ELSEVIER

Journal of Clinical Epidemiology 58 (2005) 1211–1216

Journal of Clinical Epidemiology

Clinical experience did not reduce the variance in physicians' estimates of pretest probability in a cross-sectional survey

Amos Cahan[a], Dan Gilon[b], Orly Manor[c], Ora Paltiel[c,d,*]

[a]*Faculty of Medicine, The Hebrew University of Jerusalem, Jerusalem, Israel*
[b]*Heart Institute, Hadassah University Hospital, Jerusalem, Israel*
[c]*School of Public Health, The Hebrew University of Jerusalem, Jerusalem, Israel*
[d]*Department of Hematology, Hadassah University Hospital, Jerusalem, Israel*

Accepted 28 February 2005

Abstract

Background and Objectives: In light of the increasing popularity of the threshold approach in clinical decision-making, this study assesses the role of expertise in physicians' agreement in estimating the probability of disease in patients.

Methods: A cross-sectional survey of physicians of different specialties, attending weekly staff meetings in four teaching hospitals in Jerusalem, Israel. An anonymous questionnaire describing three case scenarios of patients with chest pain was administered and participants were asked to estimate pretest probabilities of disease.

Results: Eighty-six physicians (practicing cardiology, internal medicine, and family medicine, as well as general practitioners and internists) out of 125 approached (response rate 69%). The mean estimated probabilities were very similar for residents and specialists; however, the standard deviation was higher for specialists in all three cases: 20.7, 21.0, and 19.1 among specialists and 16.4, 20.5, and 14.9 among residents, respectively.

Conclusion: This study, based on case scenarios, did not find that medical expertise improved agreement among doctors when estimating the probability of disease in patients—despite the common belief that senior physicians should have smaller interobserver differences in probability estimates. The wide variation observed calls into question the applicability of the threshold approach.

Keywords: Threshold approach; Variance; Pretest probability; Probabilistic-reasoning; Clinical decision-making

1. Introduction

In the face of increasing knowledge on the one hand and clinical uncertainty on the other, rational medical decision-making is ever a challenge. The *threshold approach* [1] has been proposed to improve decision-making by applying probabilistic thinking to clinical practice.

Briefly, the approach involves setting probability thresholds to guide management using Bayes' theorem [2,3]. These thresholds reflect the risks and benefits associated with using diagnostic tests or treating a disease, as opposed to expectant management. Two thresholds are set: the *testing threshold* indicates the lowest pretest probability (P) for the presence of disease for which investigation is indicated to rule it out and the *test–treatment threshold* indicates the highest P for which further investigation is indicated before treatment. Performing additional diagnostic tests is advised only if P lies between the two thresholds.

The threshold approach requires that physicians estimate P, which is based on the prevalence or probability of the disease in specific populations and clinical settings. The post-test probability (P^*) can then be calculated based on information, particularly sensitivity and specificity or likelihood ratios (LR) [2,4–9], regarding the performance characteristics or validity of the diagnostic test, taking P into account. The P^* is then compared to the thresholds, and further investigation is carried out if indicated.

The physician's estimate of P is key to the entire decision-making process. For this process to be valid, P must be estimated fairly accurately [1], and there should also be reasonable consensus among doctors with respect to its value. Lack of consensus would lead to wide fluctuations in estimates of P^*, resulting in wide variability of therapeutic and diagnostic decisions.

Nevertheless, achieving accuracy and reproducibility in the estimation of P is challenging, as medicine is practiced

* Corresponding author. Tel. +972-2-6777601; fax: +972-2-6449145.
E-mail address: ora@vms.huji.ac.il (O. Paltiel).

0895-4356/05/$ – see front matter
doi: 10.1016/j.jclinepi.2005.02.014

otherwise than traditionally taught. Textbooks focus mainly on diseases, but patients present with symptoms, and the doctor must consider the possible diseases that may account for them. In other words, the diagnostic process is inductive rather than deductive.

Adapting this inverted way of thinking is cardinal to successful patient management. Formal decision-making may involve complicated calculations, but most doctors decide intuitively [10]. They master this art during the residency or apprenticeship period, by learning through practice. Given that clinical reasoning is an acquired and continuously evolving expertise, it appears reasonable that a senior physician's differential diagnosis, constructed after evaluating a patient, should more realistically reflect the possible hierarchical explanations of that patient's condition. Furthermore, it follows that the differential diagnoses constructed by experienced doctors would be closer to each other than to those constructed by junior practitioners.

Surprisingly, little has been written about the accuracy and consensus of doctors' probability estimations, and even less about the role of training in improving these skills. The few studies that have evaluated the influence of experience on clinical decision-making have shown equivocal results [10–15]. In one study, cardiologists estimating the prognosis of patients with heart disease were correct, on average, less frequently than statistical computer software that used an epidemiologic database [10]; however, no significant difference was found between participants according to their clinical experience. Another study that compared the estimated probability of ischemic heart disease (IHD) based on case presentations with physicians' own estimation given 5 years earlier did not demonstrate a within-subject difference attributable to clinical experience [12]. When physicians were compared to parents assessing the probability of acute otitis media and need for antibiotic treatment according to written case scenarios, no major difference was found between them in the mean estimation and range [13]. Another study [14] presented cardiologists, general practitioners, and students with case vignettes describing actual patients with and without heart failure (subjects and controls). The participants were asked to decide which vignette described a patient with heart failure. No major differences were found among the three groups in their diagnostic accuracy; however, the range of probabilities assigned to each specific vignette by each of the three groups was large: ~70%. The variance between doctors was not reported.

On the other hand, a study that assessed the approach of gynecologists to genital herpes infection, described a significantly different management strategy according to degree of clinical experience [15]. A study that assessed doctors who described themselves as being skillful in evaluating girls with suspected sexual abuse showed that experienced doctors provided, on average, more reliable diagnoses [16].

In the context of the threshold approach, one may ask whether an estimated P given by an experienced doctor is comparable with that of a junior in terms of its accuracy and reproducibility. Although the reliability and validity of diagnostic tests are strictly measured, evaluated, and standardized before approval for clinical use, the P estimation of the doctor who eventually orders these tests and interprets them is not.

Our objectives were to determine whether, when assessing P based on identical information (a written case scenario describing the history, physical examination, and ECG), differences would be found between senior and junior doctor populations in mean estimated P and variance (s^2).

2. Materials and methods

2.1. Participants

For 6 months (January 1–June 30, 2001), 125 doctors practicing cardiology, internal medicine, and geriatrics at the four teaching hospitals in Jerusalem, Israel, were approached in weekly staff meetings and asked to participate in the study. Doctors practicing family medicine and general practitioners (GPs) were approached while attending a weekly lecture day for community-based doctors at Hadassah University Hospital. The participants included specialists, residents, and interns.

2.2. Instruments

Participants were asked to complete an anonymous questionnaire (Appendix A) in which three clinical cases dealing with chest pain were presented, including the history, physical examination and electrocardiogram description. Based on each case, the doctors were asked to estimate the percentage probability that the patient had active coronary artery disease. A further questionnaire, analyzed separately, assessed the presence of cognitive biases in probability assessment [17].

2.3. Data analysis

The mean estimated probability and the variance were computed for each of the cases for all of the participants as well as for specialists and residents separately. Due to the small number of participating interns and GPs, the responses of these doctors were not included the data analysis. The interquartile range for the mean estimated probability was calculated for specialists and residents for each of the three cases. For each case, the ratio between variance among specialists and variance among residents was computed, and a 95% confidence interval was calculated for the ratio. In cases where the distribution of probability estimates was considerably skewed from normal, we used a power transformation to transform the