Does Prevalence Matter to Physicians in Estimating Post-test Probability of Disease? A Randomized Trial

Thomas Agoritsas, MD, Delphine S. Courvoisier, PhD, Christophe Combescure, PhD, Marie Deom, MD, and Thomas V. Perneger, MD, PhD

Division of Clinical Epidemiology, University Hospitals of Geneva, Geneva 14, Switzerland.

BACKGROUND: The probability of a disease following a diagnostic test depends on the sensitivity and specificity of the test, but also on the prevalence of the disease in the population of interest (or pre-test probability). How physicians use this information is not well known.

OBJECTIVE: To assess whether physicians correctly estimate post-test probability according to various levels of prevalence and explore this skill across respondent groups.

DESIGN: Randomized trial.

PARTICIPANTS: Population-based sample of 1,361 physicians of all clinical specialties.

INTERVENTION: We described a scenario of a highly accurate screening test (sensitivity 99% and specificity 99%) in which we randomly manipulated the prevalence of the disease (1%, 2%, 10%, 25%, 95%, or no information).

MAIN MEASURES: We asked physicians to estimate the probability of disease following a positive test (categorized as <60%, 60–79%, 80–94%, 95–99.9%, and >99.9%). Each answer was correct for a different version of the scenario, and no answer was possible in the "no information" scenario. We estimated the proportion of physicians proficient in assessing post-test probability as the proportion of correct answers beyond the distribution of answers attributable to guessing.

KEY RESULTS: Most respondents in each of the six groups (67%–82%) selected a post-test probability of 95–99.9%, regardless of the prevalence of disease and even when no information on prevalence was provided. This answer was correct only for a prevalence of 25%. We estimated that 9.1% (95% CI 6.0–14.0) of respondents knew how to assess correctly the post-test probability. This proportion did not vary with clinical experience or practice setting.

CONCLUSIONS: Most physicians do not take into account the prevalence of disease when interpreting a positive test result. This may cause unnecessary testing and diagnostic errors.

Electronic supplementary material *The online version of this article (doi:10.1007/s11606-010-1540-5) contains supplementary material, which is available to authorized users.*

Received April 13, 2010
Revised July 26, 2010
Accepted October 1, 2010
Published online November 4, 2010

KEY WORDS: Bayes' theorem; predictive value of tests; prevalence; sensitivity and specificity; diagnosis; risk assessment; probability; evidence-based medicine.
J Gen Intern Med 26(4):373–8
DOI: 10.1007/s11606-010-1540-5

INTRODUCTION

Estimating the probability of a disease following a diagnostic test is a basic clinical skill that guides treatment decisions and investigation strategies. According to Bayes' theorem, the post-test probability depends on the specificity and the sensitivity of the test, and on the prevalence of the disease in the population of interest (i.e., pre-test probability)[1].

Updating disease probability estimates based on test results is not intuitive[1–4]. Many physicians misunderstand basic concepts of test accuracy[5–8], whether they are presented as sensitivity and specificity, likelihood ratios, or graph displays[6]. Another difficulty lies in estimating pre-test probabilities[9,10]. How physicians use available information about disease prevalence is not well known. A previous study has suggested that post-test probability is generally overestimated, especially when disease prevalence is low, but this study was limited by a small and selected sample of physicians, and by a narrow range of disease prevalence[5]. No study was able to determine the proportion of physicians who are proficient in estimating post-test probability according to prevalence.

In this study, we explored whether physicians' estimates of post-test probability vary with disease prevalence in a hypothetical diagnostic situation. We also estimated the proportion of physicians who can correctly apply Bayes' theorem in deriving the post-test probability and explored whether this skill varies with sociodemographic characteristics, practice setting, or clinical experience.

METHODS

Study Design and Subjects

We conducted a mail survey among physicians of all clinical specialties working in Geneva, Switzerland, whether prac-

ticing in a public hospital or in private practice, either still in training or board certified. Participants were identified from two databases, one including most private practitioners in Geneva, the other comprising all salaried staff at the Geneva University Hospitals. After exclusion of duplicate records and physicians who did not work with patients (e.g., pathologists, public health specialists, etc.), we sent the survey package to 2,745 eligible participants on October 2007, with up to two reminders in the next 4 months. This study was approved by the Research Ethics Committee of Geneva University Hospitals.

Questionnaire

The questionnaire assessed doctors' opinions on various practice-related issues, which are not analyzed in this paper. Respondents were also asked to report on their age, sex, year of birth, year of graduation, specialty (either completed or planned), and practice setting (private practice or public hospital practice).

In addition, the survey included the following hypothetical vignette (translated from French):

> *"As a school doctor, you perform a screening test for a viral disease in a primary school. The properties of the test are very good:*
>
> - *among 100 children who have the disease, the test is positive in 99, and falsely negative in only 1, and*
> - *among 100 children who do not have the disease, the test is negative in 99, and falsely positive in only 1.*
>
> *On average, about x % of children are infected without knowing it."*

This vignette was produced in six different versions, in which the prevalence of the disease took the following values: 1%, 2%, 10%, 25%, and 95%; the sixth version gave no information on prevalence. Each participant was randomized to receive only one of the six versions of the vignette and was blinded to the existence of alternative versions. At the end of the vignette, respondents were asked the following multiple-choice question: *"If the test for one of the children is positive, what is the probability that he actually has this viral disease?"* Possible answers were on a five-point scale: (1) <60%, (2) 60–79%, (3) 80–94%, (4) 95–99.9%, and (5) >99.9%. For each version, one option was the correct post-test probability. This correct answer depended only on the parameters described in the vignette, namely: the sensitivity, specificity, and disease prevalence. Indeed, the post-test probability was 50.0% for a prevalence of 1%, 66.9% for a prevalence of 2%, 91.7% for 10%, 97.1% for 25%, and 99.95% for a prevalence of 95% (Fig. 1). No answer was possible in the "no information" scenario. We did not explicitly name sensitivity and specificity, but rather expressed them as natural frequencies[11], in order to overcome possible misunderstandings about their definitions[12]. We described a hypothetical disease to avoid the bias of previous knowledge. The vignette was pre-tested for understanding and readability in face-to-face interviews with 15 physicians before the survey.

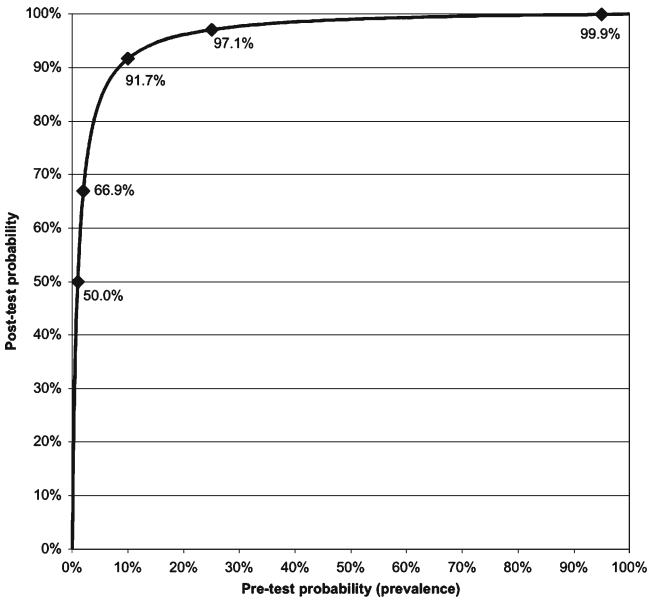

Figure 1. Probability modifying plot: relationship between pre-test probability (i.e., prevalence of a disease) and post-test probability, for a positive diagnostic test with 99% of sensitivity and specificity. The diamonds show the post-test probabilities corresponding to the different values of prevalence that were randomly manipulated in our questionnaire. This relationship is described by the following equation:

$$\text{Post-test probability} = \frac{\text{Sensitivity} * \text{Prevalence}}{(\text{Sensitivity} * \text{Prevalence}) + (1 - \text{Specificity}) * (1 - \text{Prevalence})}$$

Statistical Analysis

We compared respondents who answered the vignette with those who skipped it but filled out the rest of the questionnaire. Then we compared the six groups of patients randomly allocated to the different versions of the vignette.

We examined the frequency distributions of estimated post-test probabilities across versions of the vignette (Table 1, correct answers are marked with an *). We hypothesized that our sample is composed of two populations: (1) those who knew how to estimate post-test probability and all of whom answered correctly, and (2) those who did not know but could answer correctly by chance. We sought to estimate the proportion of respondents in the first population. To do so, we performed a median polish of the column proportions of our data (Table 1, excluding the subgroup without information on prevalence). Median polish is an exploratory data analysis technique for two-way tables, described by Tukey[13]. It is similar to a two-way ANOVA, but uses medians instead of means, which makes it robust to outliers. This is an important property, as we expected one outlier per row and per column, the correct answer. This procedure removes the row effect of a table (i.e., the observed distribution of post-test probability estimates independent of the version of the vignette) and the column effect (i.e., the effect due to each different version), and yields residual proportions (see online appendix for details). We hypothesized that there would be only one larger residual for each version of the vignette, corresponding to the correct answer, which captures the proportion of respondents who correctly estimate post-test probability beyond the distribution of answers attributable to guessing. We computed the median

Table 1. Answer Frequencies (and Column %) in the Five Proposed Categories of Post-test Probability According to the Prevalence of the Disease (Pre-test Probability)

Post-test probability	Prevalence of the disease (randomized by group of physicians)						
	1%	2%	10%	25%	95%	Not mentioned	Total
>99.9%	28 (13.1)	20 (8.7)	30 (13.0)	22 (9.6)	46 (20.4)*	35 (15.0)	181 (13.3)
95 to 99.9%	150 (70.4)	184 (80.3)	154 (66.7)	184 (80.3)*	160 (71.1)	191 (81.6)	1,023 (75.2)
80 to 94%	7 (3.3)	5 (2.2)	37 (16.0)*	7 (3.1)	12 (5.3)	4 (1.7)	72 (5.3)
60 to 79%	2 (0.9)	13 (5.7)*	2 (0.9)	11 (4.8)	1 (0.4)	1 (0.4)	30 (2.2)
<60%	26 (12.2)*	7 (3.1)	8 (3.5)	5 (2.2)	6 (2.7)	3 (1.3)	55 (4.0)
Total	213 (100.0)	229 (100.0)	231 (100.0)	229 (100.0)	225 (100.0)	234 (100.0)	1,361 (100.0)

**For a test with 99% sensitivity and specificity (i.e., likelihood ratio of 99), correct answers are marked with an* * (i.e., for a prevalence of 1% the post-test probability is 50.0%, for a prevalence of 2% → 66.9%, for a prevalence of 10% → 91.7%, for a prevalence of 25% → 97.1%, and for a prevalence of for 95% → 99.95%)

of these five residuals as the overall proportion of physicians proficient in estimating post-test probability. The 95% confidence interval on this proportion was obtained by non parametric bootstrap (bias-corrected and accelerated bootstrap, 10,000 iterations)[14]. Finally, we explored this proportion and its 95% confidence interval across subgroups of respondents. Descriptive analyses were performed with SPSS 15.0 software. Median polish and bootstrap were performed using R 2.9.2[15].

Table 2. Respondents' Characteristics and Estimated Proportion of Physicians that Correctly Assess Post-test Probability Across Subgroups of Respondents

	N* (column %)	Proportion of physicians that skilfully estimate post-test probability % (95% CI)§
All respondents	1,361 (100%)	9.1 [6.0 to 14.0]
Sex		
Female	520 (38.2)	9.0 [5.6 to 17.7]
Male	841 (61.8)	8.9 [4.4 to 14.8]
Age		
≤35 years	288 (21.3)	7.6 [2.4 to 16.1]
36–50 years	551 (40.5)	9.8 [4.1 to 17.3]
>50 years	521 (38.3)	4.4 [0.0 to 8.7]
Practice setting		
Private practice	748 (55.0)	7.2 [3.9 to 12.7]
Public hospital, in training	477 (35.0)	9.8 [5.8 to 18.8]
Public hospital, senior	136 (10.0)	11.0 [0.0 to 19.2]
Specialty		
Primary care physicians†	519 (38.3)	12.4 [7.5 to 21.1]
Internal medicine specialists	208 (15.3)	5.1 [0 to 15.2]
Technical specialists‡	392 (28.9)	6.8 [3.1 to 14.3]
Psychiatrists	237 (17.5)	5.4 [0.0 to 13.3]
Number of years after diploma		
≤10 years	357 (26.5)	6.7 [1.5 to 12.8]
11–20 years	347 (25.8)	8.4 [1.3 to 15.7]
21–30 years	394 (29.3)	7.7 [1.0 to 13.4]
>30 years	247 (18.4)	8.0 [4.1 to 21.2]

**The total can be different from N=1361 because of missing values ranging from 0 to 16*

†*Including general practitioners, general internists, and paediatricians*

‡*Including surgeons, gynecologists and obstetricians, anesthesiologists, ENT, ophthalmologists, radiologists, and dermatologists*

§*Computed among the 1,127 respondents who received information on prevalence. See methods for a detailed explanation of this estimation. Confidence intervals (CI) were obtained by non-parametric bias corrected and accelerated bootstrap*

RESULTS

Sample Characteristics

Of 2,745 eligible physicians, 1,544 (56.2%) returned completed questionnaires, and 1,361 (88.1% of 1,544) answered the diagnostic vignette and were included in the analysis. Vignette respondents were younger than the 183 non-respondents (mean age 46.6 vs 52.5 years, $p<0.001$) and had more recently graduated (mean number of years 20.0 vs 24.2 $p<0.001$). A higher proportion was in training (35.0% vs 19.7%, $p<0.001$) and worked as primary care physicians (38.3 vs 29.5, $p=0.01$). Respondents' characteristics are shown in Table 2, column 1.

There were no differences among the six groups of physicians allocated to the different versions of the vignette in terms of response rate and for all covariates tested (data not shown, available upon request).

Clinical Diagnostic Vignette

The distribution of post-test probability estimates was similar among the five groups of physicians allocated to the different versions of the vignette that provided information on disease prevalence (Table 1). A majority of respondents (between 66.7% and 80.3% across versions) selected a post post-test probability of 95–99.9% regardless of the prevalence of disease. This answer was correct for a prevalence of 25% and incorrect in all other groups. The same distribution was observed in the sixth group, which received no information on prevalence.

A slightly higher proportion of respondents selected the correct answer in each of the five groups where information on prevalence was provided (correct answers are marked with an * in Table 1). The proportion of correct answers varied across the versions from 5.7% to 20.4%, except in the version with a prevalence of 25%, where 80.3% of respondents selected the correct post-test probability of 95–99.9%.

Using Tukey's median polish, we estimated that the proportion of respondents who knew how to estimate post-test probability was 9.1% (95% CI 6.0–14.0; see online appendix). This proportion did not substantially vary according to sex, age, number of years after graduation, clinical specialty, or practice setting (Table 2, column 2). Primary care physicians

performed slightly better (12.4%) compared to other specialties (5.1 to 6.8%), but all confidence intervals overlapped widely.

DISCUSSION

Most physicians did not use information about the prevalence of the disease correctly when interpreting a positive result of a diagnostic test. In a scenario that described a highly accurate screening test (sensitivity 99% and specificity 99%), most selected a post post-test probability of 95–99.9%, regardless of the prevalence of disease. The same distribution was found when no information on prevalence was provided, even though it is impossible to derive the post-test probability in this case. Overall, only about one out of ten physicians knew how to correctly assess post-test probability, a proportion that did not vary significantly across subgroups of respondents.

Comparison with Other Studies

In an earlier study, 31 physicians and 19 non-physicians at a university hospital evaluated the post-test probability of disease based on three levels of prevalence (0.001%, 0.1%, and 10%)[5]. The respondents generally overestimated post-test probability. Our results are compatible with this finding. In addition, by manipulating disease prevalence over a wider range, we were able to show that physicians not only overestimate post-test probabilities, but that a large majority actually ignore information on prevalence.

Three other studies confirm that doctors are not very good at estimating post-test probabilities. Only 22% of Swiss general practitioners chose the correct range of post-test probability (<25%) when presented with a positive screening test (sensitivity and specificity of 95%) and a disease prevalence of 1%[8], and only 17% of US pediatricians estimated a nearly correct post-test probability when faced with a negative test result (sensitivity 50% and specificity 95%) and a disease prevalence of 30%[7]. In a third study conducted among US students, house officers, and physicians, more than 50% largely overestimated post-test probability (>50%, the correct answer being 9%) when presented with a positive screening test (sensitivity and specificity of 95%) and a disease prevalence of 0.5%[16]. These results are more favorable than our estimate of 9% of physicians proficient in estimating post-test probability, but they do not account for responding correctly by chance, as they did not manipulate disease prevalence.

Interpretation of the Results and Implications

What might be the reasons for such poor performance among physicians? Firstly, some doctors may not know that post-test probability depends on disease prevalence. This is corroborated by our finding that most physicians who were given no information on prevalence still guessed at post-test probabilities, although it is impossible in this case.

A second explanation is that physicians misunderstand basic notions of test accuracy[12,17]. Most respondents selected a post post-test probability of 95–99.9%, perhaps because they confused post-test probability with sensitivity or specificity. Such confusion is reflected by the inconsistency of the language used everyday to express diagnostic reasoning[17].

Thirdly, physicians may have difficulties with the arithmetic component of post-test probability estimation[1,4]. Two main strategies can be used to reach a correct estimate. One relies on the direct application of Bayes' theorem. Prevalence is converted into pre-test odds [odds = prevalence/(1-prevalence)], and the positive likelihood ratio is obtained from test characteristics [LR of a positive test = sensitivity/(1-specificity)]. The product of pre-test odds and the LR is the post-test odds, which are then converted into post-test probability [probability = odds/(1 + odds)]. An alternative strategy is to construct a two-way contingency table, with an arbitrary total number of patients (e.g., 10,000), and estimate each cell from the prevalence and test characteristics. Post-test probability is then obtained by dividing the number of true positives by the number of all positive tests[11,18]. Fagan's nomogram[19], computer software or handheld devices can be used as effective shortcuts for these computations. Given our mail survey approach, participants in our study would have had the opportunity to use these aids or even to look for assistance on how to provide the correct answer. A similar study in a classroom or specific practice setting may thus have shown an even lower proportion of correct answers, which would only reinforce our conclusions regarding the inability of physicians to correctly estimate post-test probability.

Given the recent emphasis on clinical epidemiology and evidence-based medicine in medical curricula, one could hypothesize that more recently trained physicians would be more proficient at estimating post-test probability. This was not confirmed by our results. Moreover, two studies conducted among medical students suggested that traditional teaching of Bayesian reasoning may not be effective enough in improving diagnostic performance[18,20].

Conversely, Bayesian skills might be acquired through clinical experience, i.e., through regular exposure to the interpretation of diagnostic tests[3,21,22]. However, we found that more experienced physicians did not perform better. Clinical experience may improve the estimation of pre-test probability at the bedside, but does not lead to an accurate application of Bayes' rule.

Does the inability to correctly estimate post-test probability affect patient care? This depends on the clinical decision that follows this estimation. We did not investigate this relationship, but an earlier study showed that post-test probability estimates were indeed associated with subsequent intended management [7]. An important barrier to clinicians using Bayes' theorem may be that many correct diagnoses and decisions can be made using non-probabilistic, yet effective cognitive processes, both analytical[4,23] and non-analytical [24,25]. However, diagnostic errors remain common, as they range from <5% in radiological or pathological specialties up to 10–15% in most other clinical fields[24]. Inaccurate probability estimation is considered a substantial contributor to these errors [4,10,23,26,27]. In light of our results and of previous work, increased risk of diagnostic error and unnecessary testing may result either from the overestimation of positive predictive value when pre-test

probability is low (as in our scenario) or the overestimation of negative predictive value when pre-test probability is high. Of course, clinical decisions rarely depend on the interpretation of a single test, many tests provide more information than the presence or absence of a disease (e.g., imaging procedures)[1], and complex clinical decisions generally involve more than one clinician or specialist[23]. Nevertheless, in several clinical situations incorrect post-test estimation would still result in unnecessary testing, increased patient anxiety, and diagnostic errors[2,4,10,20]. Finally, poor understanding of probability thinking may also lead to the inability to interpret and make use of available evidence on new diagnostic procedures correctly[4,28].

Study Limitations and Strengths

The main limitation of our study is that the evidence is based on a hypothetical vignette. Although vignette studies are used increasingly to assess physician practice variation in diagnostic and treatment decisions[6–8,29], they constitute a simplification of real clinical practice. On the other hand, the correct answer to our multiple-choice vignette depends only on the parameters presented (namely: the sensitivity, specificity, and disease prevalence) and should not be influenced by the clinical context. Moreover, the use of a hypothetical vignette allowed us to isolate specific aspects of diagnostic testing[29], in this case the prevalence of disease, which cannot be manipulated in real clinical situations, at least not over a wide range. A better identification of specific barriers in the evidence-based diagnosis is essential, as most attempts to improve it have remained unsuccessful[4,7,20].

Another limitation is possible selection bias. Although typical for physicians' surveys[30], we obtained a moderate response rate of 56%, which may affect absolute estimates of physicians' abilities. However, it is unlikely that selection bias would influence the comparisons of the randomly assigned values of prevalence, as response rate was comparable across the different versions of the scenario.

As with any local study, generalizability to other contexts is uncertain. In particular, physicians' proficiency may differ according to their specific training. Yet, contemporary Swiss medical curricula and residency programs are comparable to those of many other countries, especially in their emphasis on clinical epidemiology and evidence-based medicine. Moreover, among the studies on probability estimation that we discussed above, three used physician samples from the USA and showed fairly comparable performance[5,7,16]. For these reasons, we believe our results are generalizable.

The main strength of this study is the random manipulation of disease prevalence on a wide range of values, which yielded a comprehensive perspective on how this information is used—or in this case, ignored—by physicians. This also allowed us to estimate the overall proportion of physicians who are proficient in estimating post-test probability. Finally, we surveyed a large population-based sample of physicians, from various clinical specialties, practice settings, and clinical experience, in contrast with previous research on post-test probability estimation[5–8,16].

Conclusion and Future Prospects

We found that most physicians did not use prevalence of disease when they estimated post-test probability, independently of clinical experience, practice setting, or recent exposure to teaching in evidence-based diagnosis. This may cause unnecessary testing and diagnostic errors. Regular training is probably needed to improve familiarity with probabilistic diagnostic reasoning. Further research should explore novel approaches to help physicians implement Bayes' rule in clinical practice. Wider use of graphical presentations of diagnostic information[31], testing of new didactic models[18,32], and development of computerized knowledge resources for evidence-based diagnosis[4] are avenues for further study.

Acknowledgments: *The authors thank Véronique Kolly, Research Nurse at the University Hospitals of Geneva, for her help and guidance in conducting the survey.*

Funding: *The survey was funded by The Research and Development Program of University Hospitals of Geneva. There was no external funding.*

Presentations: *Part of this paper was presented as an abstract and poster at the 78th Annual Meeting of the Swiss Society of Internal Medicine, Basel, Switzerland, in May 2010*

Conflicts of Interest: *None Disclosed.*

Corresponding Author: *Thomas Agoritsas, MD; Division of Clinical Epidemiology, University Hospitals of Geneva, Gabrielle Perret-Gentil 6 1211, Geneva 14, Switzerland (e-mail: thomas.agoritsas@hcuge.ch).*

REFERENCES

1. **Phillips B, Westwood M.** Testing our understanding of tests. Arch Dis Child. 2009;94(3):178–179.
2. **Ghosh AK, Ghosh K, Erwin PJ.** Do medical students and physicians understand probability? QJM. 2004;97(1):53–55.
3. **Reid MC, Lane DA, Feinstein AR.** Academic calculations versus clinical judgments: practicing physicians' use of quantitative measures of test accuracy. Am J Med. 1998;104(4):374–380.
4. **Richardson WS.** We should overcome the barriers to evidence-based clinical diagnosis! J Clin Epidemiol. 2007;60(3):217–227.
5. **Lyman GH, Balducci L.** The effect of changing disease risk on clinical reasoning. J Gen Intern Med. 1994;9(9):488–495.
6. **Puhan MA, Steurer J, Bachmann LM, ter Riet G.** A randomized trial of ways to describe test accuracy: the effect on physicians' post-test probability estimates. Ann Intern Med. 2005;143(3):184–189.
7. **Sox CM, Doctor JN, Koepsell TD, Christakis DA.** The influence of types of decision support on physicians' decision making. Arch Dis Child. 2009;94(3):185–190.

8. **Steurer J, Fischer JE, Bachmann LM, Koller M, ter Riet G.** Communicating accuracy of tests to general practitioners: a controlled study. BMJ. 2002;324(7341):824–826.
9. **Attia JR, Nair BR, Sibbritt DW, et al.** Generating pre-test probabilities: a neglected area in clinical decision making. Med J Aust. 2004;180 (9):449–454.
10. **Richardson WS.** Five uneasy pieces about pre-test probability. J Gen Intern Med. 2002;17(11):882–883.
11. **Hoffrage U, Gigerenzer G.** Using natural frequencies to improve diagnostic inferences. Acad Med. 1998;73(5):538–540.
12. **Young JM, Glasziou P, Ward JE.** General practitioners' self ratings of skills in evidence based medicine: validation study. BMJ. 2002;324 (7343):950–951.
13. **Mosteller F, Tukey J.** Data analysis and regression, a second course in statistics: Addison-Wesley publishing company; 1977.
14. **Efron B, Tibshirani R.** An introduction to the Bootstrap: Chapman & Hall; 1993.
15. R Development Core Team. R: A language and environment for statistical computing. http://www.R-project.org (Accessed on 3 October 2010). R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. 2008.
16. **Schwartz WB, Gorry GA, Kassirer JP, Essig A.** Decision analysis and clinical judgment. Am J Med. 1973;55(3):459–472.
17. **Bianchi MT, Alexander BM.** Evidence based diagnosis: does the language reflect the theory? BMJ. 2006;333(7565):442–445.
18. **Kurzenhauser S, Hoffrage U.** Teaching Bayesian reasoning: an evaluation of a classroom tutorial for medical students. Med Teach. 2002;24 (5):516–521.
19. **Fagan TJ.** Letter: nomogram for Bayes' theorem. N Engl J Med. 1975;293(5):257.
20. **Noguchi Y, Matsui K, Imura H, Kiyota M, Fukui T.** A traditionally administered short course failed to improve medical students' diagnostic performance. A quantitative evaluation diagnostic thinking. J Gen Intern Med. 2004;19(5 Pt 1):427–432.
21. **Gill CJ, Sabin L, Schmid CH.** Why clinicians are natural Bayesians. BMJ. 2005;330(7499):1080–1083.
22. **Kahneman D, Slovic P, Tversky A, eds.** Judgment under uncertainty: heuristics and biases. Cambridge, UK: Cambridge University Press; 1982.
23. **Graber M, Gordon R, Franklin N.** Reducing diagnostic errors in medicine: what's the goal? Acad Med. 2002;77(10):981–992.
24. **Berner ES, Graber ML.** Overconfidence as a cause of diagnostic error in medicine. Am J Med. 2008;121(5 Suppl):S2–23.
25. **Norman GR, Eva KW.** Diagnostic error and clinical reasoning. Med Educ. Jan;44(1):94-100.
26. **Croskerry P.** The importance of cognitive errors in diagnosis and strategies to minimize them. Acad Med. 2003;78(8):775–780.
27. **Kassirer JP, Kopelman RI.** Cognitive errors in diagnosis: instantiation, classification, and consequences. Am J Med. 1989;86(4):433–441.
28. **Grijalva CG, Poehling KA, Edwards KM, et al.** Accuracy and interpretation of rapid influenza tests in children. Pediatrics. 2007;119(1):e6–11.
29. **Veloski J, Tai S, Evans AS, Nash DB.** Clinical vignette-based surveys: a tool for assessing physician practice variation. Am J Med Qual. 2005;20 (3):151–157.
30. **Asch DA, Jedrziewski MK, Christakis NA.** Response rates to mail surveys published in medical journals. J Clin Epidemiol. 1997;50 (10):1129–1136.
31. **Whiting PF, Sterne JA, Westwood ME, et al.** Graphical presentation of diagnostic information. BMC Med Res Methodol. 2008;8:20.
32. **Van den Ende J, Bisoffi Z, Van Puymbroek H, et al.** Bridging the gap between clinical practice and diagnostic clinical epidemiology: pilot experiences with a didactic model based on a logarithmic scale. J Eval Clin Pract. 2007;13(3):374–380.