Biochimica et Biophysica Acta 1844 (2014) 874–883

Contents lists available at ScienceDirect

Biochimica et Biophysica Acta

journal homepage: www.elsevier.com/locate/bbapap

Review

Biomarker research with prospective study designs for the early detection of cancer ☆

B. Pesch [a,*,1], T. Brüning [a,b,1], G. Johnen [a], S. Casjens [a], N. Bonberg [b], D. Taeger [a], A. Müller [b], D.G. Weber [a], T. Behrens [a,b]

[a] Institute for Prevention and Occupational Medicine of the German Social Accident Insurance (IPA), Ruhr University Bochum, Germany
[b] Protein Research Unit Ruhr within Europe (PURE), Ruhr University Bochum, Germany

ARTICLE INFO

Article history:
Received 30 April 2013
Received in revised form 6 November 2013
Accepted 10 December 2013
Available online 17 December 2013

Keywords:
Bias
Diagnostic marker
Prospective study
Screening
Study design
Validation

ABSTRACT

This article describes the principles of marker research with prospective studies along with examples for diagnostic tumor markers. A plethora of biomarkers have been claimed as useful for the early detection of cancer. However, disappointingly few biomarkers were approved for the detection of unrecognized disease, and even approved markers may lack a sound validation phase. Prospective studies aimed at the early detection of cancer are costly and long-lasting and therefore the bottleneck in marker research. They enroll a large number of clinically asymptomatic subjects and follow-up on incident cases. As invasive procedures cannot be applied to collect tissue samples from the target organ, biomarkers can only be determined in easily accessible body fluids. Marker levels increase during cancer development, with samples collected closer to the occurrence of symptoms or a clinical diagnosis being more informative than earlier samples. Only prospective designs allow the serial collection of pre-diagnostic samples. Their storage in a biobank upgrades cohort studies to serve for both, marker discovery and validation. Population-based cohort studies, which may collect a wealth of data, are commonly conducted with just one baseline investigation lacking serial samples. However, they can provide valuable information about factors that influence the marker level. Screening programs can be employed to archive serial samples but require significant efforts to collect samples and auxiliary data for marker research. Randomized controlled trials have the highest level of evidence in assessing a biomarker's benefit against usual care and present the most stringent design for the validation of promising markers as well as for the discovery of new markers. In summary, all kinds of prospective studies can benefit from a biobank as they can serve as a platform for biomarker research. This article is part of a Special Issue entitled: Biomarkers: A Proteomic Challenge.

1. Introduction

Biomarkers can be used for the early detection of disease in asymptomatic subjects, the diagnosis of the disease in tissue samples, and the evaluation of response to therapy in patients. Here, we refer to diagnostic markers for the non-invasive detection of early stages of cancer to initiate treatment at curable stages. Expert groups like the U.S. Preventive Services Task Force (USPSTF) and the Early Detection Research Network (EDRN) detail the standards for the early detection of cancer. Diagnostic markers aimed for screening should also comply with established WHO standards [77].

Although a plethora of candidate markers have been suggested from discovery studies, very few positive findings could be reproduced, and a disappointingly small number of molecular assays were approved for application in clinical practice. For example, HPV-DNA testing was recommended for cervix cancer screening and fecal immunochemical testing or the detection of cancer cell DNA in stool for colorectal cancer (CRC) screening [68]. Widely used markers like the prostate-specific antigen (PSA) are still subject of debate regarding its usefulness for screening [29]. The nuclear matrix protein 22 (NMP22) was approved for the early detection of bladder cancer by the U.S. Food and Drug Administration (FDA) prior to sound validation. Subsequently, prospective studies like UroScreen failed to show acceptable specificity of NMP22 in asymptomatic subjects [28].

The detection of early stages of cancer is based on the paradigm that the disease develops along advancing deviations from the normal status. Diagnostic markers need to indicate the development of the cancer before the occurrence of clinical symptoms, which requires the analysis of markers in pre-diagnostic samples in easily accessible proxy tissues. Invasive access to the target organ is not possible in

☆ This article is part of a Special Issue entitled: Biomarkers: A Proteomic Challenge.

* Corresponding author at: Institute for Prevention and Occupational Medicine of the German Social Accident Insurance, Institute of the Ruhr University Bochum (IPA), Bürkle-de-la-Camp-Platz 1, 44789 Bochum, Germany. Tel.: +49 234 302 4536; fax: +49 234 302 4505.
E-mail address: pesch@ipa-dguv.de (B. Pesch).

[1] Both authors contributed equally to the manuscript.

1570-9639/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.bbapap.2013.12.007

asymptomatic subjects. However, the majority of studies aimed at detecting new markers in the discovery phase are performed with cross-sectional comparisons of symptomatic cases and hospitalized patients as "convenience" controls. Inattentiveness to methodological issues of study design and performance has been repeatedly claimed as a major reason for false-positive findings in marker research [60–62]. This likely leads to biased performance measures and lacking data about the cancer-predictive value of markers in pre-diagnostic samples, which can only be estimated reliably in longitudinal studies, recruiting clinically asymptomatic subjects at baseline. Prospective studies are large, long-lasting and, hence, costly and represent the bottleneck in marker research [13]. However, they can be upgraded with a biobank to serve as a platform in marker research to overcome the shortcomings of cross-sectional studies.

Many excellent reviews have detailed the phases of marker research (e.g. [47,61]). In a previous review we explained more general epidemiological principles in marker research [8]. Here we address the question how well-designed cohort studies in the general population, screening studies in clinical settings, and other prospective designs can be employed for marker research [62]. We briefly introduce the two major steps, marker discovery and validation, together with performance measures and typical failures of prospective studies.

2. Prospective studies in the framework of marker research

2.1. Marker discovery

Marker discovery starts with the identification of candidate markers from larger sets of tested biomarkers that may be associated with the disease under study. An initial step is the investigation of molecular signatures in the target organ comparing affected and healthy tissue parts. Then markers are assessed in tissue samples from diseased and non-diseased subjects. Because access to the target tissue from non-diseased subjects is hardly feasible with invasive methods, body fluids may serve as proxy tissue. A marker for the early detection of small or pre-malignant lesions requires sensitive assays to detect exfoliated cells or cancer-related molecules in the blood or other body fluids. Finally, a robust assay has to be developed for the validation of the diagnostic marker.

The discovery step employs classification algorithms to identify differences in marker levels by disease status. Various statistical methods are applied to search for the best-discriminating markers from large datasets of "omics" data, with examples given in [31,70]. The number of markers frequently outnumbers the number of investigated subjects. This may lead to overfitting as a major problem in marker discovery [6]. A reproduction of the best-discriminating markers found in a single study requires independent investigations to exclude candidates detected by chance. The discovery of markers is oftentimes data-driven and not based on biological hypotheses. A stronger translation of biological considerations into marker discovery would improve the identification of promising markers [40].

At first glance, the comparison of diseased with non-diseased subjects does not require profound epidemiological knowledge. However, the dilemma of false-positive findings of candidate markers is associated with methodological problems of the design and conduct of simple cross-sectional studies [8,61,62]. Major limitations are the lack of pre-diagnostic samples from cases, the confounding of samples from cases by therapeutic measures, the recruitment of hospitalized patients as "convenience controls", insufficient attention to pre-analytical factors, and batch problems from an unbalanced collection and handling of samples from cases and controls. Sensitivity estimates derived from symptomatic cases may be over-optimistic, because marker levels are likely higher in advanced than in early stages of the disease [17]. The marker specificity may be underestimated if hospital controls are enrolled, particularly if they suffer from diseases of the target organ. For example, a sensitivity of 48% and a specificity of 91.5% was reported for methylated *SEPT9* in detecting colorectal cancer in a longitudinal study of asymptomatic subjects compared with 90% sensitivity and 88% specificity in a case–control study [17,74].

In line with EDRN and other experts we recommend prospective studies already for marker discovery as detailed for nested case–control comparisons, if pre-diagnostic samples are stored in a biobank [47]. All steps of marker research in body fluids can be performed with prospective designs as depicted in Fig. 1. Fig. 2 shows an example for such an integrated approach that is based on the design of a randomized controlled trial (RCT).

2.2. Marker validation

Validation refers to the evaluation of the performance of diagnostic markers in prospective studies to detect the disease earlier than by symptoms and to assess the overall benefit of a diagnostic marker against "usual care". Here we refer to usual care as the clinical examinations commonly applied to detect a disease, for example the cytological investigation of the cell pellet in a urine sample in screening for bladder cancer [55]. Please note that some authors also refer to "validation" for the verification of results in the discovery phase. Validation studies for diagnostic markers enroll "clinically" healthy subjects. Repeated examinations are necessary to determine the marker in pre-diagnostic samples at various follow-ups, because marker levels are associated with the tumor growth as shown in Fig. 3. The number of incident cases that occur prospectively in the cohort is the limiting factor. For example, among 7941 participants enrolled in a routine study aimed at detecting colorectal cancer, 53 cases were observed in a single screen [17]. The study population should be at excess risk to ensure the occurrence of a sufficient number of cases. A major risk factor is age, but sex, smoking status, occupation, or other factors associated with the future development of the disease may also define the target population.

In the first step of the validation process, an elevated marker level is not yet considered for a decision on diagnostic workup. After having collected cases during follow-up, a case–control comparison may be nested into the cohort. The marker is determined in pre-diagnostic samples from diseased subjects and a subsample of controls who have not developed the disease, but not in the entire cohort. Such a nested design was applied to validate diagnostic markers for ovarian cancer within the framework of the Prostate, Lung, Colorectal, and Ovarian Cancer (PLCO) trial [79].

In the second step, a positive marker test is used to initiate further diagnostic workup. Due to the invasiveness of most diagnostic methods, only markers with good performance, particularly with a high specificity in previous phases, should be used for a clinical decision. For example, the UroVysion test was validated in the UroScreen study for the early detection of bladder cancer [5,12].

The most rigorous design to assess the overall benefit of diagnostic markers in comparison to usual care is a randomized controlled trial in which the participants are randomly assigned to different trial arms. This allows a comparison of the disease-specific mortality and the rate of side effects in the marker arm vs. the usual-care arm (Fig. 2).

An assay suitable for clinical practice or screening should be robust against test-modifying factors and of moderate costs when compared to standard methods. A detailed cost-benefit analysis was provided for the early detection of bladder cancer with tumor markers [44]. BladderChek is a point-of-care assay providing in-office results for an elevated NMP22 concentration at reasonable costs [44]. However, the marker level in spot urine is influenced by a variety of factors related to the urine sample like density or cellularity [53]. By contrast, UroVysion is an expensive FISH assay that can only be performed in specialized laboratories. FISH generates complex data about chromosomal instability in atypical cells that are also targeted in the much cheaper urine cytology [12]. High costs are also associated with the Epi proColon assay that measures aberrant methylation in circulating *SEPT9* DNA compared with cost-efficient fecal tests [5,17]. Costly assays