

Markers for nutrition studies: review of criteria for the evaluation of markers

Jan de Vries · Jean-Michel Antoine · Tomasz Burzykowski ·
Alessandro Chiodini · Mike Gibney · Gunter Kuhnle ·
Agnès Méheust · Loek Pijls · Ian Rowland

Received: 19 October 2012 / Accepted: 18 June 2013 / Published online: 17 August 2013
© Springer-Verlag Berlin Heidelberg 2013

Abstract

Introduction Markers are important tools to assess the nutrition status and effects of nutrition interventions. There is currently insufficient consensus in nutrition sciences on how to evaluate markers, despite the need for properly evaluating them.

Objectives To identify the criteria for the evaluation of markers related to nutrition, health and disease and to propose generic criteria for evaluation.

Method The report on “Evaluation of Biomarker and Surrogate Endpoints in Chronic Disease” from the Institute of Medicine was the starting point for the literature search. Additionally, specific search strategies were developed for Pubmed.

Results In nutrition, no set of criteria or systematic approach to evaluate markers is currently available. There is a reliance on the medical area where statistical methods

have been developed to quantify the evaluation of markers. Even here, a systematic approach is lacking—markers are still evaluated on a case-by-case basis. The review of publications from the literature search resulted in a database with definitions, criteria for validity and the rationale behind the criteria. It was recognized that, in nutrition, a number of methodological aspects differ from medical research.

Conclusions The following criteria were identified as essential elements in the evaluation of markers: (1) the marker has a causal biological link with the endpoint, (2) there is a significant association between marker and endpoint in the target population, (3) marker changes consistently with the endpoint, e.g., in response to an intervention, and (4) change in the marker explains a substantial proportion of the change in the endpoint in response to the intervention.

J. de Vries
Jan de Vries, De Vries Nutrition Solutions, Reuvekamp 26,
7213-CE Gorssel, The Netherlands

J.-M. Antoine
Danone Research, Route Départementale 128,
91767 Palaiseau Cedex, France

T. Burzykowski
Center for Statistics, Hasselt University, Agoralaan,
3590 Diepenbeek, Belgium

T. Burzykowski
International Drug Development Institute (IDDI), Avenue
Provinciale, 30, 1340 Louvain-la-Neuve, Belgium

A. Chiodini (✉) · A. Méheust
ILSI Europe a.i.s.b.l., Avenue E. Mounier 83, Box 6,
1200 Brussels, Belgium
e-mail: publications@ilsieurope.be

M. Gibney
School of Agriculture and Food Science, University College
Dublin, Belfield Dublin 4, Ireland

G. Kuhnle · I. Rowland
Hugh Sinclair Human Nutrition Group, University of Reading,
Berkshire RG6 6AP, UK

Present Address:
A. Méheust
Oxylane Research, 4 Boulevard de Mons, 59650 Villeneuve
d’Ascq, France

L. Pijls
Nestlé Research Center, Vers-chez-les-Blanc, PO Box 44,
1000 Lausanne, Switzerland

Keywords Marker · Criteria · Evaluation · Nutrition research

Introduction

Markers are a cornerstone in research on bodily function, health and disease, including pharmaceutical and nutrition research. They may reflect or ‘mark’ an exposure, a status, a function or a risk factor. As such, they can be used as outcomes in studies on the effects of a food or food constituent on status, functions or risk factors. One of the conclusions of the European Commission-funded project PASSCLAIM, coordinated by ILSI Europe, was that there is a need for adequate markers in nutrition sciences [1].

Therefore, ILSI Europe decided to start an activity called Marker Initiative on Nutrition Research that aims to identify and review the criteria on the validation of markers. The initiative distinguishes 3 steps:

The first step combined two parallel approaches:

- One approach was identifying criteria for validity of markers by reviewing the available methodological/theoretical concept of marker validation as described in the literature. The outcomes of the literature research are illustrated in the present manuscript.
- The second approach combined different groups of experts from different fields of nutrition research as addressed by ILSI Europe. They identified criteria for validation of markers, based on the analysis of markers most commonly used in their field.

The second step was a workshop that aimed to achieve consensus on the criteria to evaluate markers. This workshop gathered nutrition scientists from and outside the ILSI Europe Marker Initiative to discuss the criteria identified in the preliminary work done in the two parallel approaches of the first step.

The final step (ongoing) will be the combination of the consensus set of criteria for evaluating markers in nutrition research, with guidance on how to use them. The application of criteria on markers used in different fields would enable the identification of markers that fulfil the criteria, and those for which future research is needed in order to meet the criteria.

Overall, this Marker Initiative aims at producing a practical toolkit for the evaluation of markers in nutritional research. As previously stated, this paper reviews how different research areas covering health-related issues, including nutrition, medical research, drug development and exposure assessment, evaluate whether a certain measurement is a meaningful marker of a certain aspect of current or future health or function. For this review, nutrition research includes research on food intake, food

and nutrient exposure, nutritional status and effects of nutrition on physiological functions. The aims of this activity are to make:

- An inventory of the criteria, found in the literature, for validation of markers of nutrition, health and disease.
- A proposal for the generic criteria for the validity of markers of nutrition, health and disease to be used in nutrition research.

Analytical validity is, although clearly an important criterion, outside the scope of this paper.

Definition framework

In reviewing the literature, a broad definition of the concept of marker is applied, comprising criteria for evaluating markers of dietary intake, dietary exposure, nutritional status and physiological functions. The inventory, as mentioned above, was set up to include measures on biological materials and information gathered, e.g., via questionnaires. Although most information retrieved in the study covers measures on biological materials, in this publication, it was chosen not to speak of “biomarkers” but of “markers”.

A great diversity in terminology and definitions in the field of marker assessment was recognized. Potischman defined a biomarker as “any biological specimen that is an indicator of nutritional status with respect to intake or metabolism of dietary constituents. It can be biochemical, functional or clinical index of status of an essential nutrient or another dietary constituent” [2]. Potischman proposed that markers of exposure be used to validate dietary measurement, or as a surrogate of dietary intake, or as an integrated measure of nutritional status, and that they should be evaluated according to precision, accuracy, sensitivity, specificity to the nutrient, and variability between subjects and temporality [2–4].

Although Potischman has proposed a definition for nutritional status markers, no definition framework could be discovered for the nutrition research area, covering dietary intake, nutritional status, nutrient exposure and effects of nutrition interventions on physiological and/or pathological outcomes. Therefore, for this review, it was decided to adhere to the definition of the Biomarker Definition Working Group (BDWG) that was proposed in 2001 and to comply with the definition framework of surrogate markers [5–7].

According to the BDWG, a marker is “a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes or pharmacologic responses to a therapeutic intervention” [5]. A clinical endpoint is a measure that captures information on “how a patient feels, functions or survives” [8]. In

medical research, three types of markers are distinguished: prognostic, predictive and surrogates [7]. Prognostic markers predict the likely course of disease for a patient, irrespective of the treatment. Predictive markers predict patient's response to treatment. Finally, a surrogate marker is used to replace a clinical endpoint to obtain a faster, simpler and/or less expensive evaluation of the efficacy of an experimental treatment and is, therefore, also called a surrogate endpoint. A marker is termed "validated" if it "has been demonstrated by robust statistical methods to be associated with a given clinical endpoint (prognostic markers), to predict the effect of a therapy on a clinical endpoint (predictive markers), or to be able to replace a clinical endpoint to assess the effects of a therapy (surrogate markers)" [7]. The statistical methods used for the validation depend on the type of the marker.

The glossary, present at the end of the manuscript, lists terms and definitions that are most often used in the discussions on marker evaluation.

Materials and methods

It was decided to use the report of the Institute of Medicine (IOM) on "Evaluation of Biomarkers and Surrogate Endpoints in Chronic Disease" as a starting point because a committee of the IOM was convened "to generate recommendations on the qualification process for biomarkers with a focus on risk biomarkers and surrogate endpoints in chronic disease" [9]. This committee was asked to work on this task because the Food and Drug Administration (FDA) is exploring the development of a framework for validating modifiable risk factors (markers) for chronic diseases, such as cancer, heart disease, diabetes and others that can be the subject of a health claim. The task of IOM was closely related to the purpose of ILSI Europe transversal 'Marker Initiative on Nutrition Research'. In addition to the IOM list of references, and as their process of literature retrieval was not described in the IOM report, a PubMed database search was performed to ensure that no critically important articles were missing. The search string(s) for PubMed included terms in the field of nutrition, medicine, pharmacology, toxicology, genomics, proteomics and metabolomics, and terms related to all types of markers, including questionnaires and other measures of non-biological nature. Publications on the validity of analytical methods to analyse biological or other samples were not included. Therefore, the following steps were adopted in order to perform the literature search (see Fig. 1):

The references in Table A-1 at page 254 of the IOM report were screened by title and summary/commentary, on relevance in the light of the aforementioned terminology.

Then, a first search in PubMed was performed (see search string 1, time limits from 1948 to 2011).

Search string 1 (date of search: 4 August 2011, numbers of hits are mentioned in brackets):

1. (biomarker* or "risk factor*" or "surrogate endpoint*" or "predictive marker*").ti,ab. (321.978)
2. biological markers (119.856)
3. 1 or 2 (416.813)
4. valid*.ti,ab. (296.207)
5. criteri*.ti,ab. (323.000)
6. 3 and 4 and 5 (1.466)
7. (health or physiological or function or nutrition).ti,ab. (2.271.384)
8. 6 and 7 (372)
9. limit 8 to humans (336)

The results of this search were compared on authors and title with the literature list of the IOM report [9]. Because of the very little overlap observed in the literature retrieved (one paper only between the results of search 1 and the references in the IOM report), a second PubMed search (Search string 2) was performed (Note: due to the large numbers of publications found in the literature, the second search was limited to 2010 and 2011):

Search string 2 (date of search: 8 August 2011, numbers of hits are mentioned in brackets):

1. (biomarker* or "risk factor*" or "surrogate endpoint*" or "predictive marker*").ti,ab. (322193)
2. Biological Markers (119.856)
3. 1 or 2 (417.028)
4. criteria*.ti,ab. (323.149)
5. (valid* or evaluat*).ti,ab. (1.988.377)
6. 3 and 4 and 5 (6.835)
7. limit 6 to humans (6.299)
8. (health* or physiological* or function* or nutrition* or CVD or CHD or cardiovasc* or cancer* or oncolo* or tumor* or tumour* or HIV or AIDS or genetic*).ti,ab. (5.456.419)
9. exp neoplasms/or exp digestive system diseases/or exp respiratory tract diseases/or exp cardiovascular diseases/or exp "nutritional and metabolic diseases"/ or exp endocrine system diseases/or exp immune system diseases/(6.297.638)
10. 8 or 9 (9.388.890)
11. 7 and 10 (5.070) (and limited to 2010–2011)

One scientist reviewed the titles and abstracts of every record retrieved on (1) aim of the paper and (2) discussion on criteria to validate markers. Full articles were considered to be relevant if the information in the title and the abstract indicated that the publication discusses criteria on the validation, qualification and/or evaluation of (bio) markers. Articles were discarded when describing a study

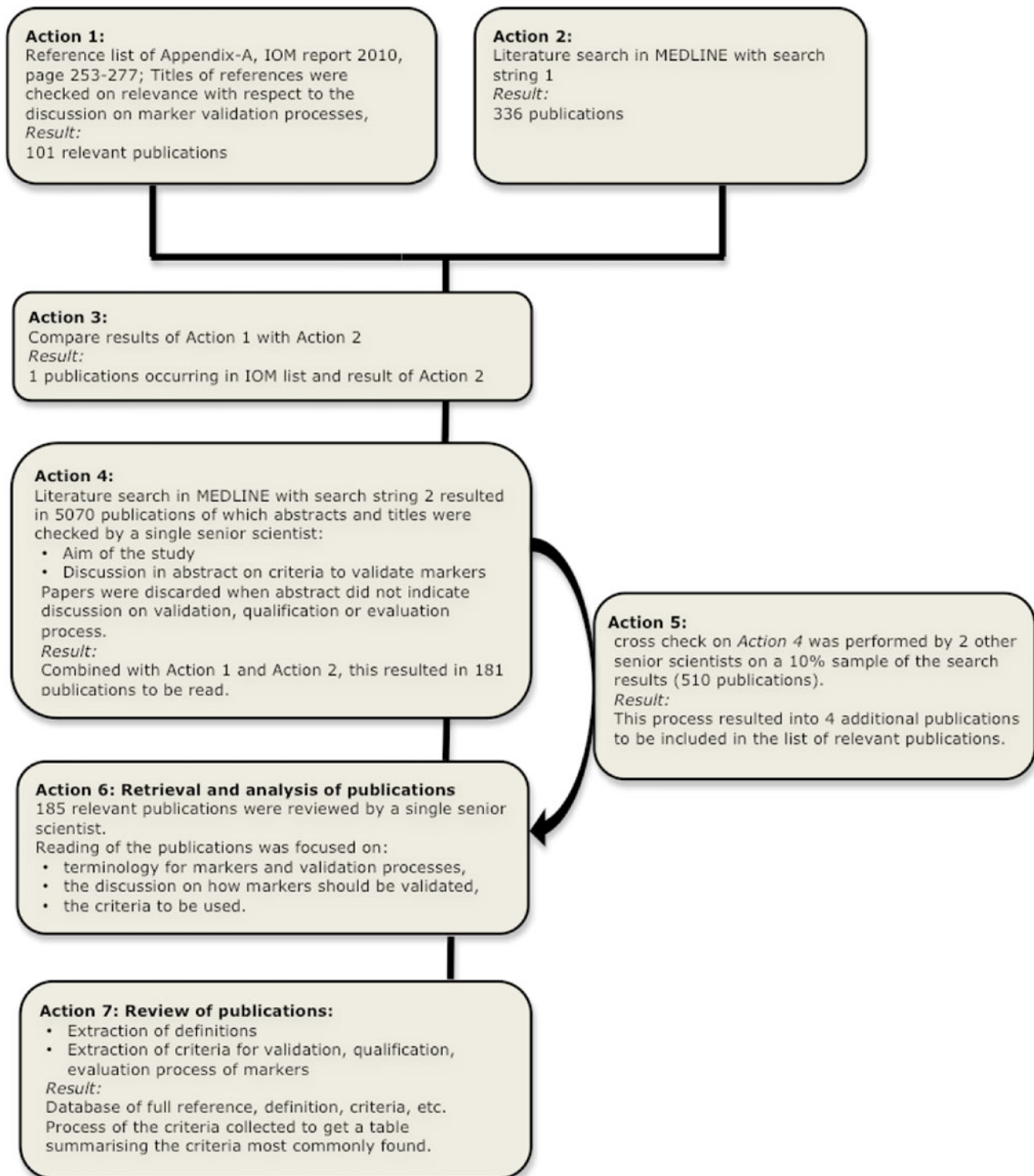


Fig. 1 Flowchart of the search procedure

to describe/assess a specific marker, but the abstract did not indicate criteria for a validation, qualification and/or an evaluation process.

Two other scientists followed the same procedure on a random 10 % sample of the search results, in order to

cross-check that most relevant articles were picked-up from the search, validating the process of retrieving literature. This resulted in 4 additional papers on a check of ± 500 titles and abstracts to be added to the list of relevant publications. On the basis of this result, the review

produced was considered adequate for the purpose of the assessment.

The reading of the full text publications focused on validation processes, discussion on how markers should be validated, qualified and/or evaluated, and what criteria should be used for this process. The terminology used in these publications was also captured. This resulted in a database with information about the criteria used for the qualification and evaluation of markers, corresponding rationale of the markers—if presented—and definitions used for the terminology extracted from the selected publications.

The database was constructed with the following elements:

- Full reference: ability to go back to the original publications.
- Definitions: create an overview of the definitions used in the area of marker validation.
- Criteria for validity (what): create an overview of the criteria that are proposed and used for validation, qualification and evaluation of markers.
- Rationale behind the criteria: the “why” behind the criteria that are proposed in the validation, qualification and evaluation processes.
- Methodology: types of scientific methods that contribute to the validation, qualification and evaluation of markers.
- Context of the paper: physiological and/or technical context (e.g. cancer research, bone metabolism).

It was then checked whether publications retrieved with the word ‘biomarker’ in the search string were different from those retrieved using the word ‘marker’, in order to keep a broader perspective in this review. It came out that 18 articles were different from those proposed when using ‘biomarker’ and these were not related to diet or nutrition.

Results

Publications resulting from the search strategy

The PubMed search string 1 resulted in 336 publications and screening of the references in the IOM report [9] yielded 101 publications. Because only one paper was identified in both searches, a search string 2 was developed, which resulted in several thousands of publications related in some way to the validation of markers. To achieve a number of publications that could reasonably be reviewed, search string 2 was limited to 2010–2011. The total search process yielded 1,184 publications of which, concluded from the titles and the abstracts, 185 were considered to be relevant enough for full text review.

The survey was unable to provide one publication proposing a set of criteria or a systematic approach to assess or qualify markers to be used in nutrition research either for dietary exposure or for research on status, functions or health consequences. For drug development, a clear framework of types of markers has been developed in the last 2 decades [5–7]. This framework focused on surrogate markers. Most publications referred to criteria to assess markers in drug development or the curing of diseases and, consequently on surrogate markers, with the exception of markers for exposure. The assessment of markers in drug development on their meaningfulness is mostly, if not completely, performed on a case-by-case basis, due to a lack of consensus on the quantification in the evaluation process. Dietary exposure data have been discussed in relation to their validity; however, these publications lack a discussion on the criteria for validation [2–4].

Properties required for evaluating surrogate markers

The 1987 meeting of the Biometrics Society started the discussion on how to establish valid ‘surrogates’ [10]. In his landmark publication, Prentice laid out a definition for a surrogate endpoint, being “a response variable for which a test of the null hypothesis of no relationship to the treatment groups under comparison is also a valid test of the corresponding null hypothesis based on the true endpoint” and also presented criteria for a surrogate endpoint (see Table 1) [11]. These criteria have been the basis for further discussion on criteria and the development of statistical evaluation methods. The review process of the selected literature resulted in a database on the definitions and theoretical criteria to assess the meaningfulness of markers in scientific processes. Table 1 summarises the theoretical criteria extracted from the selected publications. These criteria were put in relation to prognostic markers, predictive markers and surrogate endpoints.

Developments in statistical methods on the evaluation of surrogate markers and surrogate endpoints

The Prentice criteria [11] can be evaluated by using data from a single clinical trial in which the surrogate and the true endpoints have been observed. However, the fourth criterion requires proving a null hypothesis of no effect of treatment after adjustment for the surrogate. Recognizing the difficulty of this approach, Freedman et al. [19] proposed to replace this fourth criterion by the estimation of the proportion of the treatment effect on the true endpoint mediated by the surrogate (Proportion Explained, PE). According to this proposal, for a valid surrogate endpoint, the PE should be close to 1.

Table 1 Summary of the properties required for different types of (bio) markers, extracted from the selected literature

Publication	Criteria identified in publications	Proposed properties		
		Prognostic	Predictive	Surrogate endpoint
Prentice [11]	Treatment has a significant impact on the surrogate endpoint			X
	Treatment has a significant impact on the true endpoint			X
	The surrogate endpoint has a significant impact on the true endpoint			X
	The full effect of treatment upon the true endpoint is captured by the surrogate			X
Lippman [12]	Differential expression of the biomarker in normal and high-risk tissue	X	X	X
	Ability to analyse the biomarker in small tissue specimens	X	X	X
	Quantitative degree or pattern correlating the biomarker with the stage of disease; and Availability of preclinical or early clinical data supporting modulation by study agent	X	X	X
Boissel [13]	Show biological plausibility	X	X	X
	Show significant association between biomarker and clinical endpoint in the target population	X	X	X
	Show that the biomarker changes consistently in response to therapy, preferably in a predictable and dose-dependent fashion that agrees with the known mechanism of action of the therapeutic intervention		X	X
	Show that changes in the biomarker with treatment explain a substantial proportion of the efficacy on the clinical endpoint		X	X
Rolan [14]	Analysis of data on the statistical properties of the surrogate relating to (1) reproducibility, (2) accuracy and (3) bias of measurement	X	X	X
	The analysis may include (1) observer error and (2) observer variability	X	X	X
	The marker must have a close correlation with the clinical endpoint in a stable environment	X	X	X
	The marker must be sensitive to interventions		X	X
	The marker must share a causal mechanism with the clinical endpoint		X	X
Temple [15]	Qualitative surrogates: Both surrogate and clinical endpoint move in the same direction			X
	Quantitative surrogates: The surrogate fully captures the effect of the treatment on the outcome			X
	Evidence that the surrogate biochemical, anatomic and/or morphological, or pathophysiologic endpoint is on the causal pathway to the adverse outcome or is a regular finding associated with that outcome and is plausibly related to a common causal factor			X
	A history of successful intervention with a pharmacologically-related agent or (even better) a range of pharmacologically-unrelated agents		X	X
Lesko [16]	Biological plausibility should provide a mechanistic basis for using the surrogate endpoint			X
	Epidemiological or survey studies of the natural history of the disease should support surrogate status by establishing the statistical relationship between the biomarker and the clinical endpoint under basal conditions (30)			X
	Adequate and well-controlled clinical trials should provide an estimate of the expected benefit in terms of clinical endpoints that can be derived mathematically or mechanistically from an estimate of the change in the potential surrogate endpoint. Ideally, an appropriate dose–response or exposure–response relationship would be established as supplemental support for surrogate status			X
	The analysis should include a consideration of potential adverse reactions unrelated to the clinical endpoints predicted by the surrogate endpoint			X
	An exposure–response model should be developed that mathematically describes and predicts relationships between drug doses or plasma concentrations and surrogate endpoints and clinical outcomes. Verification of these predictions is important	X	X	X
	The development and validation of biomarkers and surrogate endpoints should be built into the drug development process, beginning with the preclinical phase		X	X
	It may be helpful to conduct a meta-analysis of multiple clinical trials to look across and within studies to determine the consistency of effects following interventions with various drug classes and within different stages of the disease			X

Table 1 continued

Publication	Criteria identified in publications	Proposed properties		
		Prognostic	Predictive	Surrogate endpoint
Mayeux [17]	Content validity, which shows the degree to which a biomarker reflects the biological phenomenon studied	X	X	X
	Construct validity, which pertains to other relevant characteristics of the disease or trait, for example, other biomarkers or disease manifestations	X	X	X
	Criterion validity, which shows the extent to which the biomarker correlates with the specific disease and is usually measured by sensitivity, specificity and predictive power	X	X	X
	Monitorability of the marker (development of the marker)	X	X	X
Sistare [18]	Specific fit-for-purpose need, the initial goal and the study designs must be synchronized and focused and not broadened in an attempt to address all deficiencies in current conventional biomarkers	X	X	X
	Strengths and limitations of (bio) markers	X	X	X
	Initially establish claims describing the utility of biomarkers	X	X	X
	Full capacity of a biomarker: It is essential that the study protocol generates samples that mimic the most pressing problem that the biomarker is intended to address (study design)	X	X	X

Buyse and Molenberghs showed that PE is ill-defined because, mathematically, it is not proportional. They proposed to replace it by two measures: the relative effect (RE) and the adjusted association (AA). RE is the ratio of the effects of treatment upon the true and the surrogate endpoints. RE could be used to construct a model allowing a prediction of the effect of treatment on the true endpoint, based on the effect on the surrogate endpoint [20].

In a single-trial setting, the use of RE for the prediction purposes requiring adopting assumptions that cannot be tested. This limitation does not apply in a multiple trials setting, i.e., in the so-called meta-analytic approach to the validation of surrogate endpoints [21]. In a nutshell, the approach consists of using data from multiple clinical trials in which both surrogate and true endpoints have been observed. Trial-specific treatment effects are estimated and analysed by using a linear model. The model predicts the treatment effect on the true endpoint, with the effect on the surrogate as an input. The precision of the prediction is quantified by the coefficient of determination, R^2 . The closer value of R^2 to 1, the more predictive the model and the higher the validity of a surrogate endpoint.

To interpret the clinical meaningfulness of the values of R^2 , the concept of a surrogate threshold effect (STE) has been developed [22]. STE is the lowest treatment effect upon the surrogate endpoint that predicts a significant treatment effect upon the true endpoint. If STE is small enough within the usual range of treatment, then the surrogate endpoint may be deemed useful. This is because a small STE indicates narrow prediction limits for the treatment effect on the true endpoint. Hence, the surrogate may “reasonably predict” the effect of treatment on the

true endpoint, as postulated in the definition of the Biomarkers Definitions Working Group [9].

Proposals for systematic approaches to evaluate markers

There is substantial literature discussing the qualitative framework for the evaluation of markers in the medical area. Some attempts have been made to provide them from a qualitative to a quantitative framework for marker evaluation. Below, a list is provided with proposals for quantitative evaluation procedures that we have identified in our literature research:

- Ransohoff 2007 (summarised in Table 1 in the publication) [23] and Lassere 2007 (ranking system with different variables) [24]
- Boussein 2008 (four-step approach) [25]
- Altar 2008 (summarised in Table 1 in the publication, systematic framework developed by the Pharmaceutical Research and Manufacturers of America) [26]
- Wagner 2008 (from exploration to surrogacy) [6]
- Liang 2009 (4 criterions to fulfil) [27]
- Hlatky 2009 (6 phases for the evaluation of novel risk markers) [28]
- Early Detection Research Network (EDRN) 2009 (5 phases, phasing of marker development is developed as a tool to clinically detect cancer before symptoms appear and to identify people at risk; see Fig. 2) [29]
- Doust 2010 (3 phases: exploration, qualification and evaluation of clinical utility) [30]
- IQWiG report 2011 (flow diagram to determine the validity of markers in oncology) [31].

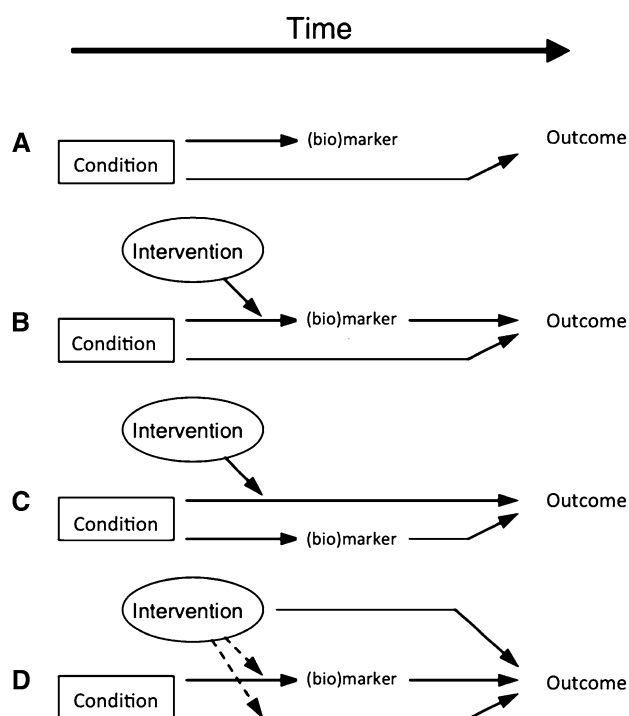


Fig. 2 Reasons for failure of surrogate markers in predicting endpoints. Adapted from Fleming 1996 [60]. **a** The surrogate is not in the causal pathway of the process. **b** Of several causal pathways, the intervention affects only the pathway mediated through the surrogate. **c** The surrogate is not on the pathway of the intervention's effect or is insensitive to its effect. **d** The intervention has mechanisms of action independent of the process that results in the outcome. Dotted lines = mechanisms of action that might exist

The above-mentioned schemes are based on sets of criteria for validation, qualification and evaluation processes that are summarized in Table 1. Lassere [24], Altar [26] and the EDNRN [29] have proposed grading and/or weighing evidence in their evaluation process of markers. Wagner [6] proposed an evaluation process tailored to the degree of certainty required in different contexts of use, which is illustrated with an inverted pyramid of marker qualification.

Discussion

This review summarises the developments of marker evaluation in the medical and nutrition sciences. Because in nutrition sciences no definition framework is available according to this review, it was decided to use the definitions from the medical sciences. The majority of the papers in the medical area on evaluation of markers cover prognostic and predictive markers and surrogate endpoints.

The discussion will first focus on the small number of findings in the nutrition sciences. Then, the concept of association versus causality is discussed in the light of commonly found criteria for the evaluation. The discussion

continues to elaborate on the development of statistical methods for surrogate markers and the proposals that have been launched for a systematic approach in the evaluation of markers. The applicability of the medical approach on evaluation of markers in nutrition sciences will be considered, taking into account the potential pitfalls in the use of surrogate markers as identified from medical studies. Finally, the applicability of the medical evaluation system approaches will be discussed with respect to single and multiple marker systems in nutrition sciences.

Findings in nutrition sciences

This review of existing recent literature on criteria for the evaluation of markers in nutrition, and life science at large, generated little data dedicated to nutrition research. In the recent past, several activities of ILSI covered the validity of exposure markers in nutrition [2–4, 32–47]. These papers discuss the validity of specific markers as such in relation to dietary intake/exposure but did not intend to create consensus on criteria to be used for the validation of markers. Most publications were dealing with surrogate markers and/or markers for drug development.

This review indicates that no specific definition framework exists for nutrition research. Only Potischman has proposed a definition for a nutritional status marker [2]. Neither has there been any organized discussion on the development of criteria to validate markers used in nutrition research, including intake markers, status markers, exposure markers and markers of physiological effects on nutrition intervention. It can be concluded that nutrition research strongly relies on the development of criteria for the assessment of markers in the medical area. For drug development, a consensus exists on the framework of types of surrogate markers, which is based on the definitions of prognostic marker, predictive marker and surrogate endpoints [5–7].

Association or causality: criteria for evaluation

The need for a proper evaluation process for surrogate markers is based on the search for causality between the surrogate marker and the endpoint. In 1965, Sir Bradford Hill proposed nine characteristics that relate to causality [48]. They do not prove causality, but the more the criteria apply, the more likely that an observed association is of a causal nature.

1. *Strength of the association*: causation is supported if the relative risk due to the exposure is very large.
2. *Consistency of observed association*: has it been repeatedly observed by different persons, in different places, circumstances and times?
3. *Specificity of the association*: causation is supported if an exposure appears to cause only a specific effect.

4. *Temporal relationship of the association*: causation is supported if exposure precedes the effect.
5. *Biological gradients*: a clear dose–response curve admits of a simple explanation and puts the observation in a clearer light.
6. *Biologically plausible*: it will be helpful if the causation is biologically plausible; according to Hill, this is a feature that cannot be demanded.
7. *Coherence*: the cause-and-effect relationship should not seriously conflict with the generally known facts of the natural history of the disease.
8. *Experiment*: occasionally, evidence that reducing or removing the exposure decreases the effect can be used to draw conclusions about causality.
9. *Analogy*: in some circumstances, comparison between weaker evidence of causation between an exposure and its effect and strong evidence of causality between another exposure and its similar effect is appropriate.

These nine characteristics should not be considered a checklist but provide an approach for studying association before we cry causation [48]. According to Biesalski, the nine characteristics are an approach to the interpretation of the data sets that provide a way to compensate for data gaps [49]. In this review, a set of commonly used criteria to evaluate the meaningfulness of a marker to its intended endpoint is identified. These have a lot in common with the characteristics of Sir Bradford Hill in 1965. This is logical, as a certain measure being a marker of a certain endpoint may be based on a causal relationship between this marker and the endpoint. Commonalities observed are as follows:

- The surrogate marker must share a causal biological mechanism with the endpoint (biological plausibility).
- Significant association between surrogate marker and endpoint in the target population.
- Surrogate marker changes consistently with the endpoint in response to the intervention.
- Change in the surrogate marker explains a substantial proportion of the change in the endpoint in response to the intervention.

Statistical considerations on surrogate markers and surrogate endpoints

The criteria described in Table 1 are more characteristics than criteria because they lack quantification. These qualitative characteristics offer a conceptual framework for the evaluation of surrogate markers. The quantification of a conceptual framework is necessary before the characteristics can be used in an evaluation process. Characteristics become criteria and are fully operational when a decision is taken on their thresholds. For the drug development area,

several methods have been developed to quantify characteristics. This process started with the landmark publication of Prentice who proposed that (1) a treatment should have a significant impact on the surrogate endpoint, (2) a treatment should have a significant impact on the true endpoint, (3) the surrogate endpoint should have a significant impact on the true endpoint, and (4) the full effect of treatment upon the true endpoint is captured by the surrogate [11].

Among many proposals, only few were of any practical value. It is clear that the optimal surrogate marker, capturing the full effect of an intervention upon the endpoint, is a rather hypothetical concept. Therefore, Prentice's definition [19] is of a limited practical value because it is not possible to confirm the null hypothesis of no effect based on the observed data. Thus, evidence from clinical trials with non-significant treatment effects cannot be used, even though such trials may be consistent with a desirable relationship between both the surrogate and the endpoint. The operational criteria, derived by Prentice, aimed at addressing this issue. However, Buyse and Molenberghs showed that the last two of Prentice's criteria were necessary and sufficient for binary responses, but not in general [20]. More importantly, the fourth criterion also requires proving a null hypothesis. Hence, although the criteria seem appealing and are easy to apply, they are not a valid approach.

PE, proposed by Freedman et al. [19], attempted to rectify the problems related to the use of Prentice's definition. However, although practically feasible, the use of PE does not offer a valid approach either. This is because it is an ill-defined measure [20]. PE could be replaced by the use of RE and AA [20]; especially, the latter measure is of interest, as it could predict the effect of treatment on the true endpoint. However, in a single-trial setting, the use of RE requires making untestable assumptions. Hence, though estimation of RE is practically feasible, the use of the measure is not very appealing. This limitation is removed by the extension of the concept of RE and AA to the case of multiple trials. The meta-analytic approach to the validation of surrogate endpoints is feasible from a practical point of view [21]. In fact, it has been used to validate a range of candidate endpoints in oncology [50–55].

The key measure in the meta-analytic approach is the trial-level R^2 . The desired value of R^2 is 1, but it is not possible to attain it in practice. Hence, values as close to 1 as possible are sought. However, there is no universally accepted threshold value for R^2 to deem a candidate surrogate endpoint acceptable. This is partly due to the fact that it is difficult to interpret the values of R^2 . STE addresses this problem [22]. The choice of a threshold value for STE may be easier to define than for R^2 , because the former measure is expressed on a clinically meaningful

scale. Nevertheless, there are no universal rules for choosing the threshold, as the choice depends on the disease and treatments under study.

Given the discrepancies between the pharmaceutical and nutrition frameworks, it seems plausible to assume that additional research on the validation criteria, assessing the discrepancies, might be warranted. To this aim, the use of the causal inference approach could be investigated. The approach has recently become the focus of intensive research within the pharmaceutical clinical trials context. While it has not led to any well-established set of criteria yet, it can potentially offer a solution for the situations when, for instance, the validation can only use data from epidemiological studies.

Surrogate marker evaluation systems

Statistical evaluation could probably be done objectively, although evidence for a putative surrogate cannot be solely based on the statistical validity of a surrogate [9]. An objective evaluation for the biological plausibility, which we propose to include in any evaluation process, and the correlation between both, might be even more difficult. Combined evaluation of different types of evidence is not easy. Lassere's scheme is a step towards a combined evaluation of different types of evidence in this direction [56]. She constructed a systematic approach for the evaluation of markers containing the following domains of criteria:

- Study design (ranking from 0 to 5)
- Target outcome (idem)
- Statistical evaluation (idem)
- Penalties (−1 to −3 reduction points in case some information is missing).

Its usability can be questioned, but one of the values of the attempt lies in that it illustrates many difficulties an overall evaluation scheme illustrates, e.g., what domains to include, how to weigh the different domains, and what thresholds to use. The quest for a framework “regulating” this issue, at least at this point and within the drug development domain, has not led to any commonly accepted procedure or scheme [9]. However, a standardised method to present marker validation data, i.e., to make it easier to compare different markers and to assess their validity for specific applications, would be of great support for the nutrition sciences.

Applicability in the nutrition research

A major question in the evaluation process is “what are the issues to be taken into account when evaluating a marker for its intended purpose”? Markers are effective only to the

degree that they are used in the appropriate context. Drug use is mostly employed for looking at disease treatment, whereas in nutrition, there is a need for markers of reduction in disease risk and markers for health—effects that are only partly comparable with the drug approach, and which may need to be evaluated differently [18, 49, 57–59]. The IOM committee recommends that the same degree of scientific rigour should be used for the evaluation of markers across regulatory areas, whether they are proposed for use in the arenas of drugs, medical devices, biologics, or foods and dietary supplements [9].

The assessment of a marker needs evidence-based science and a well-defined endpoint. However, as already mentioned, it is not easy to define what an endpoint is, and what a marker is because most endpoints can be considered as marker for a more integrative function. For instance, nutrition interventions most often result in multiple effects on sets of physiological processes instead of the improvement of one specific target. The definition of the endpoint may therefore depend on the purpose of the study, i.e., the context of use.

Secondly, does every marker need to fulfil all criteria, and does each criterion contribute to the same extent in qualifying a marker? Literature is mainly focused towards surrogate markers to explore the efficacy of drugs, and we were unable to find a rationale for adequate weighing of each criterion to fit its purpose. For example, plausibility may be required as a golden standard and may not be mandatory for assessing a risk factor when there is a very good correlation between a marker and a risk. Practicality is a key criterion for epidemiology, while it is less important for research, as well as inter-individual variability that can be compensated to a great extent in epidemiology. Accuracy is a key factor to assess a status, while precision is a key factor to monitor changes.

The third point is that scientific data rarely provide 100 % certainty: correlation tends towards 1, repeatability tends towards 100 %, precision is fair, accuracy must be good, but what is the threshold to deny that a criterion is fulfilled? The meaningfulness of a marker, or the certainty of such meaningfulness, is often considered to be a dichotomous phenomenon: it is meaningful or not for a specific purpose. In drug development, this requirement has proved unrealistic [7]. In fact, such dichotomy exists only if it is decided to put a ‘breaking point’ (X) somewhere on a continuous scale of meaningfulness. The totality of data will point somewhere on a scale ranging from *very unlikely* to *very likely* meaningful. Below a certain breaking point X, there is reasonable doubt as to whether the measure is meaningful, and above that point, one is beyond reasonable doubt. In the latter case, it is concluded that the parameter and, therefore the effects, on such a parameter are meaningful; if not, the relevance of the parameter remains

hypothetical. This approach, however, raises several questions, such as where to place X on the scale of “meaningfulness”, and how do we decide on this? Is this a scientific or a ‘risk–benefit management’ decision? Some argue that X may be relatively low for communications about health and diet in non-commercial settings and high for such communications in a commercial setting. Others may say that X should depend on the presence and magnitude of any potential risk involved. If no risk, then X can be lower; but how much lower? Should X be higher if the marker is related to an intervention to cure a disease, compared to an intervention that is intended to prevent disease onset [60] or to improve mental or physical performance?

Potential problems with the use of surrogate markers

Although the use of surrogate markers is promising in nutrition research, we need to understand that not all surrogate markers may be fit for every purpose. A study, in which, e.g., a margarine containing sitostanol ester decreases both LDL-cholesterol levels and the incidence of cardiovascular diseases, does not prove that lowering LDL-cholesterol itself by the intervention decreases the incidence of and mortality due to cardiovascular diseases. The reason is that such a study does not exclude the possibility that the intervention had—at least—two separate effects, one being the effect on cholesterol and one on cardiovascular disease. One cannot conclude whether the effect on cardiovascular disease was fully, partially, or in no way caused by the effect on cholesterol. Fleming [61, 62] elegantly explains the potential reasons for failure of markers in predicting endpoints, a concept that was also described by Boissel [13] (see Fig. 2).

Longer time intervals between the measurement of the surrogate marker and the outcome, and increasing complexity of the effect of the intervention on the surrogate marker and the outcome may result in more reasons for failure of the surrogate marker in predicting the outcome. Reasons for failure of surrogate markers in predicting endpoints may further increase when using multiple markers in study designs.

Multiple markers

Sets of markers that emerge from omics techniques are recent examples of the use of multiple markers, tools aiming at managing the limitations of single markers. Over the last decade, there has been a significant development in the biology of genomics, transcriptomics, proteomics and metabolomics, collectively known as omics. In the field of nutrition, there has been a great expectation that this field would deliver significant advances in marker discovery. By

and large, the reality has fallen far short of expectations [63]. Most of the interest in this area has been in the discovery of diagnostic and prognostic markers for disease. In human nutrition, this expectation that the omics area would reveal nutrient-sensitive markers also prevailed in the early part of the last decade, but again, very little progress has been made. There are a number of methodological problems facing nutrition research, which do not apply to medical research. Perhaps the largest difference is that the development of a disease involves a significant level of metabolic perturbation. Thus, when comparing cases to controls, there are marked metabolic differences that can be readily detected with any one of the omics technologies. Moreover, following successful therapeutic intervention, the reversal of this significant metabolic perturbation is also readily detected with omics technologies. In human nutrition, the effects of variation in nutrient intake are subtle compared to those in medicine or in pharmacology, and moreover, their effects are very diverse [58]. Whereas the cholesterol lowering drugs, such as the statins, have very specific effects, changing dietary fats to alter blood cholesterol impinges on many other areas of metabolism. The agreement on the significance of a level of change in gene expression was informally agreed-upon but not validated by any scientific rationale, as far as we were able to ascertain.

A second methodological problem is that for transcriptomics, most available material is from peripheral blood mononuclear cells that are heavily dominated by genes involved in the immune system. Again, for metabolomics, blood tends to be dominated by endogenous metabolites, such as lipid traffic from adipose tissue to liver or amino acid traffic from muscle to liver. Subtle changes in multiple metabolites due to diet are unlikely to be detected by standard metabolomics. Thus, there is a growing interest in the use of targeted metabolomics where, for example, a total lipid profile is determined. Urinary metabolomics tends to be dominated by the excretion of plant phytochemicals, and there is interest in this as a potential marker of food intake [64].

In human nutrition, markers are generally used either to identify long-term dietary patterns (e.g. red blood cell folate or cholesteryl ester eicosapentaenoic acid) or to identify a nutrient-sensitive factor, which if altered indicates an increased or decreased risk of a disease (elevated plasma low-density lipoprotein cholesterol and heart disease, elevated plasma glucose and type 2 diabetes). At present, there is little evidence that the omics technologies will add to this in the short term, but it is likely that in the course of time and with targeted omics technologies, some improvements will be made. One area that is attracting attention is the area of stress testing individuals, most often with a test meal. Thus, whereas inter-individual differences

might be small in the fasting state, the real variability is seen in the postprandial state [65]. This will pose challenges in that most of the epidemiological data linking diet and chronic disease use fasting values and very few link challenge tests to disease development. Moreover, we need to first understand if and to what extent, e.g., postprandial variability is relevant for current or future health or physiological function.

Conclusions

The discussion on the meaningfulness of markers started only late 1980s and primarily focuses on drug development. This paper reviewed the state of the art for processes for evaluating markers in general for medicine and nutrition.

Definition framework

In the nutrition sciences, a diversity of markers is used, based on experiences and tradition without a proper framework of definitions or criteria to evaluate these markers for their intended purpose, not for markers that estimate exposure and intake, nor for markers that estimate future outcomes related to nutrition interventions.

Nutrition science mostly relies on the medical research area with respect to the development of and evaluation of (bio) markers. In the medical sciences, there are some qualitative criteria. However, markers for medical purposes are still evaluated on a case-by-case basis. Improving uniformity in terminology in nutrition research probably would improve systematic assessment of markers.

Proposed criteria

Most common theoretical criteria in the assessment of the meaningfulness of surrogate markers are as follows:

1. The marker must share a causal biological mechanism with the endpoint (biological plausibility).
2. Significant association between marker and endpoint in the target population.
3. Marker changes consistently with the endpoint in response to the intervention.
4. Change in the marker explains a substantial proportion of the change in the endpoint in response to the intervention.

These criteria can better be considered as concepts of thinking because they lack quantification.

Due to the limited discussion on criteria to validate dietary intake markers, dietary exposure markers and nutritional status markers, no criteria could be proposed for these types of markers.

Statistical developments

To statistically evaluate surrogate markers and surrogate endpoints, several methods have been developed. The key method used at this moment is the meta-analytical approach that leads to an estimate that results in a trial-level R^2 with an optimal value of 1. The value of 1 is not obtainable in practice, and there is no universal rule for choosing a threshold value.

The adoption of markers in nutrition research needs a combination of biological and statistical considerations [7]. The criteria from the medical sciences are not yet easily applicable to evaluate markers in nutrition.

Systematic approach for the assessment of validity

In general, there is a lack of systematic approach. In its recent overview “Evaluation of biomarkers and surrogate endpoints in chronic disease”, IOM does not come up with a system for the evaluation of markers and concludes that, currently, the evaluation of markers is not based on uniform standards or processes but rather on the gradual development of consensus in the scientific community [9]. Some authors have proposed a systematic approach to assess the meaningfulness of markers, but these proposals have not been implemented into consensual systematic review systems for marker assessment.

Gaps in applying approaches from medicine and drug development into nutritional context

- Evidence for markers/surrogates in drug development comes from clinical trials. In nutrition research, this type of data is lacking. Many nutrition intervention studies do not use hard endpoints but surrogate markers as endpoints. With this approach, the outcomes of a nutrition intervention study rely on the validation of those surrogate markers in relation to their hard endpoints that are provided by medical research.
- In drug development, the target population generally comprises patients suffering from a disease, and the target intervention is the treatment of the disease. In nutrition, the target is prevention of the disease in the population at large. The latter is somehow similar to, e.g., the area of infectious diseases with vaccination as the intervention. Marker/surrogate research is much less advanced for vaccination trials than, e.g., for oncological ones.
- Exceptions can be made for enteral and parenteral nutrition, which are products dedicated to patients. However, usually these interventions are supportive to medical treatment but are not considered to be medical treatment as such.

- In drug development, there is usually a well-established set of endpoints one might consider for the evaluation of a treatment’s efficacy. In nutrition, the choice and/or definition of an endpoint are often less clear.
- There are proposals for systematic approaches for the evaluation of markers in drug development/medicine, but not in nutrition.
- No consensus criteria in the nutrition research area exist.

Acknowledgments This work was conducted by an expert group of the European branch of the International Life Sciences Institute (ILSI Europe). The expert group received funding from the ILSI Europe Functional Foods Task Force. Industry members of this task force are listed on the ILSI Europe website at www.ilsieurope.eu. For further information about ILSI Europe, please email info@ilsieurope.be or call +32-2-7710014. The opinions expressed herein and the conclusions of this publication are those of the authors and do not necessarily represent the views of ILSI Europe, nor those of its member companies.

Conflict of interest Alessandro Chiodini is employed by ILSI Europe. None of the other authors have a conflict of interest to report.

Glossary

Marker (including biomarker) A characteristic that can be objectively measured and evaluated as an indicator of daily dietary intake, daily habits, biological and/or pathological processes or responses to interventions that may affect daily habits, biological and/or pathological processes [5].

Biological marker (biomarker) A characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes or pharmacologic responses to a therapeutic intervention [5]. *This definition is now commonly used although variations on this definition are not uncommon* [16, 17, 66–68].

Prognostic marker Biomarker that forecasts the likely course of disease irrespective of the treatment [7].

Predictive marker A marker that forecasts the likely response to a specific treatment [7].

Surrogate endpoint A marker that is intended to substitute for a clinical endpoint. A surrogate endpoint is expected to predict clinical benefit (or harm or lack of benefit or harm) based on epidemiologic, therapeutic, pathophysiologic or other scientific evidence [5]. *According to*

Buyse, surrogate end points require data demonstrating both that the surrogate is prognostic of the true end point, and that the effect of treatment on the surrogate correlates with that of the true end point [7].

Clinical endpoint A characteristic or variable that reflects how a patient feels, functions or survives [5].

Endpoint A characteristic or variable that reflects any physiological and pathological function of an individual. (This is a definition brought up by the expert group of Approach A.)

Target engagement marker A marker that occurs early in a pathophysiologic cascade and informs on physical or biological interactions with the molecular target of the drug and, thus, assesses how hard a drug hits the target [6].

Disease related marker A marker that occurs late in the pathophysiologic cascade and is linked to clinical benefit and, thus, assesses the effect of a drug on a particular disease [6].

Validation A process that demonstrates by robust statistical methods that a marker is associated with a given clinical endpoint (prognostic markers), predicts the effect of a therapy on a clinical endpoint (predictive markers) or is able to replace a clinical endpoint to assess the effects of a therapy (surrogate markers) [7].

Qualification A process that assesses the available evidence on associations between the marker and the (clinical) endpoint, including data showing effects of interventions on both the marker and the outcomes [9].

Evaluation A utilization phase that builds a contextual analysis on the specific use proposed and the applicability of the available evidence for this use. This must include a determination of whether the validation and qualification conducted provide sufficient support for the use proposed [9].

References

1. Aggett PJ, Antoine JM, Asp NG, Bellisle F, Contor L, Cummings JH, Howlett J, Müller DJG, Persin C, Pijls LTJ et al (2005) Passclaim, consensus on criteria. *Eur J Nutr* 44:1–30
2. Potischman N, Freudenheim JL (2003) Biomarkers of nutritional exposure and nutritional status: an overview. *J Nutr* 133:S3873–S3874
3. Potischman N (2003) Biologic and methodologic issues for nutritional biomarkers. *J Nutr* 133:3875S–3880S
4. Marshall JR (2003) Methodologic and statistical considerations regarding use of biomarkers of nutritional exposure in epidemiology. *J Nutr* 133:S3881–S3887
5. Biomarkers definitions working group (2001) Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin Pharmacol Ther* 69:89–95
6. Wagner JA (2008) Strategic approach to fit-for-purpose biomarkers in drug development. *Annu Rev Pharmacol Toxicol* 48:631–651
7. Buyse M, Sargent DJ, Grothey A, Matheson A, de Gramont A (2010) Biomarkers and surrogate end points—the challenge of statistical validation. *Nat Rev Clin Oncol* 7:309–317
8. Temple RJ (1995) A regulatory authority's opinion about surrogate endpoints. In: Nimmo WS, Tucker GT (eds) *Clinical measurement in drug evaluation*. John Wiley and Sons, New York, pp 3–22
9. National research council (2010) *Evaluation of biomarkers and surrogate endpoints in chronic disease*. The National Academies Press, Washington
10. D'Agostino RB, Campbell MJ, Greenhouse JB (2006) Surrogate markers: back to the future. *Stat Med* 25:181–182
11. Prentice RL (1989) Surrogate endpoints in clinical trials: definition and operational criteria. *Stat Med* 8:431–440
12. Lippman SM, Lee JS, Lotan R, Hittelman W, Wargovich MJ, Hong WK (1990) Biomarkers as intermediate end points in chemoprevention trials. *J Nat Cancer Inst* 82:555–560
13. Boissel JP, Collet JP, Moleur P, Haugh M (1992) Surrogate endpoints: a basis for a rational approach. *Eur J Clin Pharmacol* 43:235–244
14. Rolan P (1997) The contribution of clinical pharmacology surrogates and models to drug development—a critical appraisal. *Br J Clin Pharmacol* 44:219–225
15. Temple R (1999) Are surrogate markers adequate to assess cardiovascular disease drugs? *JAMA* 282:790–795
16. Lesko LJ, Atkinson AJ (2001) Use of biomarkers and surrogate endpoints in drug development and regulatory decision making: criteria, validation, strategies. *Annu Rev Pharmacol Toxicol* 41:347–366
17. Mayeux R (2004) Biomarkers: potential uses and limitations. *NeuroRx* 1:182–188
18. Sistare FD, Dieterle F, Troth S et al (2010) Towards consensus practices to qualify safety biomarkers for use in early drug development. *Nat Biotechnol* 28:446–454
19. Freedman LS, Schatzkin A (1992) Sample size for studying intermediate endpoints within intervention trials or observational studies. *Am J Epidemiol* 136:1148
20. Buyse M, Molenberghs G (1998) Criteria for the validation of surrogate endpoints in randomized experiments. *Biometrics* 54:1014–1029
21. Buyse M, Molenberghs G, Burzykowski T, Renard D, Geys H (2000) The validation of surrogate endpoints in meta-analyses of randomized experiments. *Biostatistics* 1:49–67
22. Burzykowski T, Buyse M (2006) Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation. *Pharmaceutical Statist* 5:173–186
23. Ransohoff DF (2007) How to improve reliability and efficiency of research about molecular markers: roles of phases, guidelines, and study design. *J Clin Epidemiol* 60:1205–1219
24. Lassere MN, Johnson KR, Boers M et al (2007) Definitions and validation criteria for biomarkers and surrogate endpoints: development and testing of a quantitative hierarchical levels of evidence schema. *J Rheumatol* 34:607–615
25. Bouxsein ML, Delmas PD, Bouxsein ML, Delmas PD (2008) Considerations for development of surrogate endpoints for anti-fracture efficacy of new treatments in osteoporosis: a perspective. *J Bone Mineral Res* 23:1155–1167
26. Altar CA, Amakye D, Bounos D et al (2008) A prototypical process for creating evidentiary standards for biomarkers and diagnostics. *Clin Pharmacol Ther* 83:368–371
27. Liang MH, Simard JF, Costenbader K et al (2009) Methodologic issues in the validation of putative biomarkers and surrogate endpoints in treatment evaluation for systemic lupus erythematosus. *Endocr Metab Immune Disord Drug Targets* 9:108–112
28. Hlatky MA, Greenland P, Arnett DK et al (2009) Criteria for evaluation of novel markers of cardiovascular risk: a scientific statement from the American Heart Association. *Circulation* 119:2408–2416
29. National Cancer Institute. Division of Cancer Prevention (2008) *The Early Detection Research Network*. http://edm.nci.nih.gov/docs/progress-reports/edm_4th-report_200801.pdf. Accessed 6 September 2011
30. Doust J, Doust J (2010) Qualification versus validation of biomarkers. *Scand J Clin Lab Invest* 242:40–43
31. Validity of surrogate endpoints in oncology (2011) *Validity of surrogate endpoints in oncology*. <http://www.iqwig.de>. Accessed 21 September 2011
32. Blanck HM, Bowman BA, Cooper GR et al (2003) Laboratory issues: use of nutritional biomarkers. *J Nutr* 133:888S–894S
33. Mayne ST (2003) Antioxidant nutrients and chronic disease: use of biomarkers of exposure and oxidative stress status in epidemiologic research. *J Nutr* 133:S3933–S3940
34. Mason JB (2003) Biomarkers of nutrient exposure and status in one-carbon (methyl) metabolism. *J Nutr* 133:S3941–S3947
35. Lampe JW (2003) Isoflavonoid and lignan phytoestrogens as dietary biomarkers. *J Nutr* 133:S3956–S3964
36. Livingstone MBE, Black AE (2003) Markers of the validity of reported energy intake. *J Nutr* 133:895S–920S
37. Bingham SA (2003) Urine nitrogen as a biomarker for the validation of dietary protein intake. *J Nutr* 133:921S–924S
38. Arab L (2003) Biomarkers of fat and fatty acid intake. *J Nutr* 133:925S–932S
39. Hambidge M (2003) Biomarkers of trace mineral intake and status. *J Nutr* 133:948S–955S
40. Foreword Willett W (2009) The validity of dietary assessment methods for use in epidemiologic studies. *Br J Nutr* 102:S1–S2
41. Henríquez-Sánchez P, Sánchez-Villegas A, Doreste-Alonso J et al (2009) Dietary assessment methods for micronutrient intake: a systematic review on vitamins. *Br J Nutr* 102:S10–S37
42. Serra-Majem L, Pfrimer K, Doreste-Alonso J et al (2009) Dietary assessment methods for intakes of iron, calcium, selenium, zinc and iodine. *Br J Nutr* 102:S38–S55
43. Øverby NC, Serra-Majem L, Andersen LF (2009) Dietary assessment methods on n-3 fatty acid intake: a systematic review. *Br J Nutr* 102:S56–S63
44. Ortiz-Andrellucchi A, Doreste-Alonso J, Henríquez-Sánchez P, Cetin I, Serra-Majem L (2009) Dietary assessment methods for micronutrient intake in pregnant women: a systematic review. *Br J Nutr* 102:S64–S86

45. Ortiz-Andrellucchi A, Henríquez-Sánchez P, Sánchez-Villegas A et al (2009) Dietary assessment methods for micronutrient intake in infants, children and adolescents: a systematic review. *Br J Nutr* 102:S87–S117
46. Ortiz-Andrellucchi A, Sánchez-Villegas A, Doreste-Alonso J et al (2009) Dietary assessment methods for micronutrient intake in elderly people: a systematic review. *Br J Nutr* 102:S118–S149
47. Hooper L, Ashton K, Harvey LJ et al (2009) Assessing potential biomarkers of micronutrient status by using a systematic review methodology: methods. *Am J Clin Nutr* 89(suppl):S1953–S1959
48. Hill AB (1965) The Environment and Disease: association or Causation? *Proc R Soc Med* 58:295–300
49. Biesalski HK, J Aggett P, Anton R, et al (2010) Scientific substantiation of health claims: Evidence-based nutrition. 26th Hohenheim Consensus Conference, September 11, 2010 Nutrition 27:S1-S20
50. Burzykowski T, Molenberghs G, Buyse M et al (2001) Validation of surrogate end points in multiple randomized clinical trials with failure time end points. *J Roy Stat Soc* 50:405–422
51. Collette L, Burzykowski T, Carroll KJ et al (2005) Is prostate-specific antigen a valid surrogate end point for survival in hormonally treated patients with metastatic prostate cancer? Joint research of the European Organisation for Research and Treatment of Cancer, the Limburgs Universitair Centrum, and Astra-Zeneca Pharmaceuticals. *J Clin Oncol* 23:6139–6148
52. Sargent DJ, Wieand HS, Haller DG et al (2005) Disease-free survival versus overall survival as a primary end point for adjuvant colon cancer studies: individual patient data from 20,898 patients on 18 randomized trials. *J Clin Oncol* 23:8664–8670
53. Buyse M, Burzykowski T, Carroll K et al (2007) Progression-free survival is a surrogate for survival in advanced colorectal cancer. *J Clin Oncol* 25:5218–5224
54. Burzykowski T, Buyse M, Piccart-Gebhart MJ et al (2008) Evaluation of tumor response, disease control, progression-free survival, and time to progression as potential surrogate end points in metastatic breast cancer. *J Clin Oncol* 26:1987–1992
55. Buyse ME, Squifflet P, Allard SE et al. (2008) Leukemia-free survival as a surrogate end point for overall survival in the evaluation of maintenance therapy for patients with acute myeloid leukemia in complete remission. *Hematologica* 93(suppl 1):209
56. Lassere MN (2008) The Biomarker-Surrogacy Evaluation Schema: a review of the biomarker-surrogate literature and a proposal for a criterion-based, quantitative, multidimensional hierarchical levels of evidence schema for evaluating the status of biomarkers as surrogate endpoints. *Stat Methods Med Res* 17:303–340
57. Lee JW, Devanarayan V, Barrett YC et al (2006) Fit-for-purpose method development and validation for successful biomarker measurement. *Pharm Res* 23:312–328
58. Blumberg J, Heaney RP, Huncharek M et al (2010) Evidence-based criteria in the nutritional context. *Nutr Rev* 68:478–484
59. Raiten DJ, Namasté S, Brabin B et al (2011) Executive summary—biomarkers of nutrition for development: building a consensus. *Am J Clin Nutr* 94:633S–650S
60. Gallagher AM, Meijer GW, Richardson DP et al (2011) A standardised approach towards PROving the efficacy of foods and food constituents for health CLAIMs (PROCLAIM): providing guidance. *Br J Nutr* 106:S16–S28
61. Fleming TR, DeMets DL (1996) Surrogate end points in clinical trials: are we being misled? *Ann Intern Med* 125:605–613
62. Fleming TR (2005) Surrogate endpoints and FDA's accelerated approval process. *Health Aff (Millwood)* 24:67–78
63. Abu-Asab MS, Chaouchi M, Alesci S et al (2011) Biomarkers in the age of omics: time for a systems biology approach. *OMICS* 15:105–112
64. O'Sullivan A, Gibney MJ, Brennan L (2011) Dietary intake patterns are reflected in metabolomic profiles: potential role in dietary assessment studies. *Am J Clin Nutr* 93:314–321
65. van Ommen B, Keijer J, Heil SG, Kaput J (2009) Challenging homeostasis to define biomarkers for nutrition related health. *Mol Nutr Food Res* 53:795–804
66. Rasnake CM, Trumbo PR, Heinonen TM (2008) Surrogate endpoints and emerging surrogate endpoints for risk reduction of cardiovascular disease. *Nutr Rev* 66:76–81
67. Goodsaid F, Frueh F (2007) Biomarker qualification pilot process at the US Food and Drug Administration. *AAPS J* 9:E105–E108
68. Molenberghs G, Burzykowski T, Alonso A, Buyse M (2004) A perspective on surrogate endpoints in controlled clinical trials. *Stat Methods Med Res* 13(3):177–206