotoxic agents to which a patient would respond allowed inaccurate genomic signatures to guide care in clinical trials *(23)*. Transparently available data and an explanation about how data have been used could help avoid such situations *(24)*. Nevertheless, one study showed that out of 18 microarray studies, the authors were able to access data for 10 and could reproduce quantitative results for just 2 *(25)*.

There are also numerous challenges with data analysis, and it has been suggested that clinicians and statisticians should work together to incorporate clinical and demographic information with assay data in all the steps of the research: experimental design, data visualization, preprocessing, and biomarker identification *(11)*. Overinterpretation or "spin" has been commonly reported in diagnostic accuracy studies *(2, 14, 15)* regardless of bias and pitfalls in "-omics" research. Researchers are vulnerable to "spin," both as producers and as consumers, because they are generally unaware of the preanalytical and analytical processes that may have biased the research findings. Misleading claims about biomarker use may be picked up by prominent physicians and be fed to laypersons, sometimes via further "spun" media headlines, which can in turn attract venture capitalists to secure private funding to continue development of potentially useless biomarkers. Ultimately, "spin" can confuse all types of readers and policy makers and potentially harm patients and increase healthcare costs.

Our results showed that studies that had overinterpreted the applicability of the test were just as frequently cited as studies that did not overinterpret and they were equally likely to have progressed toward clinical validation. Similarly, progress was not more common for tests that had been described originally in high-impact journals. Although unexpected, these negative findings are still important as they highlight the limited progress made in terms of clinical application regardless of spin and publication practices.

We initially planned to quantify which of the 107 tests are now commercially available for use in clinical or public health practice but were unable to do so because the lack of regulatory standards for the evaluation of diagnostic tests *(26)*. This is in stark contrast with the regulation of pharmacological evaluations, and it is a problem that is even more marked among new molecular technologies *(27)*, in which regulatory authorities in different regions of the world recognize different classes of medical devices *(28)*. In the European Union, there is no official agency tasked with harmonizing requirements and regulating medical devices. While the Food and Drug Administration undertakes this task for the US, in the European Union, medical device regulation is the

responsibility of each of the 27 member states; therefore, there is a wide variety of sources and websites that include information on molecular tests *(29)*. Furthermore, medical devices are not directly subject to any premarket authorization by a regulatory authority. They are subject to a conformity assessment, which for medium- and high-risk devices involves a commercial, independent third party, known as a "notified body." There are around 80 notified bodies across Europe designated and monitored by member states under control of national authorities, but manufacturers are free to choose any notified body to carry out the conformity assessment.

In light of this complex situation, we limited our evaluation of progress in clinical application to an analysis of published research detected in our search of the Scopus search. This decision may have affected the results because it is possible for test development to progress with studies financed by manufacturers not resulting in academic publications. Similarly, it is possible that studies on the same test have been carried out, but they do not cite the original paper and hence were not detected in the Scopus search, or that further diagnostic studies of the tests were carried out but never published. Chalmers and Glasziou acknowledge that research studies with disappointing results are more likely to be never published or published in gray literature *(8)*. Prospective trial registration initiatives have been successful in providing a tool for evaluating potential publication biases and reducing research waste in drug discovery research, but diagnostic accuracy studies are rarely registered *(30)*. We showed that approximately 1 in 5 of the 118 studies carrying out further diagnostic research on the 2006 biomarker studies was potentially research waste, but it should be acknowledged that this sample is limited in size and representative only of the 2006 studies. Furthermore, the proportion of research waste related to these 2006 biomarker studies may indeed be higher if we consider all 4257 studies that cited them in the following 10 years, as mentioned above. We should also acknowledge that this analysis was limited to discovery of biomarkers to aid diagnosis, and we did not perform analyses regarding translation of potential biomarkers to clinically useful prognostic tools or for aiding treatment choice. Given the challenges observed in diagnostic research, we feel it is unlikely that significant progress has been made in these other arenas, but that is not to say that money and resources pumped into biomarker discovery research have all been in vain. Improved analytical techniques for many classes of molecules, particularly low molecular weight analytes, have stemmed from the biomarker discovery world and have been successfully integrated into clinical practice *(31)*.

In conclusion, despite intense research on our sample of 107 molecular diagnostic tests, few made significant advances toward clinical application in the 10 years from their discovery. Of the 4000+ articles citing these molecular diagnostic tools over the 10-year period, <5% were diagnostic studies on the same test, and of those that did continue diagnostic research on the tool, 20% were classified as research waste. Strong regulatory standards for the evaluation of molecular diagnostic tools are desperately needed.

**Author Contributions:** *All authors confirmed they have contributed to the intellectual content of this paper and have met the following 4 requirements:* (*a*) *significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data;* (*b*) *drafting or revising the article for intellectual content;* (*c*) *final approval of the published article; and* (*d*) *agreement to be accountable for all aspects of the article thus ensuring that questions related to the accuracy or integrity of any part of the article are appropriately investigated and resolved.*

L.A. Parker, statistical analysis; I. Hernandez-Aguado, financial support; B. Lumbreras, statistical analysis.

**Authors' Disclosures or Potential Conflicts of Interest**: *Upon manuscript submission, all authors completed the author disclosure form. Disclosures and/or potential conflicts of interest:*

**Employment or Leadership:** None declared.
**Consultant or Advisory Role:** None declared.
**Stock Ownership:** None declared.
**Honoraria:** None declared.
**Research Funding:** CIBER Epidemiología y Salud Pública (CIBERESP), Madrid.
**Expert Testimony:** None declared.
**Patents:** None declared.

**Role of Sponsor:** The funding organizations played no role in the design of study, choice of enrolled patients, review and interpretation of data, preparation of manuscript, or final approval of manuscript.

## References

1. Joyner MJ, Paneth N. Seven questions for personalized medicine. JAMA 2015;314:999–1000.
2. Lumbreras B, Parker LA, Porta M, Pollán M, Ioannidis JPA, Hernández-Aguado I. Overinterpretation of clinical applicability in molecular diagnostic research. Clin Chem 2009;55:786–94.
3. Bossuyt PM. Where are all the new omics-based tests? Clin Chem 2014;60:1256–7.
4. Ioannidis JPA. Why most clinical research is not useful. PLoS Med 2016;13:e1002049.
5. Ioannidis JPA, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, Moher D, et al. Increasing value and reducing waste in research design, conduct, and analysis. Lancet 2014;383:166–75.
6. Ioannidis JPA. Is it possible to recognize a major scientific discovery? JAMA 2015;314:1135–7.
7. Dechartres A, Ravaud P. Better prioritization to increase research value and decrease waste. BMC Med 2015;13:244.
8. Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. Lancet 2009;374:86–9.
9. Issaq HJ, Waybright TJ, Veenstra TD. Cancer biomarker discovery: opportunities and pitfalls in analytical methods. Electrophoresis 2011;32:967–75.
10. Bass BP, Engel KB, Greytak SR, Moore HM. A review of preanalytical factors affecting molecular, protein, and morphological analysis of formalin-fixed, paraffin-embedded (FFPE) tissue: how well do you know your FFPE specimen? Arch Pathol Lab Med 2014;138:1520–30.

11. Morris JS, Baggerly KA, Gutstein HB, Coombes KR. Statistical contributions to proteomic research. Methods Mol Biol 2010;641:143–66.
12. Drucker E, Krapfenbauer K. Pitfalls and limitations in translation from biomarker discovery to clinical utility in predictive and personalised medicine. EPMA J 2013;4:7.
13. Pepe MS, Etzioni R, Feng Z, Potter JD, Thompson ML, Thornquist M, et al. Phases of biomarker development for early detection of cancer. J Natl Cancer Inst 2001;93:1054–61.
14. McGrath TA, McInnes MDF, van Es N, Leeflang MMG, Korevaar DA, Bossuyt PMM. Overinterpretation of research findings: evidence of "spin" in systematic reviews of diagnostic accuracy studies. Clin Chem 2017;63:1353–62.
15. Ochodo EA, de Haan MC, Reitsma JB, Hooft L, Bossuyt PM, Leeflang MMG. Overinterpretation and misreporting of diagnostic accuracy studies: evidence of "spin." Radiology 2013;267:581–8.
16. Committee on the Review of Omics-Based Tests for Predicting Patient Outcomes in Clinical Trials; Board on Health Care Services; Board on Health Sciences Policy; Institute of Medicine. Evolution of translational omics: lessons learned and the path forward. Washington DC: National Academy of Sciences; 2012. http://www.ncbi.nlm.nih.gov/pubmed/24872966 (Accessed August 2018).
17. Grossman HB, Soloway M, Messing E, Katz G, Stein B, Kassabian V, et al. Surveillance for recurrent bladder cancer using a point-of-care proteomic assay. JAMA 2006;295:299–305.
18. Franco-Alvarez de Luna F, Ruiz P, Gutiérrez J, Casal M. Evaluation of the GenoType Mycobacteria Direct assay for detection of Mycobacterium tuberculosis complex and four atypical mycobacterial species in clinical samples. J Clin Microbiol 2006;44:3025–7.
19. Jordan JA, Durso MB, Butchko AR, Jones JG, Brozanski BS. Evaluating the near-term infant for early onset sepsis: progress and challenges to consider with 16S rDNA polymerase chain reaction testing. J Mol Diagn 2006;8:357–63.
20. Ilyin SE, Belkowski SM, Plata-Salamán CR. Biomarker discovery and validation: technologies and integrative approaches. Trends Biotechnol 2004;22:411–6.
21. Shuford CM, Walters JJ, Holland PM, Sreenivasan U, Askari N, Ray K, et al. Absolute protein quantification by mass spectrometry: not as simple as advertised. Anal Chem 2017;89:7406–15.
22. Madoz-Gúrpide J, López-Serra P, Martínez-Torrecuadrada JL, Sánchez L, Lombardía L, Casal JI. Proteomics-based validation of genomic data: applications in colorectal cancer diagnosis. Mol Cell Proteomics 2006;5:1471–83.
23. Potti A, Dressman HK, Bild A, Riedel RF, Chan G, Sayer R, et al. Retraction: genomic signatures to guide the use of chemotherapeutics. Nat Med 2011;17:135.
24. Baggerly K. Disclose all data in publications. Nature 2010;467:401.
25. Ioannidis JPA, Allison DB, Ball CA, Coulibaly I, Cui X, Culhane AC, et al. Repeatability of published microarray gene expression analyses. Nat Genet 2009;41:149–55.
26. Mustafa RA, Wiercioch W, Cheung A, Prediger B, Brozek J, Bossuyt P, et al. Decision making about healthcare-related tests and diagnostic test strategies. Paper 2: a review of methodological and practical challenges. J Clin Epidemiol 2017;92:18–28.
27. Lay JO, Liyanage R, Borgmann S, Wilkins CL. Problems with the "omics." TrAC Trends Anal Chem 2006;25:1046–56.
28. George B. Registration of medical devices. Perspect Clin Res 2010;1:90–3.
29. Campillo-Artero C. A full-fledged overhaul is needed for a risk and value-based regulation of medical devices in Europe. Health Policy 2013;113:38–44.
30. Korevaar DA, Hooft L, Askie LM, Barbour V, Faure H, Gatsonis CA, et al. Facilitating prospective registration of diagnostic accuracy studies: a STARD initiative. Clin Chem 2017;63:1331–41.
31. Grebe SK, Singh RJ. LC-MS/MS in the clinical laboratory–where to from here? Clin BioChem Rev 32:5–31.