INTRODUCTION

Diagnostic and prognostic markers are extensively used in clinical practice. Tumor markers are substances produced by cancer or normal cells; that are produced at much higher levels in cancerous conditions. These substances can be found in the blood, urine, stool, tumor tissue, or other tissues or bodily fluids of some patients with cancer. Thus far, more than 20 different tumor markers have been characterized and are in clinical use. Common examples are the use of Prostate-specific antigen (PSA) or Ki-67 for screening of prostate cancer, CA 15-3 for ovarian cancer, and CA 125 for ovarian cancer. Development of diagnostic and prognostic markers has been discussed by the statistical community and different approaches have been suggested. The USA Food and Drug Administration (FDA) provides regulatory guidance for submission of tumor marker premarket notifications (1996).10

METHODS

We conducted a general literature search in order to review and describe the state-of-the-art in the design of studies for the development and validation of diagnostic and prognostic markers (General approach). We compared our findings with the FDA pragmatic point of view (FDA guidance)12. We finally conducted a search in the FDAs In Vitro Diagnostic Product Database13 to find the characteristics of the studies that were granted FDA approval to commercialize a medical device related to a tumor marker (FDA approved). All published reports for approved markers in prostate, bladder, testicular, and ovarian cancers were selected. The review focused on the following topics: marker development, sample size estimation, data selection, cutoff selection, and validation.

RESULTS

The search provided 10 reports published between 1999 and 2014: prostate cancer (n=5), bladder cancer (n=2), ovarian cancer (n=1), and testicular cancer (n=2). All studies were diagnostic studies. Development plan

General approach: Development and validation is generally divided in three broad phases: development, retrospective validation, and prospective validation. (Table 1)

FDA guidance: Prospective studies are preferred. For retrospective studies, use an appropriate statistical sampling plan to ensure representativeness and unbiasedness of the data.

FDA approved: Only three out of ten reports described the use of prospective designs; four reports included retrospective designs and the other three combined data from prospective and retrospective studies. Sampling methods were not described in all of them.

Sample size

General approach: The sample size must be adequate to ensure that the test’s accuracy is estimated with good precision.13 Three main formulas are proposed for sample size calculation. (Table 2)

FDA guidance: Sample size must have sufficient statistical power to determine clinical importance. It is better to overestimate sample size than to underestimate it, although it should be realistically obtainable. “Low” sample size can be considered for diseases with low prevalence.

FDA approved: Sample size estimation methods were never reported.

Alternatively, sample size was very large among all studies.

Representative data

General approach: Include samples from individuals for which the use of the device is intended. Include samples from individuals with diseases or conditions that may cause false positive or false negative results with the device.6

FDA guidance: Same requirements as the general approach. Use data from at least three centers; at least one from the USA.

FDA approved: All studies were performed in a specifically selected population. Only three out of ten reports contained confounding populations for the analysis, including populations with different cancer types, gender and other diseases.

Cutoff selection

General approach: The cutoff is generally selected to obtain either a large sensitivity, a large specificity or a tradeoff between these two parameters depending on the consequences of the diagnosis. Conduct subgroup analysis to assess the performance of the test on different population groups.12

FDA guidance: Describe how the cutoff point was determined and what performance characteristics it was intended to produce. Identify different cutoffs to define the equivocal zones, otherwise justify its absence. Describe the population characteristics for those groups that the cutoff is applicable or any differences among them.

FDA approved: Eight of the reviewed reports specified a cutoff value. From these, one was established without justification and five to achieve specific sensitivity or specificity. The use of an equivocal zone was never discussed. Test performance was described for different subgroups in only one report.

Validation

General approach: Prediction models require internal and external validation.

FDA guidance: There are no explicit requirements in the 1999 guidance about model validation. Selection of cutoff and model performance is determined during development and no independent validation is requested.

FDA approved: No internal, neither external validation was reported in any of the reports.

DISCUSSION

Our review suggests that the statistical community is far more precise and demanding than the FDA at the time to develop diagnostic and prognostic tumor markers. Although FDA is very aware about the importance of marker validation, e.g., guidance on development of markers for acute cellular rejection (ACR) in heart transplant; marker validation is particularly overlooked in the development of tumor markers as reported in the published reports. Our conclusions are limited by the difficulty of information available in the published reports.

Table 1. Study phases for prognostic and diagnostic marker development

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Exploratory</td>
<td>Phase I</td>
<td>Phase II</td>
<td>Phase I</td>
<td>Phase I “Exploratory”</td>
</tr>
<tr>
<td>Retrospective validation</td>
<td>Phase II</td>
<td>Phase IIb</td>
<td>Phase II-III</td>
<td>Phase II “Challenge”</td>
</tr>
<tr>
<td>Prospective validation</td>
<td>Phase IIc</td>
<td>Phase III</td>
<td>Phase IV</td>
<td>Phase III “Clinical”</td>
</tr>
<tr>
<td>Impact study (clinical trial)</td>
<td>Phase IV</td>
<td>Phase III</td>
<td>Phase V</td>
<td>-</td>
</tr>
<tr>
<td>Normalization (healthy subjects)</td>
<td>-</td>
<td>Phase I</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 2. Sample size calculation formulas

Machten 2009

Number of subjects necessary to show that a given probability, $P$, differs from a target value $P_{target}$. Given a significance level $\alpha$ and a power (1-$\beta$) against the specified alternative $P_{true}$:

$$n = \frac{\alpha z_{\alpha/2} (1-P_{target}) + \beta z_{\beta} (P_{true}(1-P_{true}))}{(P_{true}-P_{target})^2}$$

Zhou 2011

Zhou and Hess provide the number of patients needed for a study so that, with probability $\alpha$, the width of a two-sided 95% confidence interval CI for the accuracy of the test will be $\alpha$ or less.

$$n = \frac{z_{\alpha/2}^2 + z_{\beta}^2}{2 \cdot \alpha \cdot \beta}$$

Hess 2012

The number of patients needed for a study so that, with probability $\alpha$, the difference in the proportions is estimated with a confidence interval $CI = p_1 - p_2$.

$$n = \frac{2 \cdot (p_1 + p_2) z_{\alpha/2}^2}{\alpha \cdot (p_1 - p_2)}$$

References


