



OPINION ARTICLE

How difficult is the validation of clinical biomarkers? [v1; ref status: indexed, <http://f1000r.es/5a1>]

Jan Voskuil

Everest Biotech Ltd, Upper Heyford, OX255HD, UK

v1 First published: 28 Apr 2015, 4:101 (doi: [10.12688/f1000research.6395.1](https://doi.org/10.12688/f1000research.6395.1))
 Latest published: 28 Apr 2015, 4:101 (doi: [10.12688/f1000research.6395.1](https://doi.org/10.12688/f1000research.6395.1))

Abstract

Recent developments of introducing stratified medicine/personal health care have led to an increased demand for specific biomarkers. However, despite the myriads of biomarkers claimed to be fit for all sorts of diseases and applications, the scientific integrity of the claims and therefore their credibility is far from satisfactory. Biomarker databases are met with scepticism. The reasons for this lack of faith come from different directions: lack of integrity of the biospecimen and meta-analysis of data derived from biospecimen prepared in various ways cause incoherence and false indications. Although the trend for antibody-independent assays is on the rise, demand for consistent performance of antibodies (both in choice of antibody and how to apply it in the correct dilution where applicable) in immune assays remains unmet in too many cases. Quantitative assays suffer from a lack of world-wide accepted criteria when the immune assay is not ELISA-based. Finally, statistical analysis suffer from coherence both in the way software packages are being scrutinized for mistakes in the script and remaining invisible after small-scale analysis, and in the way appropriate queries are fed into the packages in search for output that is fit for the types of data put in. Wrong queries would lead to wrong statistical conclusions, for example when data from a cohort of patients with different backgrounds are being analysed, or when one seeks an answer from software that was not designed for such query.

Open Peer Review

Referee Status:

	Invited Referees	
	1	2
version 1 published 28 Apr 2015	 report	 report
1 Bjorn LDM Brücher, Theodor-Billroth-Academy Germany		
2 Ijaz S Jamall, Risk-Based Decisions, Inc. USA		
Discuss this article		
Comments (0)		

Corresponding author: Jan Voskuil (voskuil@everestbiotech.com)

How to cite this article: Voskuil J. **How difficult is the validation of clinical biomarkers? [v1; ref status: indexed, <http://f1000r.es/5a1>]** *F1000Research* 2015, 4:101 (doi: [10.12688/f1000research.6395.1](https://doi.org/10.12688/f1000research.6395.1))

Copyright: © 2015 Voskuil J. This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the [Creative Commons Zero "No rights reserved" data waiver](https://creativecommons.org/licenses/by/4.0/) (CC0 1.0 Public domain dedication).

Grant information: The author(s) declared that no grants were involved in supporting this work.

Competing interests: The author is the Chief Scientific Officer of Everest Biotech Ltd, a research antibody manufacturer specialised in peptide-generated reagents from goat. Although the author highlights the value of peptide-generated antibodies in animals larger than rodents or rabbits as cost-effective alternative to monoclonal antibodies under specific circumstances, this notion should not be deemed as sole advertisement for Everest antibodies, since goat and other large animals are being used by other manufacturers.

First published: 28 Apr 2015, 4:101 (doi: [10.12688/f1000research.6395.1](https://doi.org/10.12688/f1000research.6395.1))

First indexed: 11 May 2015, 4:101 (doi: [10.12688/f1000research.6395.1](https://doi.org/10.12688/f1000research.6395.1))

Introduction

Clinical biomarkers have been around for a long time now, and the field is moving rapidly. In addition to genetic and protein markers, we now also have microRNAs, epigenetic markers, lipids, metabolites, and imaging markers. Some are extremely useful as a (companion-) diagnostic; others may serve as a mere indicator. However, there are problems. There is confusion on the nomenclature and on the way how biomarkers are meant to be validated and used. A proposal published in 2006 was meant to create some clarity and consistency in the matter¹. The biggest obstacle by far is that Biomarker validation and qualification depend on confirmation at different locations (different labs). There are issues with consistency in the preparation of the biological material used in the different studies, and with consistency in the choice of antibody when required. It should also be noted that in quantitative immunohistochemistry (IHC) one needs a standard in the quantification method². A recent opinion paper reveals yet another layer of complexity: The statistical analysis is prone to wrong conclusions down to coding errors in the software³. It may not be a surprise then that one another led to the observation that only about 11% of preclinical research papers demonstrated reproducible results⁴. It is time to take stock and to address the different levels of disturbance complicating the process of biomarker validation and qualification.

Biological material

The integrity of the tissue specimens will determine the quality of the biomarker's measurements, especially when biomarkers are instable. Post-mortem samples in particular will never represent samples from living individuals because of the post-mortem delay. As the post-mortem delay will differ from individual to individual, the level of decay will vary dramatically per sample. For this reason, post-mortem samples are best fit for qualitative analysis. Quantification of any biomarker in post-mortem samples should be interpreted with extra care⁵.

Plasma samples can be prepared in different ways: they can be prepared either by citrate, by ethylenediaminetetraacetic acid (EDTA) or by heparin. In addition, biomarkers can be tested in serum and in whole blood. It is clear that levels of biomarkers will need to be compared between equally treated samples in order to avoid variations in noise from the different ways the samples were prepared⁶. Since this principle is universal, it will be true for any other tissue types.

For microscopy, tissue slides and cell suspensions have to be prepared in line with the required assay before they can be investigated. Fixatives (alcohols, aldehydes), embedding materials (paraffin, LR White, etc) and temperatures (frozen vs heated) have profound effects on the integrity of the tissues and cells and they will determine the success of the assay. Again, consistency in the tissue preparation, tissue sections and cells to be analysed is paramount^{7,8}. Mega-data analysis may get skewed when data are collated from samples treated in different ways.

A systematic approach to record and keep biospecimen has been proposed and is aimed to become the new standard: Biospecimen Reporting for Improved Study Quality (BRISQ) guidelines provide a tool to improve consistency and to standardize information on the biological samples⁹.

Antibody choice

Mass-spec and RT-PCR quantifications will be robust by the consistency of the assay material. However, the robustness of immune assays depends highly on the choice of antibodies used in the assay. Once an antibody has been successfully validated in one assay, this assay is defined by this antibody. Change of antibody will potentially change the outcome altogether as demonstrated in the past^{10,11}. When an antibody needs changing, the assay is no longer validated and the validation procedure will have to be repeated with the new antibody. For this reason the preference goes to monoclonal antibodies. The rationale behind this preference is that the clone number of the antibody would define its characteristics: the expectation then is that the assay will remain validated because the antibodies remain identical when using antibodies from the same clone number, no matter which vendor they are from. Unfortunately this is a myth. Depending on the vendor (and sometimes depending on the catalogue number) the formulations, all with the same clone number, will differ: the antibody may be purified from ascitic fluid, from culture media, or not purified at all (just ascitic fluid or just culture supernatant). These different formulations will have an effect on the way the antibody needs to be diluted to avoid non-specific background¹². Therefore, the monoclonal antibody needs to be revalidated in the same assay when the original formulation is no longer available. But even subsequent batches from the same formulation show some level of differences, thus undermining the main argument of preference to use monoclonal antibodies in standard assays. A peptide-generated polyclonal antibody from a larger animal than rabbit (for large size batches) may serve as a cost-effective alternative because the batch-to-batch variation of such antibody is limited by the size of the immunizing peptide unlike other polyclonal antibodies¹².

Assay development

When a new assay is being developed a monoclonal antibody may not be always readily available. Then a peptide-generated polyclonal antibody may serve as a good and cost-effective alternative. However, peptide polyclonal antibodies need a new round of validation when a new batch from a different animal arrives, just like different formulated monoclonal antibodies.

During assay development it is essential to dilute the antibody far enough to avoid non-specific background, but it needs to be strong enough to allow measuring a dynamic range, especially when the assay is quantitative. When the assay is dependent on a secondary antibody, this antibody needs validation as well (with and without primary) so to assess its non-specific signals (noise)¹².

Specificity needs to be addressed by comparing specimen spiked and un-spiked with the intended protein of interest (analyte) at various quantities. The signals need to be proportionate to the spiked quantities. In addition, specimen known not to have any of the analyte needs to be compared with specimen known to have the analyte at natural levels¹³.

Detection and cut-off values

Sensitivity is commonly attributed to the antibody used in an assay, but this is a misunderstanding. Sensitivity is determined by the detection method of which the antibody/or primary and secondary antibodies may take part in. If levels of the analyte are low, a higher

sensitivity is required. This increased sensitivity is usually not accomplished by increasing the antibody concentration, although using an antibody with higher affinity will help to some extent. But in general the change of detection method (fluorophore, isotope, PCR, etc.) is the appropriate step to take. Together with the increase of sensitivity, the noise and background will also increase. When a change to a higher sensitivity is required, the validation should focus on a more stringent regime for keeping noise and background at bay¹².

When quantification is a requirement, cut-off values need to be put in place. Both the Lowest Levels Of Quantification (LLOQ) and Highest Levels Of Quantification (HLOQ) must be determined. Often the detection limits are determined as well, but this is only relevant for qualitative work. In IHC these values become tricky, because the intensity of signal is not just a number generated by a detector; the density of signal is combined with the location in the tissue. In addition, the surface area of quantification needs well defined boundaries. And even when all these measures are in place, the quality of the tissue and the quality of the slides can potentially jeopardize these measures and skew the results¹⁴. Diagnostics by IHC is therefore prone to misinterpretation when for one specific test consistency at all levels (same antibody at same dilution, identically prepared tissue samples, identical area surface, identical staining analysed, etc.) is not followed in all laboratories in the world.

Statistics and jumping to conclusions

Statistical analysis is notoriously used to provide the convenient evidence required by the author(s). No matter what method of statistics is used, when the input data have been selected from a larger set, any outcome will be biased and flawed by default. Only analysis of ALL data (non-selected) would yield proper results, but then they might be inconclusive or inconvenient. The pressure to publish in peer-reviewed papers force authors to present statistics in the most incomprehensible way possible, knowing that their peers will not admit their confusion and likely take the author's word for it¹⁵. Even when the statistic results are sound, they may get over-interpreted. Thus original claims were made based on prejudice and weak statistics and only over time, when more scientific details become available, a more complex picture emerged. For example how cholesterol levels are linked to cardiovascular disease^{16,17}, how cancer is not merely caused by mutations^{18,19}, how obesity is not a choice of lifestyle^{20,21} etc. Simplified claims can be (and has been) driven by apparent conflicts of interest as suggested in a study²². The reputation of biomarkers has suffered dramatically from lack of scientific integrity and as a result many scientists lost faith in the usefulness of biomarker databases. New guidelines have been introduced by publishers in order to introduce a new standard on how statistics are presented²³.

There are several statistical packages on the market for scientists and clinicians to use. However, these packages are quite advanced and need expertise handling, very much like a driver's licence is required in order to safely use a motorised vehicle on the public road. Vendors of such packages admit that their products are not always properly used (personal communications). The chosen algorithms

need to be appropriate for the type of data to be analysed: some algorithms are designed for decision making, and they are not necessarily fit for scientific fact finding. In addition, the same data entered in the same system may result in different output on different occasions simply because the wrong type of results is being asked for (personal communications with statistic analysts). Finally, subtle coding errors in the software cannot always be identified in small tests on script integrity, only to skew results when large scale data are being processed³.

Project design and personalized medical care/stratified approaches

When all the above hurdles have been successfully taken, we are not quite there yet. Each individual is different from the next, and therefore each individual has different tolerance or sensitivity to toxins and medicines. This makes the assessment of biomarkers to follow the progress of a disease, or to follow the efficacy of a therapy, difficult to analyse when a group of patients have been treated all in the same way but the individuals in the groups are so diverse in genetic and/or ethnic background that the data can still be all over the place. Only when a group is defined by a certain genetic or environmental background, would there be sufficient homogeneity to assess a biomarker for this particular defined group. For example, only recently it was found that HER2-type breast cancer patients do not benefit as well from therapies when they carry PICK3CA mutations compared to those who do not²⁴. It is like the chicken-egg (catch-22) paradigm: one has to start clinical trials in order to identify the non-responsive patients and only then one can leave them out for proper validation of a new biomarker. However, proper validation demands positive and negative controls and not allowing to select the convenient data only. Although this paradox can be dealt with properly, it is no surprise that the search for proper clinical biomarkers remains very challenging for some time to come.

Competing interests

The author is the Chief Scientific Officer of Everest Biotech Ltd, a research antibody manufacturer specialised in peptide-generated reagents from goat. Although the author highlights the value of peptide-generated antibodies in animals larger than rodents or rabbits as cost-effective alternative to monoclonal antibodies under specific circumstances, this notion should not be deemed as sole advertisement for Everest antibodies, since goat and other large animals are being used by other manufacturers.

Grant information

The author(s) declared that no grants were involved in supporting this work.

Acknowledgements

Special thanks go to the communities of LinkedIn and ResearchGate who have been helpful in their feedback on the troubles on commercial antibodies, on statistical issues around clinical research and subsequent claims and on sample preparation/preservation issues before their use in analysis.

References

1. Lee JW, Devanarayan V, Barrett YC, *et al.*: **Fit-for-purpose method development and validation for successful biomarker measurement.** *Pharm Res.* 2006; **23**(2): 312–28.
[PubMed Abstract](#) | [Publisher Full Text](#)
2. Taylor CR: **Quantitative in situ proteomics; a proposed pathway for quantification of immunohistochemistry at the light-microscopic level.** *Cell Tissue Res.* 2015; **360**(1): 109–20.
[PubMed Abstract](#) | [Publisher Full Text](#)
3. Soergel DAW: **Rampant software errors undermine scientific results [v1; ref status: approved with reservations 2, <http://f1000r.es/4w2>].** *F1000Res.* 2014; **3**: 303.
[PubMed Abstract](#) | [Publisher Full Text](#)
4. Begley CG, Ellis LM: **Drug development: Raise standards for preclinical cancer research.** *Nature.* 2012; **483**(7391): 531–3.
[PubMed Abstract](#) | [Publisher Full Text](#)
5. Nacul L, O'Donovan DG, Lacerda EM, *et al.*: **Considerations in establishing a post-mortem brain and tissue bank for the study of myalgic encephalomyelitis/chronic fatigue syndrome: a proposed protocol.** *BMC Res Notes.* 2014; **7**: 370.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
6. Tvedt TH, Rye KP, Reikvam H, *et al.*: **The importance of sample collection when using single cytokine levels and systemic cytokine profiles as biomarkers - a comparative study of serum versus plasma samples.** *J Immunol Methods.* 2015; **418**: 19–28.
[PubMed Abstract](#) | [Publisher Full Text](#)
7. Babic A, Loftin IR, Stanislaw S, *et al.*: **The impact of pre-analytical processing on staining quality for H&E, dual hapten, dual color in situ hybridization and fluorescent in situ hybridization assays.** *Methods.* 2010; **52**(4): 287–300.
[PubMed Abstract](#) | [Publisher Full Text](#)
8. Howat WJ, Lewis A, Jones P, *et al.*: **Antibody validation of immunohistochemistry for biomarker discovery: recommendations of a consortium of academic and pharmaceutical based histopathology researchers.** *Methods.* 2014; **70**(1): 34–38.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
9. Moore HM, Kelly AB, Jewell SD, *et al.*: **Biospecimen reporting for improved study quality (BRISQ).** *J Proteome Res.* 2011; **10**(8): 3429–38.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
10. Bucur O, Pennarun B, Stancu AL, *et al.*: **Poor antibody validation is a challenge in biomedical research: a case study for detection of c-FLIP.** *Apoptosis.* 2013; **18**(10): 1154–62.
[PubMed Abstract](#) | [Publisher Full Text](#)
11. Anagnostou VK, Welsh AW, Giltneane JM, *et al.*: **Analytic variability in immunohistochemistry biomarker studies.** *Cancer Epidemiol Biomarkers Prev.* 2010; **19**(4): 982–91.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
12. Voskuij J: **Commercial antibodies and their validation [v2; ref status: indexed, <http://f1000r.es/4jlp>].** *F1000Res.* 2014; **3**: 232.
[PubMed Abstract](#) | [Publisher Full Text](#)
13. Khan MU, Bowsher RR, Cameron M, *et al.*: **Recommendations for adaptation and validation of commercial kits for biomarker quantification in drug development.** *Bioanalysis.* 2015; **7**(2): 229–42.
[PubMed Abstract](#) | [Publisher Full Text](#)
14. Wolff AC, Hammond ME, Schwartz JN, *et al.*: **American Society of Clinical Oncology/College of American Pathologists guideline recommendations for human epidermal growth factor receptor 2 testing in breast cancer.** *Arch Pathol Lab Med.* 2007; **131**(1): 18–43.
[PubMed Abstract](#)
15. Bohannon J: **Who's afraid of peer review?** *Science.* 2013; **342**(6154): 60–5.
[PubMed Abstract](#) | [Publisher Full Text](#)
16. Kuivenhoven JA, Groen AK: **Beyond the genetics of HDL: why is HDL cholesterol inversely related to cardiovascular disease?** *Handb Exp Pharmacol.* 2015; **224**: 285–300.
[PubMed Abstract](#) | [Publisher Full Text](#)
17. Dashti M, Kulik W, Hoek F, *et al.*: **A phospholipidomic analysis of all defined human plasma lipoproteins.** *Sci Rep.* 2011; **1**: 139.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
18. Brücher BL, Jamall IS: **Epistemology of the origin of cancer: a new paradigm.** *BMC Cancer.* 2014; **14**: 331.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
19. Baker SG: **A cancer theory kerfuffle can lead to new lines of research.** *J Natl Cancer Inst.* 2014; **107**(2): pii: dju405.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
20. Locke AE, Kahali B, Berndt SI, *et al.*: **Genetic studies of body mass index yield new insights for obesity biology.** *Nature.* 2015; **518**(7538): 197–206.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
21. Shungin D, Winkler TW, Croteau-Chonka DC, *et al.*: **New genetic loci link adipose and insulin biology to body fat distribution.** *Nature.* 2015; **518**(7538): 187–96.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
22. Sitges-Serra A: **Clinical guidelines at stake.** *J Epidemiol Community Health.* 2014; **68**(10): 906–8.
[PubMed Abstract](#) | [Publisher Full Text](#)
23. McNutt M: **Journals unite for reproducibility.** *Science.* 2014; **346**(6210): 679.
[PubMed Abstract](#) | [Publisher Full Text](#)
24. Majewski IJ, Nuciforo P, Mitterpergher L, *et al.*: **PIK3CA Mutations Are Associated With Decreased Benefit to Neoadjuvant Human Epidermal Growth Factor Receptor 2-Targeted Therapies in Breast Cancer.** *J Clin Oncol.* 2015; **33**(12): 1334–9.
[PubMed Abstract](#) | [Publisher Full Text](#)

Open Peer Review

Current Referee Status:



Version 1

Referee Report 11 May 2015

doi:10.5256/f1000research.6861.r8491



Ijaz S Jamall

Risk-Based Decisions, Inc., Sacramento, CA, USA

The paper “How difficult is the validation of clinical biomarkers?” by Jan Voskuil is timely in that as the author points out the burgeoning growth of biomarker assays particularly in chronic diseases and notably in cancer has led to the misinterpretation of both research and clinical data for all of the reasons pointed out by the author.

As correctly noted by the author, there are huge variations in sample collection, storage, preparation, assay used, antibodies involved, and analysis—all of which can lead to wide variability in results and confound the end user of such data. The author correctly points out steps that can be taken early on in the process that can minimize or preclude the accuracy of the results so obtained. There is a need for standardization. Furthermore, the statistical analyses of biomarker data need to be stipulated before the data are collected and not afterwards, using statistical software in a black-box manner to see what results might show statistical significance, even though the significance may be serendipitous or not clinically relevant (Zapf *et al.*, 2015). The slavish adherence to standard software packages by many physicians and biomedical scientists would be amusing if the wider implications of their innumeracy were not so dire (Partin *et al.*, 2013; Tuppin *et al.*, 2012).

In order for a biomarker to be useful, it must reflect a change in concentration in the media sampled with a change in disease status. It is frequently assumed that serum or blood are the best media for the study of biomarkers but because of the number of potentially confounding variables in serum or blood, tears and saliva, because they reflect intracellular fluids, might serve as better indicators of intracellular events long before these are reflected in the blood (Pieragostino *et al.* 2015; Salvisberg *et al.*, 2014).

I recommend acceptance of this very interesting paper for indexation.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: No competing interests were disclosed.

Referee Report 06 May 2015

doi:10.5256/f1000research.6861.r8494





Bjorn LDM Brücher

Theodor-Billroth-Academy, Munich, Germany

Thank you for inviting me to review the paper from Jan Voskuil.

I enjoyed reading the manuscript “*How difficult is the validation of clinical biomarkers?*” by Jan Voskuil and I agreed to review because of the following three aspects:

1. I appreciate the open review process of *F1000Research* and that such are published, as reviewers should stop hiding behind anonymity. There is – at least to me – no criterion justifying such if science increasingly wants transparency which we all know science needs.
2. The article provided is a must read. I would assume especially within the Biotech community but also scientists and clinicians should do so.
3. After I read the article the 2nd time, I decided to write a review but not in the usual way of reviewing a manuscript for a journal, because the necessary aspects have already been included in a comprehensive manner. My intent was including comments as well as additional aspects which may be of importance from the aspect of a clinician, surgeon and scientist for helping to see the subject *biomarker* from additional and different aspects.

General

The author provides important aspects by critically evaluating the use of biomarkers. This is necessary as the author reminds us about reality and wishes in science as well as in clinical practice. The different headlines are well thought through and chosen critically while questioning issues surrounding standardization. This is even more important as reports published for biomarkers under investigation are not standardized and make the same mistakes which had been made in the 70s and 80s in terms of tumor markers. Despite the necessity for being critical, it should not be viewed as being a synonymous with negative behavior. The author takes the responsibility by addressing major obstacles and missing data and that is without doubt highly appreciated and needed.

Biomarkers in diagnosis and treatment of diseases are measured characteristics and reflect a biological state of a disease. In terms of cancer, there is hope that biomarkers will provide a detection and screening tool for diagnosis, treatment with an influence on outcome orientated patient stratification as well as on predicting and monitoring multimodal treatment. The ideal biomarker is objectively measured in a comprehensible way independent of which laboratory it is investigated in, is easily measurable, cost-effective, and evaluated as an indicator for pathological and/or biological processes and of consistent value across differences in age, gender or ethnicity. Therefore, there is a necessity to remind us to not repeat history by including nearly any protein as a biomarker.

Where are we?

Many biomarkers are already declared by many companies to determine or diagnose a disease, although no data of half-life, metabolism or different interaction by different pathways are known nor provided. Again, it seems that history repeats itself as we get into the same discussions and situations as during the 70s and 80s in regard to tumor markers and cancer. Therefore, it is of importance that the author attempts to structure this theme into the headlines biological material, antibody choice, assay development, detection of cut-off values, statistics and project design.

Project Design

Logistically I would have thought that having the sub-headline Project Design earlier in the paper as this would indicate the direction.

Biological material

Of course there is hope that biomarkers help detecting a disease or serving as a screening tool for making a diagnosis. So far, it is not clear which biological material should be used and also it is not clear if this changes occur during different stages of diseases. Further, there is no standardization on how which biological material is stored and which variables influence the quality of the assay. There is another underestimated variable which needs to be taken into account and using cancer as an example may reveal further problems:

Cancer is not one disease and contains a heterogeneous set of dysfunctions such that the information available for biomarkers are also heterogeneous. This gets worse if we remind ourselves about the following: Igarashi, et al., observed in 93 specimens investigated for tumor microvessel density (MVD) and thymidine phosphorylase (dThdPase), that tumor cells strongly stained for TP were "...often observed as a rim in the periphery of the tumor nest"¹. On the other hand, biomarkers are not just expressed within tumor cell nests. Takebayashi, et al., revealed in 1998 that normal esophageal tissue showed a TP expression rate of 12.3% compared to 50.9% in the tumor cell area of 163 investigated resected ESCC, although the percentage rate of cells expressing TP was less than 5% in 85.9% of non-neoplastic tissues². Additionally even histomorphological tumor-negative lymph nodes (pN0) showed a TP expression rate of 27.9%. These examples illustrate that use of biomarkers needs a standardization in many aspects and by this scientists and clinicians need to be involved both trying to bring together the necessary aspects for future use of biomarkers, because the examples used could also mean that the gene expression is different in terms where a biopsy is taken as well as where apart of the specimen is cut for investigating biomarkers. Now, this may even be more complicated using fluids and under which condition they had been sampled, stored: was it during an operation? Do we know if drugs influenced those biomarkers of investigation with a short half-lives?

Methods (antibody choice, ..)

To my knowledge there is no standardization of methods in use for different biomarkers. Again, we had that during the 70s and 80s in terms of tumor markers in use and it took long to resolve. Is there a standard protocol in use for determination of microRNA? Is it not that some use RT-PCR, Northern blotting, oligo-based arrays, *in situ* hybridization, different assays (together with different samples)? Are these different techniques comparable?

Statistics (detection of cut-off values, statistics, ...)

Cut-offs can be determined of course. How many papers do you know in which a group determines all available necessary variables, such as sensitivity, specificity, positive predictive value, negative predictive value and overall accuracy? The decision-making for a cut-off is ultimately a clinical decision, as the clinician determines what is most important to know. For example, if we do not want to overlook a patient then it can be assumed that it makes sense having a high sensitivity in terms of using response. How is this in terms using a ROC analysis (receiving operator curve analysis)? Using a ROC analysis for better determination of a threshold, we need to compare the different measurements of a biomarker against a gold standard. As addressed below, it gets even more complicated: if there is a standard in use, but we know that it is not justified declaring it as a gold standard, what should we do?

Furthermore, as the author points out clearly how correct observations during the past resulted in wrong conclusions, which even increased dogma behavior in cholesterol levels associated with cardiovascular disease, lifestyle and obesity and also in terms of vitamin intake and health, as well as the somatic mutation theory being the cause for cancer.

An apple found in a car is not synonymous with the proof that apples grow in cars.

Critical thinking and re-thinking is continuously needed for excellence in science as well as for useful approaches in daily clinical work. Someone may enjoy reading the recent critical remarks about the wear and tear of guidelines³. I would argue that these views are also a must reading for the future implications of biomarkers.

Clinical Response Classification

This section is of course a huge one and has multiple aspects and due to this we cannot expect that all aspects are addressed by the author. However, this aspect is extremely important, as response is in daily clinical use and therefore I would assume, there is a must as a scientific and clinical reviewer addressing some points. One major question is: Can tumor response to therapy be predicted, thereby improving the selection of patients for cancer treatment?

This is a major problem now because if a biomarker is measured against the gold standard. But, is it justified declaring clinical response evaluation serving as gold-standard? The readers need to make up their own minds about the following which have recently been addressed as well⁴. The response classification is in use since 1971 since the publication by Miller⁵. What no-one wants to see is the fact, that this response classification is based on one experiment only which was conducted by Moertel and Hanley in 1976⁶.

Experiment

16 experienced oncologists (be aware that, in 1976, there was no definition of an oncologist and there wasn't one of an experienced oncologist either) in which they covered solid wooden spheres with a layer of rubber foam and placed them on a soft mattress. The colleagues had to measure the diameter of these spheres in a random order using rulers or calipers. The analysis showed that there was an error of 25% in the measurement of the size of identical spheres in 25% of the measurements, and that an error of at least 50% occurred in 6.8% of the measurements.

This means that the clinical response classification in use was based on a single experiment, and is still in use since some 35 years. Only some minor modifications were done: the US National Cancer Institute, together with the European Association for Research and Treatment of Cancer, proposed 'new response criteria' for solid tumors; a replacement of 2D measurement with measurement of one dimension was made⁷. Tumor response was defined as a decrease in the largest tumor diameter by 30%, which would translate into a 50% decrease for a spherical lesion⁸. However, no subsequent standardized of this recommendation was carried out, and 35 years after the primary experiment, no additional studies with a structured logistical way of accurate objective assessment of treatment response have been conducted or proposed. This opens an important question: If a biomarker is measured and analyzed according to the clinical response classification above, will this reflect biology as needed?

Immune Response

Another important variable to address contains immunological biomarkers in terms of response: Immunologists might just declare a response if immune-competent cells have been decreased and, possibly, without clinical signs of improvement of patient condition, or decrease of tumor size. But is it appropriate to declare a decrease of an immune-competent cell as a biomarker of any utility?

Disease stage

The 5 year survival rates for non-metastasized localized esophageal carcinoma according to the

American Cancer Society (ACS) in stage I (IA and IB) range between 71 and 57% while they drop in Stage IIA on 46% and range in Stage III (IIIA, IIIB, IIIC) between 20 and 9%⁹. Do we know if different disease stages in cancer are associated with different metabolism of different biomarkers?

Influence of paths which may influence quality of biomarker measurements

Let us take an example growth factor-beta (TGF-beta). It is known that inhibitors of this pathway, especially TGF- β 1, block the proliferation and trigger apoptosis in malignant as well as in benign tumor cells, but with increasing tumor growth development as TGF- β 1 resistance to targeted therapy develops¹⁰⁻¹². So, does this mean, that we are not aware that the function of a biomarker quality is independent from the disease, its stage how strong or weak the biomarker under investigation influence different paths? Do different situations in which biomarkers are measured mean different necessary views and if so, how can we judge those?

Another example: Epidermal growth factor receptor (EGFR; ErbB-1; HER1 in humans) is the cell-surface receptor for members of the epidermal growth factor family (EGF-family) of extracellular protein ligands¹³. EGFR plays a critical role in tumor progression by stimulating cell cycle progression, invasion, and metastasis¹⁴. Response measured by the EGFR-antibody Cetuximab in anticancer-treated patients does not correlate with the observed degree of EGFR expression in tumor tissue in patients with metastatic colorectal cancer^{15, 16}.

How many of the declarations of so-called breakthroughs in measuring a biomarker are justified? I have no doubt, that it is important to measure different biomarkers, but the marketing and promotional one should not be goal of scientists. This may be seen as an ethical aspect as well, but it is unfortunate but necessary as increasingly observations are reported and determined correctly within publications, but afterwards statements with journalists implicate having a breakthrough result, as nearly every scientific finding these days are declared as such, which from my perspective is very unfortunate. This inflationary marketing way should not be followed.

Conclusion

Taken the aspects reviewed above together, I repeat my statement: **the article by Jan Voskuil provided is a must read**. Many more aspects are necessary taking into account for future evaluation and standardization of biomarkers under investigation.

References

1. Igarashi M, Dhar DK, Kubota H, Yamamoto A, El-Assal O, Nagasue N: The prognostic significance of microvessel density and thymidine phosphorylase expression in squamous cell carcinoma of the esophagus. *Cancer*. 1998; **82** (7): 1225-1232 [PubMed Abstract](#)
2. Takebayashi Y, Akiyama S, Natsugoe S, Hokita S, Niwa K, Kitazono M, Sumizawa T, Tani A, Furukawa T, Aikou T: The expression of multidrug resistance protein in human gastrointestinal tract carcinomas. *Cancer*. 1998; **82** (4): 661-666 [PubMed Abstract](#)
3. Sitges-Serra A: Clinical guidelines at stake. *J Epidemiol Community Health*. 2014; **68** (10): 906-908 [PubMed Abstract](#) | [Publisher Full Text](#)
4. Brücher BLDM, Bilchik A, Nissan A, Avital I, Stojadinovic A: Tumor response criteria: are they appropriate?. *journal*. 2012; **8** (8): 903-906 [PubMed Abstract](#)
5. Miller AB, Hoogstraten B, Staquet M, Winkler A: Reporting results of cancer treatment. *Cancer*. 1981; **47** (1): 207-214 [PubMed Abstract](#)
6. Moertel CG, Hanley JA: The effect of measuring error on the results of therapeutic trials in advanced cancer. *Cancer*. 1976; **38** (1): 388-394 [PubMed Abstract](#)

7. Therasse P, Arbuck SG, Eisenhauer EA, Wanders J, Kaplan RS, Rubinstein L, Verweij J, Van Glabbeke M, van Oosterom AT, Christian MC, Gwyther SG: New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. *J Natl Cancer Inst.* 2000; **92** (3): 205-216 [PubMed Abstract](#)
8. James K, Eisenhauer E, Christian M, Terenziani M, Vena D, Muldal A, Therasse P: Measuring response in solid tumors: unidimensional versus bidimensional measurement. *J Natl Cancer Inst.* 1999; **17** (6): 523-528 [PubMed Abstract](#)
9. The American Cancer Society: Survival rates for cancer of the esophagus by stage. **Last Medical Review: 03/20/2014** (Last Revised: 03/02/2015). [Reference Source](#)
10. Akhurst RJ, Derynck R: TGF-beta signaling in cancer--a double-edged sword. *Trends Cell Biol.* 2001; **11** (11): S44-S51 [PubMed Abstract](#)
11. Roberts AB, Wakefield LM: The two faces of transforming growth factor beta in carcinogenesis. *Proc Natl Acad Sci U S A.* 2003; **100** (15): 8621-8623 [PubMed Abstract](#) | [Free Full Text](#) | [Publisher Full Text](#)
12. Benson RJ: Role of transforming growth factor beta in breast carcinogenesis. *Lancet Oncol.* 2004; **5** (4): 229-239 [PubMed Abstract](#)
13. Herbst RS: Review of epidermal growth factor receptor biology. *Int J Radiat Oncol Biol Phys.* 2004; **59** (2 Suppl): 21-26 [PubMed Abstract](#)
14. Lui VW1, Grandis JR: EGFR-mediated cell cycle regulation. *Anticancer Res.* 2002; **22** (1A): 1-11 [PubMed Abstract](#)
15. Cunningham D, Humblet Y, Siena S, Khayat D, Bleiberg H, Santoro A, Bets D, Mueser M, Harstrick A, Verslype C, Chau I, Van Cutsem E: Cetuximab monotherapy and cetuximab plus irinotecan in irinotecan-refractory metastatic colorectal cancer. *N Engl J Med.* 2004; **351** (4): 337-345 [PubMed Abstract](#)
16. Saltz LB, Meropol NJ, Loehrer PJ Sr, Needle MN, Kopit J, Mayer RJ: Phase II trial of cetuximab in patients with refractory colorectal cancer that expresses the epidermal growth factor receptor. *J Clin Oncol.* 2004; **22** (7): 1201-1208 [PubMed Abstract](#)

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: No competing interests were disclosed.
