Selecting new health technologies for evaluation: Can clinical experts predict which new anticancer drugs will impact Danish health care?

Douw, Karla; Vondeling, Hindrik

Published in:
Social Science & Medicine

Publication date:
2007

Document version
Publisher's PDF, also known as Version of record

Citation for published version (APA):
Selecting new health technologies for evaluation: Can clinical experts predict which new anticancer drugs will impact Danish health care?

Karla Douw\textsuperscript{a,},* , Hindrik Vondeling\textsuperscript{b}

\textsuperscript{a}Centre for Applied Health Services Research and Technology Assessment (CAST), University of Southern Denmark, Winslowsvej 9b, DK-5000 Odense, Denmark

\textsuperscript{b}Department of Health Economics, University of Southern Denmark, J.B. Winslowsvej 9b, DK-5000 Odense, Denmark

Available online 2 October 2006

Abstract

Several countries have systems in place to support the managed entry of new health technologies. The big challenge for these so-called horizon-scanning systems is to select those technologies that require decision support by means of an early evaluation. Clinical experts are considered a valuable source of information on new health technologies, but research on the relevance of their input is scarce. In 2000, we asked six Danish expert oncologists to predict whether a sample of 19 new anticancer drugs would impact Danish health care over the next 5 years. In 2005, we assessed the accuracy of these predictions in a delayed type cross-sectional study. The specificity of the Danish experts’ prediction was 1 (95\% confidence interval 0.74–1.00) and the sensitivity was 0.63 (0.31–0.86). The negative predictive value was 0.79 (0.52–0.92) and the positive predictive value was 1 (0.57–1.00). This indicates that clinical experts have the ability to predict which new anticancer drugs are unlikely to have an impact. This information can be used to increase the efficiency of selecting new technologies for evaluation. As the experts missed 37\% of drugs that are in need of guidance, they should not be relied upon to select drugs relevant for evaluation.

\textcopyright 2006 Elsevier Ltd. All rights reserved.

Keywords: Health care technology; Experts; Oncology; Denmark; Forecasting

Introduction

Several countries have systems in place to support the managed entry of new health technologies. The big challenge for these so-called horizon-scanning systems (HSS) is to select those technologies that require decision support by means of an early evaluation of their potential impact. By definition this pertains to new technologies that, when introduced in health care, will be most in need of public planning and regulation to promote or discourage diffusion (Gelijns, Brown, Magnell, Ronchi, & Moskowitz, 2005). Clinical experts are considered a valuable source of information on new health technologies (Robert, Stevens, & Gabbay, 1999), but research on the relevance of their input is scarce. In horizon scanning, information sources to inform on new health technologies are categorized as primary, secondary, and tertiary, depending on...
the proximity to the invention. Clinical expertise is considered a secondary source and can, according to Robert et al. (1999), be used either for identification purposes or to filter information from other sources. ‘Filtering’ is essentially a first selection of identified technologies, resulting in a narrowing down of the potential number of technologies to evaluate. Filtering is recognized by HSSs and their network, EuroScan (EuroScan, 2006) as an individual step in the process of identifying and evaluating significant health technologies. Most HSSs describe it as an implicit process (Douw & Vondeling, 2006). In some systems, filtering is based on experts’ predictions of the impact of new health technologies on the health care system. With this paper we intend to contribute to the further development of filtering mechanisms by analysing the accuracy of predictions by Danish expert oncologists in 2000 on newly identified anticancer drugs. The predictions were sought as part of a preparatory study on a Danish HSS and aimed to support a first selection of technologies for early evaluation (Douw, Vondeling, Sørensen, Jørgensen, & Sigmund, 2004).

Methods

The predictions of clinical experts were assessed as a diagnostic tool in a diagnostic accuracy study. The study is a “delayed type cross-sectional study”, because the reference standard was not applied at the same time as the index test. Instead the drugs were followed up during a suitable predefined period (Knotterus & Muris, 2003), in the case of HSSs a five-year period (Robert et al., 1999). The sample included 19 anticancer drugs in treatment of breast, lung, colorectal, and skin cancer, to be administered by clinicians in hospitals. It was part of a larger sample obtained in 2000 by a postal survey to collect newly identified health technologies at all HSSs involved in EuroScan.

The index test involved the predictions of 6 clinical oncologists of the impact of the 19 anticancer drugs. The oncologists were pointed out as experts by senior executives and managers representing stakeholders in Danish health care (Douw et al., 2004). All experts were in senior managing positions, and active in research and clinical practice at 5 of 6 specialized oncology centres in Denmark (Aalborg, Aarhus, Herlev, Odense and Vejle hospital and Rigs hospital in Copenhagen). We asked the experts to predict which drugs would be new for Denmark and be introduced within 0–5 years. We furthermore asked them to fill out for 4 types of impact (i.e. clinical, financial, organisational, social and/or ethical) whether a small, big or no impact could be expected. The experts could select more than one type of impact. For example, an expert could fill out that a drug would have a big financial impact, a small clinical impact, a small organisational impact, and no social impact (Douw et al., 2004). An anticancer drug was considered for further evaluation if a majority of the experts judged it to be new, expected it to be introduced within 5 years, and to have at least one type of big impact.

The reference standard to establish the final diagnosis of impact reflects the need for decision making on the drug. A drug’s impact was considered to be predicted correctly by the experts when either a positive or negative decision was made by the Cancer Steering Group (CSG), a national committee installed by the Health Minister in 1998 to supervise the improvement of cancer diagnostics and treatment, or if actual use was documented in at least one of the 6 oncology centres in Denmark in the period 2000–2005. Use was considered as a proxy for an investment decision by hospitals as well, because of the typically high costs of new anticancer drugs (Apolone, Joppi, Bertele, & Garattini, 2005). We documented use with a postal survey among all 6 oncology centres, carried out in June 2005.

Data were analysed on the basis of Altman’s statistical package using the recommended Wilson’s method for confidence intervals (Altman, Machin, Bryant, & Gardner, 2000).

Results

Table 1 shows that a majority of experts predicted five cases to have an impact on Danish health care. In all these cases exclusively a big financial impact was predicted. In only one of the five cases a decision was made by the CSG, but all five drugs were introduced in clinical practice in at least one of the six Danish oncology centres in the period 2000–2005. A minority of experts predicted twelve of the drugs (63%) to have a big clinical impact, 13 drugs (68%) to have a big organisational impact, and 6 drugs to have a big social/ethical impact. In the cases of Gefinitib, Capecitabine and Tegafur Uracil the experts did not correctly anticipate the future.
The prevalence of drugs in the sample that was in need of decision making was 42% (8 out of 19). Table 2 shows the correspondence between the index test and the reference standard. The index test, i.e. the experts’ predictions of the impact of the new anticancer drugs, had a sensitivity of 0.63 (95% confidence interval 0.31–0.86). The specificity of the test was 1 (0.74–1.00). The negative predictive value was 0.79 (0.52–0.92) and the positive predictive value was 1 (0.57–1.00).

**Discussion**

The analysis indicates that the experts have the ability to predict which technologies are unlikely to have an impact on the health care system. Their predictions are therefore useful in informing which drugs should not be selected for evaluation. We do not advocate relying on experts’ predictions to indicate which anticancer drugs will truly have an impact, since they missed 37% of the drugs that were in need of decision making (sensitivity = 0.63). A PPV of 1 looks like a rather good result for a test but this is influenced by a high prevalence (42%) of drugs that are in need of guidance, which goes together with a high probability to encounter such cases by chance. For the 5 anticancer drugs in our sample that were judged to have an impact, always a financial impact was predicted. This is not surprising given the typically high costs of new anticancer drugs (Apolone et al., 2005). Eight of the 11 drugs correctly predicted by the Danish experts not to
have an impact, never reached the market, and would therefore never have made an impact. Three out of 14 drugs are incorrectly labelled as not being in need of decision making. This can be interpreted as a failed alarm of 21% (1-NPV), and is tolerable as drugs predicted not to have an impact will be monitored.

Simpson, Hyde, Cook, Packer, & Stevens (2004) conclude in a comparable analysis that for the English and Welsh context sensitivity is the most important result, in order not to miss any potentially significant technologies, that otherwise would enter the health care system without evaluation. As the outcome of this HSS’s selection process feeds into the topic selection process for a national guidance programme (NICE) this conclusion is understandable, but for the Danish context a high sensitivity is not the key, as experts’ information is just a first step in the selection, which will be followed by a criteria-based priority-setting procedure.

In this study the index test, as well as the reference test need to be discussed. Regarding the index test, the selection of experts and the way they were accessed are important issues. We asked stakeholders to select opinion leaders in the area of oncology (Douw et al., 2004), as in general these are considered to have appropriate domain knowledge. However, it is unclear what exactly constitutes expertise in relation to forecasting the impact of new anticancer drugs, and more research into which experts to select would therefore be needed. In this study we used a single round to elicit opinions, and used a decision rule (a majority vote of ‘positive’ answers on novelty, time horizon, and impact of the technology) to interpret the answers of the experts, as there was never unanimity on all the answering options. If not a majority vote would be applied than 89% of the anticancer drugs would be judged to have an impact. The sensitivity of the test would then be 1.00, and specificity 0.18, resulting in a high number of false positives, which would not serve filtering purposes very well. There is evidence that several rounds, as used in consensus methods such as the Delphi method or the nominal group technique (NGT), improve the accuracy of a prediction (Armstrong, 2001). These methods could also shed more light on the rationale for the answers, as experts can provide arguments for their estimates when confronted with feedback from the other experts. The value of both methods as forecasting tools is, however, not uncontested (Armstrong, 2001), therefore there is a need for studies comparing different alternatives to elicit expert opinion on the impact of new health technologies. In any method it is recommendable to match the results to observable events (Jones & Hunter, 1995), as was done in this study.

Although the true value of clinical experts’ predictions needs to be established by future studies with bigger samples and in other clinical specialties as well, this study indicates that experts’ information may be valuable as part of a process aimed at efficiently selecting technologies for evaluation.

References


