UvA-DARE (Digital Academic Repository)

The Thin Line Between Hope and Hype in Biomarker Research

Bossuyt, P.M.M.

DOI

[10.1001/jama.2011.729](https://doi.org/10.1001/jama.2011.729)

Publication date

2011

Document Version

Submitted manuscript

Published in

JAMA

[Link to publication](#)

Citation for published version (APA):

Bossuyt, P. M. M. (2011). The Thin Line Between Hope and Hype in Biomarker Research. *JAMA*, 305(21), 2229-2230. <https://doi.org/10.1001/jama.2011.729>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

The thin line between hope and hype in biomarker research

Patrick MM Bossuyt, PhD

Department of Clinical Epidemiology, Biostatistics and Bioinformatics

University of Amsterdam, The Netherlands

Address correspondence to:

Patrick MM Bossuyt, Dept. Clinical Epidemiology & Biostatistics, Academic Medical Center,

Room J1b-214; PO Box 22700; 1100 DE Amsterdam; the Netherlands

p.m.bossuyt@amc.uva.nl +31(20)566 3240 (voice) +31(20)691 2683(fax)

Words: 1,198

Biomarkers are definitely a new buzzword in medicine, and investigations of putative molecular indicators of a specific biological state have started to occupy a considerable part of health research. In the past decades, advances in molecular biology coupled with progress in genomics, proteomics and metabolomics have fuelled high hopes for the development of new medical tests. Biomarkers should enable clinicians to make an earlier or more definitive diagnosis, to identify those at risk of developing disease, to develop more fine-grained statements about prognosis, and to tune treatment selection, bringing us all closer to a form of stratified, or even personalized, medicine.

With a few exceptions, most of these promises have yet to be fulfilled. Only a small number of biomarkers have made it to routine clinical practice (1). No new major cancer biomarkers have been approved for clinical use for at least 25 years (2). Most clinical decisions still have to rely on more conventional forms of medical testing, such as existing laboratory measurements and imaging.

A number of reasons can be mentioned for the relatively slow speed of progress. For many conditions, for example, molecular biomarkers have yet to be identified. Few molecular biomarkers have been discovered for several highly prevalent conditions. There have been issues with the characterization and control of the preanalytical variability (2). Maybe one of the major reasons for the disappointment has been the suboptimal design of the studies for marker discovery and validation. Many of these suffer from methodological shortcomings, in particular in the selection of appropriate study groups. Several include extreme cases and contrast them with very healthy controls. Nevertheless, hopes have floated high, and hype has never been far away.

In this issue of the Journal, John Ionnides and his colleague Orestis Panagiotou demonstrate how frequently cited biomarker studies report effect sizes that are often higher than those in larger studies of the same biomarker, and more extreme than summary estimates in a meta-analysis of that marker.

For 29 of the 35 studies in their analysis, the meta-analysis, published later, resulted in a less optimistic estimate than the highly cited study.

Some of these stories are striking. A 1994 Lancet paper, for example, contained the findings from a study on cancer risk in 33 families with evidence of linkage to BRCA1 carriers. After comparing cancer cases other than breast or ovarian to national incidence rates, the authors reported a 4.11 relative excess risk for colon cancer in carriers. In a study published nine years later, data were summarized from more than 30 epidemiologic studies on cancer incidence in BRCA1 mutation carriers. The studies on colon cancer that had appeared after the 1994 study all had reported smaller, and often non-significant relative risks. One of these, published in JNCI in 2004, reported a non-significant odds ratio of 1.24 (3). The latter study has received only 26 citations so far, compared to 1,051 for the 1994 Lancet paper.

Another example is the story of hyperhomocystenemia. In a 1991 NEJM paper, authors reported high peak serum levels of homocysteine in 16 of 38 patients with cerebrovascular disease, 7 of 25 with peripheral vascular disease, 18 of 60 with coronary vascular disease, but in none of 27 normal subjects (4). In summary, they reported a significant odds ratio of 23.9 for coronary vascular disease in patients with hyperhomocystenemia. A meta-analysis of hyperhomocystenemia, published in the American Journal of Cardiology nine years later, was able to include 33 studies and over 16,000 patients (5). The summary odds ratio for cardiovascular disease in that systematic review was only 1.58. The NEJM paper has received 1,451 citations so far, the meta-analysis 37.

It is difficult to estimate how common these effects are. How often does a study that publishes a more extreme effect receive more attention than larger studies of the same marker, or than meta-analyses, which provide a summary estimate based on all available evidence, after critical appraisal? We do not really know. The review offered by Ioannidis and colleagues is not exactly based on what epidemiologists would call an “inception cohort”: a group of studies of a biomarker defined from the

very first evaluation. Ioannidis first selected highly cited studies, and then tried to find a matching meta-analysis, published after the highly cited study. They used an arbitrary threshold of 400 citations, regardless the date of publication of the index study, and were able to match less than half of the highly cited studies to a meta-analysis.

We also do not know enough about the dynamics of study initiation, a condition for inclusion in such an analysis. New studies to evaluate biomarkers do not happen at random. Nor are they started out of ignorance of previous studies, in most cases at least. Investigators design a study and seek funding for it because they feel they have something useful, promising or otherwise worthwhile at hand. A large number of factors drive the research agenda. Admittedly, some of these can be quite mundane, but they are not random. If one would set out a list of all potential biomarkers, the likelihood of an additional evaluation is not distributed evenly across biomarkers. Nor should it be. This may help to explain why a study with a smaller effect more often follows a study with a large effect than the other way around. In the latter case, the second study may never be started. Analogously, the likelihood that someone decides to do a meta-analysis is not homogeneous for all biomarkers.

Even more complicated processes affect the citation of previous studies. Studies that are published early definitely have a citation advantage: they have more time to accrue citations. Yet the reasons for citing previous papers, the mechanisms for copying early citations from previous papers, the hurdles before switching citations to more complete or more precise studies, all of these deserve further study. The science of the scientific reception of biomarker evaluation studies is still in its infancy.

Still, the analysis presented here shows some very convincing case studies, where more extreme, often early associations receive considerable attention and continue to do so, despite the availability of studies or meta-analyses with more precise estimates. Ioannides and Panagiotou do not explain the citation advantage of the highly cited studies, but they refer to several mechanisms that may be

responsible for the inflated effects in them. Many studies were small, so chance plays an important role. Several used a two-gate or case-control design, which is known to generate inflated results (6). Most of these deficiencies can be remedied. Better study designs exist, for example (7). The completeness and transparency of reporting may be improved through the use of STARD, REMARK and similar checklists, so editors, reviewers and readers can detect study weaknesses more easily (8).

It would be premature to become suspicious of all scientific efforts at marker discovery, and unwise to burn all marker evaluation studies at the stake from here on. Yet the analysis presented here should convince us even more in being careful to match personal hope with professional skepticism, to apply critical appraisal of study design and close scrutiny of findings where indicated, and to rely on well conducted systematic reviews and meta-analyses above anything else, whenever they exist.

Conflict of Interest: None

References

1. Rifai N, Gillette MA, Carr SA. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nature biotechnology*. 2006;24(8):971-83.
2. Diamandis EP. Cancer biomarkers: can we turn recent failures into success? *Journal of the National Cancer Institute*. 2010;102(19):1462-7.
3. Niell BL, Rennert G, Bonner JD, Almog R, Tomsho LP, Gruber SB. BRCA1 and BRCA2 founder mutations and the risk of colorectal cancer. *Journal of the National Cancer Institute*. 2004;96(1):15-21.
4. Clarke R, Daly L, Robinson K, Naughten E, Cahalane S, Fowler B, et al. Hyperhomocysteinemia: an independent risk factor for vascular disease. *The New England Journal of Medicine*. 1991;324(17):1149-55.
5. Cleophas TJ, Hornstra N, van Hoogstraten B, van der Meulen J. Homocysteine, a risk factor for coronary artery disease or not? A meta-analysis. *The American journal of cardiology*. 2000;86(9):1005-9, A8.
6. Rutjes AW, Reitsma JB, Vandenbroucke JP, Glas AS, Bossuyt PM. Case-control and two-gate designs in diagnostic accuracy studies. *Clin.Chem*. 2005;51(8):1335-41.
7. Pepe MS, Feng Z, Janes H, Bossuyt PM, Potter JD. Pivotal evaluation of the accuracy of a biomarker used for classification or prediction: standards for study design. *Journal of the National Cancer Institute*. 2008;100(20):1432-8.
8. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *BMJ*. 2003;326(7379):41-4.