



OPEN ACCESS

How much diagnostic safety can we afford, and how should we decide? A health economics perspective

David E Newman-Toker,¹ Kathryn M McDonald,^{2,3} David O Meltzer⁴

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/bmjqs-2012-001616>).

¹Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

²Center for Primary Care and Outcomes Research/Center for Health Policy, Stanford University, Stanford, California, USA

³School of Public Health, University of California, Berkeley, California, USA

⁴Department of Medicine, Center for Health and the Social Sciences, University of Chicago, Chicago, Illinois, USA

Correspondence to

Dr David E Newman-Toker, Department of Neurology, Johns Hopkins Hospital, Meyer 8-154, 600 North Wolfe Street, Baltimore, MD 21287, USA; toker@jhu.edu

Received 1 April 2013

Revised 7 August 2013

Accepted 8 August 2013



► <http://dx.doi.org/10.1136/bmjqs-2013-002387>

To cite: Newman-Toker DE, McDonald KM, Meltzer DO. *BMJ Qual Saf* 2013;**22**:ii11–ii20.

INTRODUCTION

Controlling the costs of healthcare, which now exceed US\$2.7 trillion, is an economic imperative.^{1–3} Costs of diagnostic testing probably account for more than 10% of all healthcare costs, and that fraction is rising rapidly over time,⁴ with advanced diagnostic imaging leading the way and diagnostic laboratory testing a close second.⁵ Molecular diagnostic testing for ‘personalised medicine’ may fuel disproportionate rises.^{6–7} Some diagnostic tests are misused or overused, with waste from diagnostic imaging alone estimated at more than US\$25 billion.⁸

At the same time, diagnostic errors are frequent and often result in death or disability,⁹ with recent estimates suggesting more than a million a year harmed by diagnostic error in the USA.¹⁰ For those harmed, direct costs accrue from failure to treat the true condition, inappropriate testing and treatments for the incorrectly diagnosed one, and any medicolegal costs or payments.¹¹ Indirect costs also arise from defensive medicine, increased medical liability premiums, and downstream effects.¹¹ The annual costs of ‘defensive medicine’ alone—mostly unnecessary diagnostic tests obtained to guard against malpractice law suits—are at least US\$45–60 billion^{12–13} and perhaps hundreds of billions.^{14–16}

Given these facts, public awareness campaigns (eg, ‘Choosing Wisely’) have sought to foster dialogue between doctors and patients about potential ways to improve the safety and efficiency of diagnosis.¹⁷ Nevertheless, it remains challenging to determine whether diagnostic tests are being overused or underused and when ‘more’ diagnosis is not ‘better’ diagnosis. In this article, we leverage a case study example (box 1) to explore complex inter-relationships between

diagnostic test characteristics, appropriate use, actual use, diagnostic safety and cost effectiveness. We frame our discussion around the question, ‘How much diagnostic safety can we afford?’ to assess the role of economic analysis and suggest areas for future research related to the public health imperative of better value and safety in diagnosis.

FINITE RESOURCES, INFINITE DEMAND, AND WASTE IN DIAGNOSTIC TEST USE

The demand for healthcare and technological advances in diagnosis is effectively infinite.²⁹ For acute dizziness in the emergency department (ED), the number of patients seeking care has more than doubled over the past two decades.¹⁸ The continuous upward spiral in what we ‘can do’ in medicine through technological advances propels a progressively widening gap between what we ‘actually do’ and what we ‘should do.’³⁰ For example, the fraction of ED patients with dizziness undergoing CT scans rose steadily from 9% in 1995 to over 40% currently,¹⁸ but doing so has not increased the yield of stroke or other neurologic diagnoses.^{31–32} Some advances produce societal benefits exceeding their costs, while others do not.³³ More importantly, higher costs are often unrelated to health outcomes—in other words, healthcare resources are often spent inefficiently.³⁴ Some frequently used tests appear to offer no diagnostic benefit at all,³¹ never mind a downstream health benefit. Use of diagnostic tests under such circumstances is generally considered *inappropriate use* (box 2).

Most diagnostic decisions are influenced by factors other than maximising individual patient outcomes or even total societal benefit. In our case study example, CT scans are grossly overused in

Box 1 Case study example—diagnosis of acute dizziness and vertigo in the emergency department (ED)

There are more than four million US ED visits annually for acute vertigo or dizziness at a yearly ED workup cost of over US\$4 billion.¹⁸ The roughly one million who have underlying peripheral vestibular causes are generally over-tested,¹⁹ misdiagnosed²⁰ and undertreated.²¹ Hundreds of millions of dollars are spent on brain imaging trying to ‘rule out’ dangerous central vestibular causes such as stroke,¹⁸ yet, despite that, one-third of vestibular strokes are missed initially.²² Misconceptions^{23 24} drive ED clinical practice, resulting in inappropriate use of diagnostic tests.¹⁹ Patients with inner ear conditions, such as vestibular neuritis (or labyrinthitis) and benign paroxysmal positional vertigo are often imaged and admitted unnecessarily¹⁹ instead of being diagnosed at the bedside, treated appropriately and discharged; error rates may exceed 80%.²⁰ Patients with dangerous brainstem or cerebellar strokes (representing about 5% of all dizziness and vertigo presentations in the ED²⁵) may be sent home without critical stroke treatments, sometimes resulting in serious harm.²⁶

New bedside diagnostic methods to detect stroke²⁷ have been developed, that are supported by strong evidence,²⁵ and could be disseminated,²⁸ but the logical follow-on policy question is, ‘Would it be worth pursuing a radical shift in care to try to reduce misdiagnosis?’ Decision modelling with economic analysis might help answer this question by comparing several hypothetical diagnostic strategies on patient-centred health outcomes of interest.

an effort to ‘rule out’ stroke. Studies showing small area practice variation in use of CT scans for dizziness confirm that factors other than optimal patient care are at play in decisions to order a CT.³² Diagnostic test use by physicians is driven by a mix of incentives and disincentives (box 3), with mixed results for optimal patient care.⁵² Risk aversion may be a particularly important factor contributing to wasteful diagnostic test overuse.⁴⁶ In the case of CTs for dizziness, major contributing factors include knowledge gaps regarding best evidence,²³ local standards for test-ordering (peer practices),³² patient preferences and fear of litigation for missed stroke.

Another incentive often mentioned by physicians is reduced time for patient interaction⁵³ from crowding out by documentation tasks.⁵⁴ It is often faster for a physician to order a test than to think through its appropriateness. Tackling such practical barriers will be important for reducing test overuse—for example, measuring and rewarding diagnostic quality and efficiency could provide incentives that counteract need-less test overuse.

Box 2 Definitions for inappropriate use of diagnostic tests

Underuse—The failure to provide a diagnostic test when it would have produced a favourable outcome for a patient. An example would be failure to provide pap smears to eligible patients.

Overuse—Providing a diagnostic test in circumstances where the potential for harm exceeds the potential for benefit. An example would be conventional cerebral angiography to rule out brain aneurysm in a patient with typical, uncomplicated migraine-type headaches and a normal neurologic examination.

Misuse—When an appropriate diagnostic test has been selected but a preventable complication occurs and the patient does not receive the full potential benefit of the test. An example would be pulmonary CT angiography to diagnose pulmonary embolus in a patient with dyspnoea who has a known contrast dye allergy but receives no pretreatment for a possible allergic reaction.

Modified from AHRQ Web M&M³⁵

DIAGNOSTIC SAFETY: DIAGNOSTIC TEST USE AND DIAGNOSTIC ERROR

Eliminating all diagnostic errors is impossible,⁵⁵ since diagnosis occurs under uncertainty. Efforts to reduce uncertainty toward zero result in increasing marginal costs with diminishing marginal returns for patient safety. Nevertheless, potential exists to improve diagnostic safety. In the ED dizziness/vertigo case, an estimated 35% of underlying strokes are missed,²² despite the fact that relatively simple, non-invasive, bedside physical examination tests have been shown repeatedly to identify more than 99% of strokes.^{25 27 56}

Estimates of diagnostic error-associated adverse events are 0.1% of primary care visits⁵⁷ and 0.4% of hospital admissions,⁵⁸ but the diagnostic error rate is likely much higher, since harm does not invariably occur. Ballpark estimates place the overall diagnostic error rate in the range of 10–15%.⁹ Rates appear to vary by specialty (eg, 5% in radiology vs 12% in emergency medicine),⁵⁹ by disease (eg, 2% of myocardial infarctions vs 9% of strokes),⁶⁰ and especially by clinical presentation (eg, 4% of strokes presenting traditional symptoms vs 64% of strokes presenting non-traditional symptoms),⁶¹ with atypical and non-specific presentations (such as dizziness) increasing the risk of misdiagnosis dramatically.^{60 62}

Misdiagnoses, when they relate to test ordering (~15% of misdiagnoses⁶³), generally result from underuse.⁶³ For example, this appears to be the case for young patients presenting dizziness to the ED whose strokes may be missed, sometimes with devastating consequences.²⁶ Test overuse, however, can also result in diagnostic error (eg, false positives) or

Box 3 Incentives and disincentives that drive diagnostic test-ordering behaviour by physicians

Incentives

- ▶ Patient-oriented
 - Seeking best outcomes of care³⁶
 - Responding to patient and family requests³⁷
- ▶ Algorithmic
 - Practice guidelines³⁸ or care pathways/standing orders
 - Required diagnostic performance metrics³⁹
- ▶ Sociocultural
 - Tradition⁴⁰
 - Small area practice variation³²
- ▶ Time efficiency-driven
 - Productivity targets or reduced visit length⁴¹ (since it is faster to 'just order the test' or order multiple tests 'in parallel' rather than 'in series')
 - Unavailability of prior test results (faster to repeat a test than track down results)
- ▶ Purely financial
 - Greatest profit for individual physicians, hospitals, or health systems in a fee-for-service payment system⁴²
- ▶ Risk averse⁴³
 - Desire to avoid 'chagrin'/regret⁴⁴ ⁴⁵ or uncertainty⁴⁶
 - Fear of medicolegal liability—'defensive medicine'⁴⁷

Disincentives

- ▶ Patient-oriented
 - Medical futility
 - Risks of complications from testing
- ▶ Algorithmic
 - Preapproval requirements
 - Practice guidelines⁴⁸
- ▶ Sociocultural
 - Peer pressure⁴⁹
 - Physician responsibility for healthcare stewardship²
 - Public awareness campaigns (eg, 'Choosing wisely'¹⁷)
- ▶ Time efficiency-driven
 - Unavailable locally
 - Difficult to order
 - Long delay for result
- ▶ Purely financial
 - Greatest profit for individual physicians, hospitals, or health systems in a bundled or capitated payment system⁵⁰
- ▶ Cost-containing
 - Payer policies⁵¹

overdiagnosis (ie, correct but unnecessary diagnoses). Overdiagnosis includes conditions deliberately being sought, but too mild to warrant treatment,⁶⁴ and unrelated 'incidentalomas'⁶⁵ that may beget further

inappropriate testing or treatment.⁶⁶ Tests may not be appropriate when the expected probability of disease is too low or too high for the test to resolve diagnostic uncertainty in a way that positively influences clinical decision making.⁵² Since overuse often occurs in patients with expected low disease prevalence, the risk of false positives is particularly high. Sometimes blind obedience⁶⁷ to false positive or false negative test results from advanced diagnostic technology can lead to error—this appears to be the case with CT for 'ruling out' stroke in acute dizziness.⁶⁸

TESTS PROPERTIES AND TEST-ORDERING: ACCURACY, ACTION THRESHOLDS AND DECISION ANALYSIS

We assume most clinicians intuitively understand the notion that tests are imperfect⁶⁹ and, because of the risk of both false positives (imperfect specificity) and false negatives (imperfect sensitivity), that they should be judicious in choosing and interpreting tests. Physicians probably understand the basic concept of how a test influences an estimate of pretest probability to yield a post-test probability (Bayesian logic⁷⁰) and that obtaining a diagnostic test whose post-test probability could not affect management (ie, could not result in crossing a subsequent test or treatment decision threshold⁷¹ ⁷²) is usually unjustified. Clinicians also presumably grasp that the relative health value of possible downstream outcomes of care following diagnostic tests (health utility) and the probabilities of each potential outcome vary for different patients. They are intuitively aware that these utilities and probabilities can be combined conceptually for sensible decision making about whether or not to test (expected utility analysis for the diagnostic test in that patient).⁶⁹ In practice, however, it is not clear whether these decision-analytic concepts are fully understood or applied by practicing clinicians.⁴⁴ ⁷⁰ ⁷³ ⁷⁴ For example, errors overestimating the capacity of CT to 'rule out' stroke (sensitivity only ~16% in the first 24 h after stroke onset) drive overuse in dizziness.²³

While the real-world behaviour of clinicians may sometimes reflect so-called 'irrational' psychology in decision making,⁴⁵ ⁶⁷ ⁷⁵ the complexity of such diagnostic decisions should not be underestimated. Scientific evidence about diagnostic tests rarely goes beyond the diagnostic accuracy level or immediate effect on diagnostic reasoning or therapeutic decisions,⁷⁶ even though patient-centred outcomes would allow more direct inferences about overall test utility.⁷⁷ There is often combined uncertainty in the estimates of disease probability, test sensitivity and specificity, efficacy of treatment options, and probability and health benefit of the outcomes. This degree of uncertainty usually makes a thorough, quantitative determination of the optimal decision-analytic choice (to test or not to test) during the patient encounter problematic. Future interventions to optimise test

choices for frequently occurring important decisions may need to be mostly preprepared (eg, practice guidelines or computer-based tools; in the dizziness case, perhaps prepackaged decision support using device-based physiologic diagnosis²⁸).

EXPANDED NOTION OF A 'TEST': BEDSIDE EXAMS, CLINICAL PATHWAYS, AND DOCTORS AS 'TESTS'

A medical 'diagnostic test' often connotes paraclinical tests such as blood chemistries or imaging procedures. However, every piece of information acquired during the diagnostic process can be considered a diagnostic test. Each element of history taking and physical examination is a separate diagnostic test with measurable test properties, such as sensitivity, specificity and reproducibility.^{78 79} For continuous (eg, duration of symptoms) or ordinal (eg, severity of cardiac murmur) outcomes, the specific threshold for considering the result abnormal is somewhat arbitrary, and tradeoffs between sensitivity and specificity can be represented by a receiver operating characteristic (ROC) curve.⁸⁰ Also, like laboratory and radiographic tests, bedside 'tests' have an associated cost in physician and patient time and are integral to clinical billing and reimbursement schemas.

When combined together in bundles or sequences, bedside tests are sometimes called clinical decision rules,⁸¹ or clinical algorithms.⁸² These multicomponent assessments, if routinely used together, can be treated mathematically as individual tests.⁸³ For diagnosis of stroke in acute dizziness, a battery of three non-invasive tests of eye movement function known as 'HINTS' has been shown to be 99% sensitive and 97% specific.⁵⁶ This bedside decision rule substantially improves accuracy over the current best available diagnostic alternative (early MRI) in the first 72 h after symptom onset (early MRI sensitivity ~86%), with comparison to delayed, confirmatory MRI.⁵⁶

Similarly, the routine diagnostic practices of individual physicians or groups of physicians can be assessed for their accuracy in terms of sensitivity, specificity and costs in terms of overall diagnostic bedside assessment and test usage behaviour. Representing these clinical performance characteristics in the form of ROC curves makes it clear that the goal of efforts to improve diagnostic accuracy should be to move physician ROC curves towards the upper left corner (ie, maximise the area under the curve, creating 'Deft Diagnosticians') rather than to force physicians to slide upwards along the curve by sacrificing specificity in favour of sensitivity (expensive 'Nervous Nellies'), or the reverse (dangerous 'Crazy Cowboys') (figure 1). Improving diagnostic performance might be achieved through improved diagnostic education⁸⁴ combined with low-technology (eg, test indication curves⁸⁵) or high-technology (eg, computerised diagnostic decision support⁸⁶) tools.⁸⁷ For example, ED physicians may

soon use a novel 'eye ECG' approach to diagnose stroke in dizziness.²⁸

WHAT IS DIAGNOSTIC QUALITY? APPROPRIATE USE AND VALUE IN DIAGNOSIS

High-quality diagnosis is accurate, timely, impactful, patient-centred, ethical and efficient. The importance of accuracy and timeliness to high-quality diagnosis is self-evident. Impact is crucial since 'diagnosis for diagnosis sake,' per se, offers no direct health benefit. Top-quality diagnosis considers individual patient preferences in making difficult decisions about the risks and rewards of resolving diagnostic uncertainty (ie, 'shared decision making'⁸⁸). The ethics of care for specific patients and professional responsibility for stewardship of finite societal healthcare resources should also help shape diagnostic decisions.³⁷ Education⁸⁴ and public awareness campaigns¹⁷ may play an important role in conveying these core values.

Efficiency includes parsimony (conceptual efficiency), speed (temporal efficiency), and cost effectiveness (financial efficiency). Parsimony (eg, taking a 'least moves' strategy to arriving at a correct diagnosis, or choosing not to pursue rare untreatable diagnoses) and speed are attributes we naturally associate with good diagnosticians and good diagnostic process, but it may initially be less intuitive why costs and quality are inseparable for diagnosis. It is our view that profligate, non-parsimonious, inefficient diagnosis, even if accurate, cannot be considered high-quality diagnosis. A physician who orders every imaginable test for every patient with a given symptom would not be considered a high-quality diagnostician—this would indicate an inability to properly judge pretest probabilities at the bedside. By contrast, for the parsimonious physician evaluating a patient with acute dizziness or vertigo, this means the ability to rapidly assess the risk of stroke at the bedside, rather than referring all patients for neuroimaging.

GETTING 'BANG FOR OUR BUCK' IN DIAGNOSIS: THE ROLE OF ECONOMIC ANALYSIS

Generally missing from categorical 'appropriateness' definitions are explicit considerations of the value (in health or dollars) of making a correct diagnosis, the incremental benefit of one diagnostic strategy over another, and the societal opportunity cost of recommending a particular diagnostic test be performed when total healthcare resources are finite. Also usually missing from these 'go'-'no go' assessments of average appropriateness are individual patient disease probabilities or personal preferences (utilities) for specific health outcomes, the psychological impact (positive or negative) of the search for a diagnosis (vs watchful waiting), or the knowledge of the diagnosis itself, and explicit estimation of the effects of uncertainty or risk of bias in the evidence base underlying the overall recommendation.

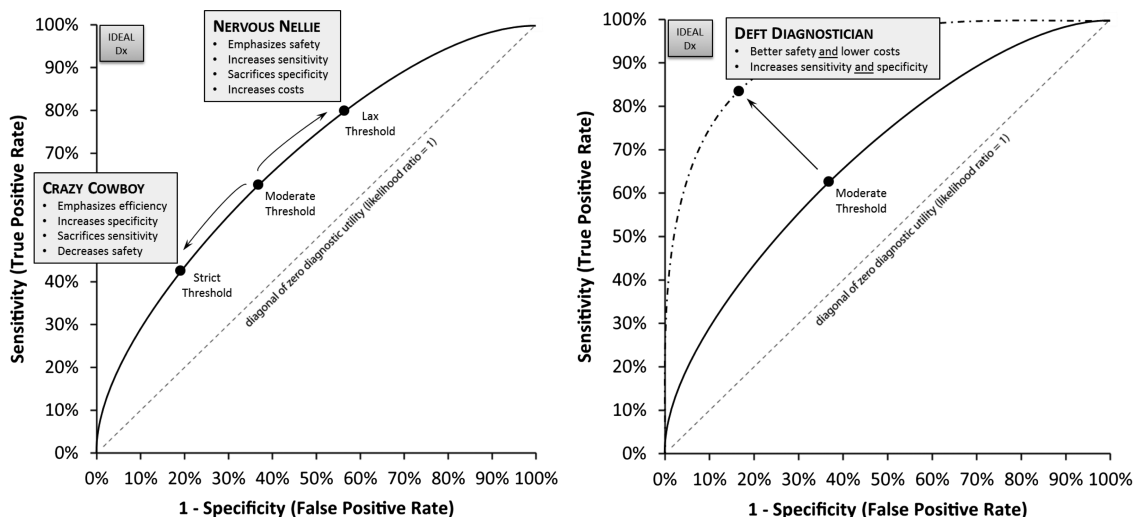


Figure 1 Tradeoffs versus improvements in diagnostic performance as illustrated by the ROC curve. (A) Performance tradeoffs (sliding along the ROC curve). In this scenario, clinicians alter their thresholds for seeking a particular diagnosis, but do not improve their overall diagnostic performance. Incentives, such as fear of malpractice litigation, drive physicians toward being ‘Nervous Nellies’, while incentives, such as productivity pressures, drive them towards being ‘Crazy Cowboys.’ (B) Performance improvements (moving the ROC curve). In this scenario, clinicians alter their diagnostic performance with regard to a particular diagnosis, rather than merely altering their threshold for decision making. Only incentives that focus on cost-effectiveness or ‘value’ (ie, combine ‘quality’ and ‘productivity’ measures) can drive clinicians to become ‘Deft Diagnosticians.’ Dx, diagnosis; ROC, receiver operating characteristic.

Economic modelling and related analytic techniques (cost–benefit, cost–effectiveness, estimated value of information) offer a robust alternative to assess the societal value of medical diagnostic testing, although some special considerations are required. The economic valuation of therapy is relatively straightforward—if a treatment improves health outcomes, its added value can be weighed against its costs in dollars and adverse effects. The value proposition for diagnosis is usually less transparent—diagnosis is more remote to the desired outcome (ie, ‘better health’ not ‘better diagnosis’, *per se*⁷⁷) and, consequently, the link between improved diagnosis of a condition and improved health is more uncertain (eg, overdiagnosis of cancer⁶⁴). For practical reasons, scientific evidence backing the use of diagnostic tests is also usually indirect, requiring a two-step inference that generally assumes, given a correct diagnosis, that the application of correct treatment will result in better outcomes.⁷⁷ Furthermore, there may be benefits to ‘knowing’ about a diagnosis even if there are no immediate treatment implications⁸⁹; and there may also be harms (as with a progressive, untreatable disorder such as Huntington’s disease).⁹⁰ These attributes lead to greater complexity in analytically assessing the value of ordering a diagnostic test, even if familiar qualitative estimates (eg, the ‘chagrin’ of making the wrong choice) are substituted for less familiar quantitative calculations of net benefit.⁴⁴

Conceptual complexities notwithstanding, we believe that as long as the psychological value (benefit or harm) of uncertain, correct and incorrect diagnoses is considered for its impact on health-related quality

of life,⁹¹ then the standard measure of health effect used in economic analyses of medical treatments, the ‘quality-adjusted life year’ (QALY), is an appropriate measure of diagnostic test outcomes. Similarly, the cost per QALY or incremental cost-effectiveness ratio (ICER) is an appropriate measure of societal value in diagnosis. Within this framework, individual diagnostic tests or overall diagnostic strategies which lead to health benefits at a cost of <US\$100 000 per QALY (or any societally sanctioned alternative threshold)^{92–94} would be considered cost effective. In a comparative effectiveness (or relative cost effectiveness) framework, diagnostic interventions offering the greatest number of QALYs per healthcare dollar spent should be endorsed at a societal level. We propose that high-value targets are those with a high burden of harm from misdiagnosis and a low cost of reducing misdiagnosis, while those with the opposite profile are of low value (figure 2).

HOW ECONOMIC ANALYSIS MIGHT HELP IMPROVE DIAGNOSTIC SAFETY: OUR CASE STUDY REVISITED

In the case of acute dizziness⁹⁵ (box 1), where diagnostic accuracy is low, leading to incorrect management and patient harm, it seems intuitive that fixing the quality problem should make economic sense. If new bedside techniques of such high sensitivity and specificity are available, one might just assume that dissemination strategies will improve quality and reduce costs (ie, be cost beneficial). Such an assumption would seem justified since there is strong evidence of unwanted practice variation³² including both

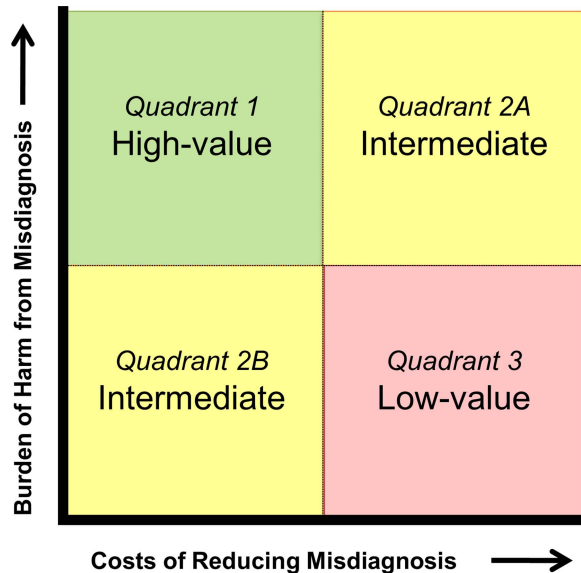


Figure 2 Societal value perspective on misdiagnosis reduction strategies.

overuse and underuse of tests, particularly neuroimaging.¹⁹ The situation suggests a compelling opportunity for resource realignment to reclaim value in diagnosis. Yet, a more nuanced economic model results in a different approach to improving diagnostic safety in dizziness than might initially be expected.

Dizziness patients at high risk for stroke would be a reasonable group to target given advances in understanding bedside stroke diagnosis and the life-threatening, time-dependent nature of the disease. The approximately 15% subgroup of patients with acute, continuous, dizziness or vertigo can be readily identified at the bedside and have about a 25% risk of stroke.²⁵ Modelling different diagnosis options reveals that better bedside diagnosis (with device-based eye movement interpretation using video-oculography (VOG) equipment²⁸) could produce cost-effective quality improvements (<US\$8000 per QALY⁹⁶), but would not be directly cost saving relative to current practice (figure 3). Diagnostic testing is less expensive in the new care approach (neuroimaging would be reduced from ~45% to <15%), but caring for stroke patients costs money in order to save lives—so correctly identifying more strokes costs more than current practice. In the cost-effectiveness framework, the shift in diagnostic strategy is not cost ‘saving’ since the net economic benefit to society of these saved or improved lives⁹⁷ (estimated at about US\$600 million per year in the USA) is not incorporated. Although perhaps counterintuitive, increased diagnostic safety will only rarely produce direct healthcare cost savings when the condition being diagnosed is expensive to treat, even if costs of diagnostic testing decline.

If one also considers, however, the impact of improved diagnosis on patients with benign vestibular disorders, a different cost–benefit picture emerges.

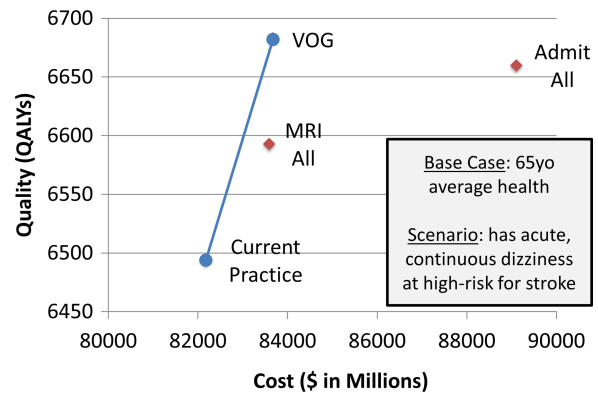


Figure 3 Case example—Improving stroke diagnosis in acute dizziness and vertigo. We modelled cost-effectiveness of a novel eye-movement physiology-based approach,⁹⁶ focusing on variable costs and effects related to diagnosis of strokes among ED patients with acute, continuous dizziness. We compared two blanket diagnostic strategies (MRI all, admit all) to current practice and our proposed strategy (bedside video-oculography (VOG)). We calculated incremental cost-effectiveness ratios (ICERs) from a societal perspective using dollars and quality-adjusted life years (QALYs). VOG could save many lives and is highly cost effective (~US\$7735 per QALY) in this subset of patients. Blue (non-dominated). Red (dominated: higher-cost, lower-quality alternatives).

This population might initially seem less important to target, since lives are not usually at stake when patients with self-limited inner ear conditions are missed. Nevertheless, these ‘benign’ conditions do reduce quality of life for patients, and treatments reclaim these losses.⁹⁸ More importantly, from an economic perspective, the societal costs of unnecessary diagnostic tests or admission for ‘stroke workups’ in these patients are enormous. With appropriate reductions in CT, modest increases in MRI, and slight decreases in overall admissions, we have estimated that total healthcare savings for ED dizziness amount to more than US\$1 billion annually in the USA alone (additional material online table 1).

This finding suggests that an intervention targeting the broader population of patients at risk for both stroke and vestibular disorders would save lives and money. Thus, our economic analysis points to a different population (all acute dizziness or vertigo) than might initially have been targeted based on common-sense approaches (subset at high stroke risk). Before pursuing a research study to prove and disseminate these techniques, it would be possible to use estimated value of perfect information analyses⁹⁹ to measure the expected value to society of any future research efforts to prove and disseminate the new approach.

CONCLUSIONS: SOCIETAL VALUE PRIORITISATION OF DIAGNOSTIC SAFETY AND QUALITY EFFORTS

Using the case example of acute dizziness, we have illustrated the potential benefit of economic analysis for guiding quality improvement approaches targeting

Table 1 Current and projected ED and hospital resource utilisation with routine VOG use

Resource usage (imaging, admission)	Current (2013 US national ^{18 95})	Projection with ED VOG use
All ED dizziness CT rate (%)	41.2	10.3
All ED dizziness MRI rate (%)	2.4	3.0
All ED dizziness admission rate (%)	18.8	17.2
Total ED/hospital workup costs (US\$)	9 242 624 941	8 198 729 820
Projected annual US healthcare savings (US\$)	–	1 043 895 121

Newman-Toker, unpublished. See online supplement for parameters and sources. These numbers are purely resource-use based, and do not consider the societal and personal benefits from lives saved due to early stroke interventions, or reduced morbidity from improved treatment of inner ear disorders. ED, emergency department; VOG, video-oculography.

reduced diagnostic error. From a societal value perspective, the most sensible approach to improving diagnostic safety and quality might be to identify the diagnostic failures (misdiagnoses and inappropriate test use) with the greatest total economic burden for society and target these first for quality assurance or improvement initiatives. While doing this across the spectrum of all conceivable problems and conditions seems daunting, conducting economic analyses for a finite set of important problems (eg, 10 most common presenting symptoms in primary and emergency care; 10 leading causes of morbidity and mortality) could help prioritise high-yield targets.

Economic analyses will only inform diagnostic safety and quality if we can define the necessary parameters to construct the analyses. As diagnostic techniques evolve, it will be critical to study not only diagnostic test properties, but the impact of different diagnostic strategies on health outcomes. Future research should seek to explicitly measure the rates of diagnostic error for common symptoms and important diseases, as well as misdiagnosis-related harms and associated costs. Standards for applying estimated value of perfect information analyses and other advanced techniques to diagnostic problems should be developed to help guide funders in determining the potential societal value of solving a particular diagnostic problem. Stakeholders, including research funding agencies, should make economic analyses priority topics for scientific inquiry related to diagnosis and diagnostic errors.

Collaborators Newman-Toker DE, Butchy GT, Lehmann HP, Aldrich EM, Chanmugam A, Frick KD.

Contributors DEN-T had full access to all the information in the study and takes responsibility for the integrity of the information and the accuracy of the content. DEN-T: conceived manuscript concept; drafted manuscript; reviewed and critically edited the manuscript; approved the final version. KM: helped conceive manuscript concept; reviewed and critically edited the manuscript; approved the final version. DM: helped conceive manuscript concept; reviewed and critically edited the manuscript; approved the final version.

Funding Agency for Healthcare Research and Quality (Grant #R13HS019252).

Competing interests None.

Provenance and peer review Commissioned; externally peer reviewed.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/3.0/>

REFERENCES

- Executive Office of the President Council of Economic Advisers. *The Economic Case for Healthcare Reform*. 2009. http://www.whitehouse.gov/assets/documents/CEA_Health_Care_Report.pdf (accessed 1 Apr 2013).
- Reuben DB, Cassel CK. Physician stewardship of health care in an era of finite resources. *JAMA* 2011;306:430–1.
- National Health Expenditure Projections 2011–2021. Centers for Medicare & Medicaid Services, 2011. <http://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/Downloads/Proj2011PDF.pdf> (accessed 1 Apr 2013).
- Feldman L. Managing the Cost of Diagnosis. *Managed Care* 2009:43–5.
- Iglehart JK. Health insurers and medical-imaging policy—a work in progress. *N Engl J Med* 2009;360:1030–7.
- Rizk E. Molecular diagnostic testing presents \$5 billion conundrum. *Managed Care* 2009.
- Schulman KA, Tunis SR. A policy approach to the development of molecular diagnostic tests. *Nat Biotechnol* 2010;28:1157–9.
- Ensuring Quality through Appropriate Use of Diagnostic Imaging. America's Health Insurance Plans, 2008. <http://www.ahip.org/Ensuring-Quality-Diagnostic-Imaging/> (accessed 1 Apr 2013).
- Graber ML. The incidence of diagnostic error in medicine. *BMJ Qual Saf*. 2013 Aug 7. doi: 10.1136/bmjqs-2012-001615. [Epub ahead of print] PubMed PMID: 23771902.
- Newman-Toker DE, Makary MA. Measuring diagnostic errors in primary care: the first step on a path forward. Comment on "Types and origins of diagnostic errors in primary care settings". *JAMA Intern Med* 2013;173:425–6.
- Saber-Tehrani AS, Lee HW, Matthews SC, et al. 25-year summary of US malpractice claims for diagnostic errors 1986–2010: an analysis from the National Practitioner Data Bank. *BMJ Qual Saf* 2013;22:672–80.
- Mello MM, Chandra A, Gawande AA, et al. National costs of the medical liability system. *Health Affairs* 2010;29:1569–77.
- Leonhardt D. Medical Malpractice System Breeds More Waste. *New York Times* 2009 September 22, 2009. http://www.nytimes.com/2009/09/23/business/economy/23leonhardt.html?_r=0 (accessed 1 Apr 2013).

- 14 Massachusetts Medical Society. Investigation of defensive medicine in Massachusetts, 2008. http://www.massmed.org/AM/Template.cfm?Section=Research_Reports_and_Studies2&TEMPLATE=/CM/ContentDisplay.cfm&CONTENTID=27797 (accessed 1 Apr 2013).
- 15 McQuillan LJ, Abramyan H. *U.S. tort liability index: 2010 report*. San Francisco: Pacific Research Institute, 2010. http://www.doh.state.fl.us/Workforce/Workforce/Council_Materials/Tort_Liability_Index_2010.pdf (accessed 1 Apr 2013).
- 16 Jackson Healthcare. A costly defense: physicians sound off on the high price of defensive medicine in the U.S. http://www.jacksonhealthcare.com/media/8968/defensivemedicine_ebook_final.pdf (accessed 1 Apr 2013).
- 17 Choosing Wisely campaign. American Board of Internal Medicine. <http://www.choosingwisely.org/> (accessed 30 Jul 2013).
- 18 Saber Tehrani AS, Coughlan D, Hsieh YH, *et al*. Rising annual costs of dizziness presentations to US Emergency departments. *Acad Emerg Med* 2013;20:689–96.
- 19 Newman-Toker DE, Camargo CA Jr, Hsieh YH, *et al*. Disconnect between charted vestibular diagnoses and emergency department management decisions: a cross-sectional analysis from a nationally representative sample. *Acad Emerg Med* 2009;16:970–7.
- 20 Kerber KA, Morgenstern LB, Meurer WJ, *et al*. Nystagmus assessments documented by emergency physicians in acute dizziness presentations: a target for decision support? *Acad Emerg Med* 2011;18:619–26.
- 21 Kerber KA, Burke JF, Skolarus LE, *et al*. Use of BPPV processes in emergency department dizziness presentations: a population-based study. *Otolaryngol Head Neck Surg* 2012;148:425–30.
- 22 Kerber KA, Brown DL, Lisabeth LD, *et al*. Stroke among patients with dizziness, vertigo, and imbalance in the emergency department: a population-based study. *Stroke* 2006;37:2484–7.
- 23 Stanton VA, Hsieh YH, Camargo CA Jr, *et al*. Overreliance on symptom quality in diagnosing dizziness: results of a multicenter survey of emergency physicians. *Mayo Clin Proc* 2007;82:1319–28.
- 24 Newman-Toker DE, Stanton VA, Hsieh YH, *et al*. Frontline providers harbor misconceptions about the bedside evaluation of dizzy patients [research letter]. *Acta Otolaryngol* 2008;128:601–4.
- 25 Tarnutzer AA, Berkowitz AL, Robinson KA, *et al*. Does my dizzy patient have a stroke? A systematic review of bedside diagnosis in acute vestibular syndrome. *CMAJ* 2011;183:E571–92.
- 26 Savitz SI, Caplan LR, Edlow JA. Pitfalls in the diagnosis of cerebellar infarction. *Acad Emerg Med* 2007;14:63–8.
- 27 Kattah JC, Talkad AV, Wang DZ, *et al*. HINTS to diagnose stroke in the acute vestibular syndrome: three-step bedside oculomotor examination more sensitive than early MRI diffusion-weighted imaging. *Stroke* 2009;40:3504–10.
- 28 Newman-Toker DE, Saber Tehrani AS, Mantokoudis G, *et al*. Quantitative video-oculography to help diagnose stroke in acute vertigo and dizziness: toward an ECG for the eyes. *Stroke* 2013;44:1158–61.
- 29 Kissick WL. *Medicine's dilemmas: infinite needs versus finite resources*. New Haven: Yale University Press, 1994.
- 30 Reuben DB. Miracles, choices, and justice: tragedy of the future commons. *JAMA* 2010;304:467–8.
- 31 Kerber KA, Schweigler L, West BT, *et al*. Value of computed tomography scans in ED dizziness visits: analysis from a nationally representative sample. *Am J Emerg Med* 2010;28:1030–6.
- 32 Kim AS, Sidney S, Klingman JG, *et al*. Practice variation in neuroimaging to evaluate dizziness in the ED. *Am J Emerg Med* 2012;30:665–72.
- 33 Cutler DM, McClellan M. Is technological change in medicine worth it? *Health affairs* 2001;20:11–29.
- 34 Skinner JS, Staiger DO, Fisher ES. Is technological change in medicine always worth it? The case of acute myocardial infarction. *Health affairs* 2006;25:w34–47.
- 35 Agency for Healthcare Research and Quality. Glossary: Underuse, Overuse, Misuse. *AHRQ Web M&M—morbidity & mortality rounds on the web* http://webmm.ahrq.gov/popup_glossary.aspx?name=underuseoverusemisuse (accessed 1 Apr 2013).
- 36 Cummins RO. Clinicians' reasons for overuse of skull radiographs. *AJR Am J Roentgenol* 1980;135:549–52.
- 37 Brett AS, McCullough LB. Addressing requests by patients for nonbeneficial interventions. *JAMA* 2012;307:149–50.
- 38 Asano TK, Toma D, Stern HS, *et al*. Current awareness in Canada of clinical practice guidelines for colorectal cancer screening. *Can J Surg* 2004;47:104–8.
- 39 Coleman K, Reiter KL, Fulwiler D. The impact of pay-for-performance on diabetes care in a large network of community health centers. *J Health Care Poor Underserved* 2007;18:966–83.
- 40 Brown SR, Brown J. Why do physicians order unnecessary preoperative tests? A qualitative study. *Fam Med* 2011;43:338–43.
- 41 Sirovich BE, Woloshin S, Schwartz LM. Too Little? Too Much? Primary care physicians' views on US health care: a brief report. *Arch Intern Med* 2011;171:1582–5.
- 42 Hillman BJ, Joseph CA, Mabry MR, *et al*. Frequency and costs of diagnostic imaging in office practice—a comparison of self-referring and radiologist-referring physicians. *N Engl J Med* 1990;323:1604–8.
- 43 Pines JM, Isserman JA, Szyld D, *et al*. The effect of physician risk tolerance and the presence of an observation unit on decision making for ED patients with chest pain. *Am J Emerg Med* 2010;28:771–9.
- 44 Feinstein AR. The 'chagrin factor' and qualitative decision analysis. *Arch Intern Med* 1985;145:1257–9.
- 45 Hozo I, Djulbegovic B. When is diagnostic testing inappropriate or irrational? Acceptable regret approach. *Med Decis Making* 2008;28:540–53.
- 46 Kassirer JP. Our stubborn quest for diagnostic certainty. A cause of excessive testing. *N Eng J Med* 1989;320:1489–91.
- 47 Studdert DM, Mello MM, Sage WM, *et al*. Defensive medicine among high-risk specialist physicians in a volatile malpractice environment. *JAMA* 2005;293:2609–17.
- 48 Chou R, Qaseem A, Snow V, *et al*. Diagnosis and treatment of low back pain: a joint clinical practice guideline from the American College of Physicians and the American Pain Society. *Ann Intern Med* 2007;147:478–91.
- 49 Berwick DM, Coltin KL. Feedback reduces test use in a health maintenance organization. *JAMA* 1986;255:1450–4.
- 50 Griner PF. Payment reform and the mission of academic medical centers. *N Engl J Med* 2010;363:1784–6.
- 51 Outpatient imaging efficiency measures: Medicare.gov, 2012. <http://www.medicare.gov/HospitalCompare/Data/ImagingEfficiency/Outpatient-Measures.aspx> (accessed 5 Jan 2013).
- 52 DeKay ML, Asch DA. Is the defensive use of diagnostic tests good for patients, or bad? *Med Decis Making* 1998;18:19–28.

- 53 Sarkar U, Bonacum D, Strull W, *et al.* Challenges of making a diagnosis in the outpatient setting: a multi-site survey of primary care physicians. *BMJ Qual Saf* 2012;21:641–8.
- 54 Ammenwerth E, Spottl HP. The time needed for clinical documentation versus direct patient care. A work-sampling analysis of physicians' activities. *Methods Inf Med* 2009;48:84–91.
- 55 Graber M, Gordon R, Franklin N. Reducing diagnostic errors in medicine: what's the goal? *Acad Med* 2002;77:981–92.
- 56 Newman-Toker DE, Kerber KA, Hsieh YH, *et al.* HINTS outperforms ABCD2 to screen for stroke in acute continuous vertigo and dizziness. *Acad Emerg Med* 2013 (in press).
- 57 Singh H, Giardina TD, Meyer AN, *et al.* Types and origins of diagnostic errors in primary care settings. *JAMA Intern Med* 2013;173:418–25.
- 58 Zwaan L, de Bruijne M, Wagner C, *et al.* Patient record review of the incidence, consequences, and causes of diagnostic adverse events. *Arch Intern Med* 2010;170:1015–21.
- 59 Graber M. Diagnostic errors in medicine: a case of neglect. *Jt Comm J Qual Patient Saf* 2005;31:106–13.
- 60 Newman-Toker DE, Robinson KA, Edlow JA. Frontline misdiagnosis of cerebrovascular events in the era of modern neuroimaging: a systematic review [abstract]. *Ann Neurol* 2008;64(Suppl 12):S17–18.
- 61 Lever NM, Nyström KV, Schindler JL, *et al.* Missed Opportunities for Recognition of Ischemic Stroke in the Emergency Department. *J Emerg Nurs*. 2012 May 25. [Epub ahead of print] PubMed PMID: 22633790.
- 62 Kostopoulou O, Delaney BC, Munro CW. Diagnostic difficulty and error in primary care—a systematic review. *Fam Pract*. 2008;25:400–13.
- 63 Schiff GD, Hasan O, Kim S, *et al.* Diagnostic error in medicine: analysis of 583 physician-reported errors. *Arch Intern Med* 2009;169:1881–7.
- 64 Welch HG, Black WC. Overdiagnosis in cancer. *J Natl Cancer Inst* 2010;102:605–13.
- 65 Orme NM, Fletcher JG, Siddiki HA, *et al.* Incidental findings in imaging research: evaluating incidence, benefit, and burden. *Arch Intern Med* 2010;170:1525–32.
- 66 Becker MC, Galla JM, Nissen SE. Left main trunk coronary artery dissection as a consequence of inaccurate coronary computed tomographic angiography. *Arch Intern Med* 2011;171:698–701.
- 67 Redelmeier DA. Improving patient care. The cognitive psychology of missed diagnoses. *Ann Intern Med* 2005;142:115–20.
- 68 Edlow JA, Newman-Toker DE, Savitz SI. Diagnosis and initial management of cerebellar infarction. *Lancet Neurol* 2008;7:951–64.
- 69 Bernstein J. Decision analysis. *J Bone Joint Surg Am* 1997;79:1404–14.
- 70 Bianchi MT, Alexander BM. Evidence based diagnosis: does the language reflect the theory? *BMJ* 2006;333:442–5.
- 71 Pauker SG, Kassirer JP. The threshold approach to clinical decision making. *N Engl J Med* 1980;302:1109–17.
- 72 Warner JL, Najarian RM, Tierney LM Jr. Perspective: Uses and misuses of thresholds in diagnostic decision making. *Acad Med* 2010;85:556–63.
- 73 Moskowitz AJ, Kuipers BJ, Kassirer JP. Dealing with uncertainty, risks, and tradeoffs in clinical decisions. A cognitive science approach. *Ann Intern Med* 1988;108:435–49.
- 74 Reid MC, Lane DA, Feinstein AR. Academic calculations versus clinical judgments: practicing physicians' use of quantitative measures of test accuracy. *Am J Med* 1998;104:374–80.
- 75 Payne VL, Crowley RS. Assessing the use of cognitive heuristic representativeness in clinical reasoning. *AMIA Annu Symp Proc* 2008:571–5.
- 76 Pearl WS. A hierarchical outcomes approach to test assessment. *Ann Emerg Med* 1999;33:77–84.
- 77 Schunemann HJ, Oxman AD, Brozek J, *et al.* Grading quality of evidence and strength of recommendations for diagnostic tests and strategies. *BMJ* 2008;336:1106–10.
- 78 Jaeschke R, Guyatt GH, Sackett DL. Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *JAMA* 1994;271:703–7.
- 79 Simel DL, Rennie D, Bossuyt PM. The STARD statement for reporting diagnostic accuracy studies: application to the history and physical examination. *J Gen Intern Med* 2008;23:768–74.
- 80 Metz CE. Basic principles of ROC analysis. *Semin Nucl Med* 1978;8:283–98.
- 81 Stiell IG, Wells GA. Methodologic standards for the development of clinical decision rules in emergency medicine. *Ann Emerg Med* 1999;33:437–47.
- 82 Feinstein AR. An analysis of diagnostic reasoning. 3. The construction of clinical algorithms. *Yale J Biol Med* 1974;47:5–32.
- 83 McGinn TG, Guyatt GH, Wyer PC, *et al.* Users' guides to the medical literature: XXII: how to use articles about clinical decision rules. Evidence-Based Medicine Working Group. *JAMA* 2000;284:79–84.
- 84 Graber ML. Educational strategies to reduce diagnostic error: can you teach this stuff? *Adv Health Sci Educ* 2009;14(Suppl 1):63–9.
- 85 Bernstein J. Test-indication curves. *Med Decis Making* 1997;17:103–6.
- 86 Singh H, Graber ML, Kissam SM, *et al.* System-related interventions to reduce diagnostic errors: a narrative review. *BMJ Qual Saf* 2012;21:160–70.
- 87 Newman-Toker DE, Pronovost PJ. Diagnostic errors—the next frontier for patient safety. *JAMA* 2009;301:1060–2.
- 88 Davey HM, Barratt AL, Davey E, *et al.* Medical tests: women's reported and preferred decision-making roles and preferences for information on benefits, side-effects and false results. *Health Expect* 2002;5:330–40.
- 89 Neumann PJ, Cohen JT, Hammitt JK, *et al.* Willingness-to-pay for predictive tests with no immediate treatment implications: a survey of US residents. *Health Econ* 2012;21:238–51.
- 90 Coustasse A, Pekar A, Sikula A, *et al.* Ethical considerations of genetic presymptomatic testing for Huntington's disease. *J Hosp Mark Public Relations* 2009;19:129–41.
- 91 Salz T, Richman AR, Brewer NT. Meta-analyses of the effect of false-positive mammograms on generic and specific psychosocial outcomes. *Psychooncology* 2010;19:1026–34.
- 92 Ubel PA, Hirth RA, Chernew ME, *et al.* What is the price of life and why doesn't it increase at the rate of inflation? *Arch Intern Med* 2003;163:1637–41.
- 93 Grosse SD. Assessing cost-effectiveness in healthcare: history of the \$50,000 per QALY threshold. *Expert Rev Pharmacoecon Outcomes Res* 2008;8:165–78.
- 94 Hirth RA, Chernew ME, Miller E, *et al.* Willingness to pay for a quality-adjusted life year: in search of a standard. *Med Decis Making* 2000;20:332–42.
- 95 Newman-Toker DE, Hsieh YH, Camargo CA Jr, *et al.* Spectrum of dizziness visits to US emergency departments:

- cross-sectional analysis from a nationally representative sample. *Mayo Clin Proc* 2008;83:765–75.
- 96 Newman-Toker DE, Butchy GT, Lehmann HB, *et al.* Diagnostic decision support to reduce stroke misdiagnosis among acutely dizzy patients: a cost effectiveness analysis [abstract]. *Neurology* 2009;72(11 Suppl 3):A185–A85.
- 97 Dranove D. *What's your life worth?: health care rationing— who lives? who dies? who decides?* Upper Saddle River, NJ: FT Prentice Hall, 2003.
- 98 Lopez-Escamez JA, Gamiz MJ, Fernandez-Perez A, *et al.* Long-term outcome and health-related quality of life in benign paroxysmal positional vertigo. *Eur Arch Otorhinolaryngol* 2005;262:507–11.
- 99 Meltzer D. Addressing uncertainty in medical cost-effectiveness analysis implications of expected utility maximization for methods to perform sensitivity analysis and the use of cost-effectiveness analysis to set priorities for medical research. *J Health Econ* 2001;20:109–29.