Getting the Gist of Health Risks

Tanner J. Caverly, MD, Allan Prochazka, MD, MSc, Ingrid Binswanger, MD, MPH, Jean S. Kutner, MD, MSPH and Dan Matlock, MD, MPH
University of Colorado School of Medicine, Division of General Internal Medicine, Denver, CO.

Abstract

Many clinicians admit to having difficulty with numbers. When tested on their ability to interpret health statistics they have performed poorly. There is increasing awareness that numeracy, a person's ability to draw meaning from numbers, is important in medical decision making and central to clinical judgment. Three powerful forces combine to promote numeracy as a core competency of modern patient care: 1) the consensus around informed patient decision-making; 2) the ascendency of evidence-based medicine; and 3) the increasing focus on high-value care. Nonetheless, interpreting health statistics receives scant attention in medical education and quality improvement efforts.

Several key statistical skills such as understanding the difference between relative and absolute risk and understanding that prevalence of a disease influences the value of a test have been shown to be important for clinical judgment. We call these skills - and other quantitative skills fundamental to patient care - Clinician Numeracy (CN). CN can be defined as "the ability to use numbers and numeric concepts in the context of taking care of patients." CN includes 3 domains: primary numeric skills, interpretive numeracy, and applied numeracy. CN is fundamental to the practice of everyday medicine because it affects the conclusions clinicians draw from the tests they order and the medical literature they read. The applied numeracy domain deals directly with patient care tasks. These tasks include: 1. risk communication tasks such as using numbers to communicate probabilistic information about potential harms and benefits to patients and 2. decision making tasks like balancing harms and benefits to make a given medical decision. Unfortunately, we cannot currently assess CN because there is no valid measure to test these skills in clinicians. Indeed, measuring the range of skills inherent in the domains of CN may be too much for a single measure. Nonetheless, a way of assessing these skills would be useful. A valid measure would help determine to what extent CN affects clinical performance and provide a means of testing improvement in these skills with educational interventions.

Fortunately, a modern theory of medical decision making with a growing body of support (fuzzy trace theory) provides a potential avenue of assessing essential numeracy skills. This theory asserts that medical decisions are most often based on gist-based intuition or "simple, bottom-line representations of the meaning of information or experience." We have developed a conceptual model utilizing this theory to show how CN might be related to medical decision making and health outcomes via the 'risk gist.' We present preliminary work on the development of the Critical Risk Interpretation Test which seeks to evaluate the appropriateness of a clinician's risk gist. Finally, we discuss how the conceptual framework we have developed allows us to test hypotheses regarding the effect of risk gist on risk communication tasks and medical decision making.

Key Words: numeracy, statistical literacy, health statistics, medical education

I. Introduction

The consensus around informed patient decision making, the importance of evidence based medicine, and the growing focus on practicing high-value care all increase the need for clinician mastery of basic statistical skills. A recent paper by an American College of Physicians workgroup demonstrates the importance of statistical skills for practicing high-value care.[1] The group emphasized that reducing low value testing was a national priority and urged physicians to more carefully weigh the benefits, harms, and downstream costs of testing. To help determine the value of a test, the group highlighted the importance of understanding a *basic statistical principle*: prevalence affects the predictive value of a test. Many of the testing scenarios on the group's 'little or no value' list involve ordering tests in a low prevalence population where false-positives are a problem. This example demonstrates how statistical skill can be central to clinical judgment and its absence a fundamental problem in the quality of healthcare. But physicians struggle when interpreting basic health statistics.[2-11] When 85 physicians were asked basic questions - to convert a decimal to a percent, a percent to a decimal, and how many heads in 1,000 coin flips - only 72% answered all three correctly.[10] More recently, 76% of a nationally representative sample of physicians believed that if cancer screening improved 5-year survival rates this proved that the screening saved lives.[11]

No studies have attempted to measure the effect of CN on clinical performance. The few relevant measures are too long to implement in busy practice settings or were developed for patients and therefore overly simplistic.[12-16] Given what we now know about how *patient* numeracy affects behaviors and outcomes[17-20] the need to measure fundamental numeracy skills among *clinicians* is clear. We utilized a conceptual framework for the CRIT and extensively reviewed the literature to find items from existing measures which fit into this framework. Because no adequately validated measure existed, we developed an instrument by using relevant existing items and by writing new items when needed. New items were iteratively modified and pilot tested. Our survey instrument is now undergoing validation. A validated metric is critical for future research investigating how these skills impact clinical performance. A measure is also necessary to evaluate whether educational interventions improve these important skills.

1.1 Context and Significance

Interpreting numerical data has become a core competency of modern patient care for three reasons:

1. <u>*Informed Patient Decision Making*</u>: There is a growing consensus about the importance of informed patient decision making.[16] The need for health professionals to clearly communicate health information to the public has been called an ethical imperative.[21] There is increasing recognition that *numerical* information is an essential part of this communication[22] and that understanding numerical information is essential for making health decisions.[17-20] Although patient preferences for participation are variable, many want detailed information and some indicate a strong preference for active participation in decision making.[23-24] Despite this, there is evidence that patients are not incorporated to the level that they desire,[25-26] which may stem from lack of skill on the part of clinicians. Patients need clinicians skilled in communicating *risk* and *uncertainty* – both quantitative concepts – to help them understand and apply health information to their context. The fact that many patients have difficulty understanding medical data strongly reinforces the importance of clinician mastery in these domains. However, we cannot be certain that more numerate clinicians are better at involving patients in decision making without a validated measure of CN to test that claim.

2. *Evidence-Based Medicine (EBM)*: Increasing emphasis on practicing evidence-based medicine further reinforces the need to efficiently interpret numerical data from research literature. For example, clinicians must be able to recognize the difference between relative risk and absolute risk when reading about treatment benefits. Skillfully applying this data to individual patients *is* evidence-based medicine.[27] Current evidence indicates that clinicians struggle with these basic numeracy tasks.[10] We cannot determine whether more numerate clinicians practice more individualized evidence-based medicine until we have a measure of CN.

3. *High-Value, Cost-Conscious Care*: At least some of the responsibility for increasing medical costs in the US lies with practicing clinicians.[28] Regional variation research suggests that almost one third of health care costs go toward interventions and technology that do not improve outcomes.[29] Also, physicians in lower cost regions of the US practice lower intensity medicine without compromising care.[30] The sixth edition of the ACP ethics manual identifies responsible stewardship of resources as an ethical responsibility of clinicians.[31] This stewardship requires that clinicians grasp and weigh numerical concepts – the *likelihood* and *magnitude* of the potential benefits and harms that result from their decision making. We cannot assess whether more numerate physicians practice higher value care until we can measure CN.

2. Construct Framework

The initial goal of this project is to validate a novel test we have developed - the Critical Risk Interpretation Test (CRIT). The test should be able to identify the interpretive skill levels among highly competent health professionals (discriminability).

A previous definition of health numeracy[32] can be modified to define clinical numeracy as "the ability to use numbers and numeric concepts in the context of taking care of patients." CN includes 3 domains: primary numeric skills, interpretive numeracy, and applied numeracy (Table 1).[32] The applied numeracy domain deals directly with patient care tasks.

Our goal was to identify key numeric concepts within this overall framework which were

Table 1. Construct Framework for Clinical Numeracy (modified from Schapira et. al.[32])

Domain Definitions	Sub-Categories of Domain
Primary Numeric Skills Ability to use basic arithmetic functions and graphs as well as perform basic calculations	•Counting •Basic math functions •Calculate absolute and relative risk, nnt •Scales and graphs
Interpretive Numeracy The ability to understand the strengths and limitations of numbers to represent health or disease states, the efficacy of an intervention, or other expected health outcomes.	•Probability and chance •Principles of scientific method •Concept of uncertainty •Graphic and verbal formats •Individual and biologic variation in expected outcomes •Estimation and sense of magnitude
Applied Numeracy *Basic Tasks* The use of numbers in day to day health care tasks. *Risk Communication Tasks* The use of numbers to communicate probabilistic information about health outcomes including risk, severity, and outcomes of disease. *Decision-Making Tasks* The use of numbers to help consider the risks and benefits of a given medical decision.	•Interpretation of lab values •Discuss and calculate medication dosage •Use prognostic or diagnostic tools •Disease Incidence •Risk factor modification •Prognosis, Survival •Information seeking •Balancing risks and benefits •Assessment of evidence •Estimation and sense of magnitude

both commonly misinterpreted by clinicians and fundamental to risk communication and medical decision-making tasks. Many skills such as calculating absolute or relative risk and discussing medication dosage do not present major challenges for most clinicians – a fact corroborated with our pre-testing and piloting of multiple items within the primary numeric and interpretive numeracy domains. To identify key numeric concepts we undertook an extensive review of research that spans disciplines – psychology, philosophy of science, behavioral economics, statistics, epidemiology, medical education, evidence-based medicine, risk perception, and risk communication. Overall, this search identified the basic components of a risk – known as cumulative incidence in epidemiology – as well as how the risk is framed and how the risk is modified after diagnostic testing – as commonly misinterpreted and miscommunicated.[3-7,9-11,33] The components of a risk are: the type of risk, the timeframe, and the denominator. We then developed the testing framework in Table 2 below.

Table 2: Testing Framework

Framework – Critical Risk Interpretation Test (CRIT)	
Test Category	**Ideal response to test item**
1. Definition:	
Risk of what? Or Reduced risk of what?	• Give more importance to the risk of dying from a disease than the risk of getting disease • Select a larger risk estimate for deaths from all causes than death from a specific disease • Give more weight to a reduction in mortality from any cause than disease-specific mortality • Give more weight to a reduction in a patient-oriented end-point than a surrogate or a combined end-point.
Risk out of how many?	• Give equal weight to an absolute number in the US population than to a proportion when they are the same.
Risk over what time frame?	• Give more weight to risk over the next 10 years than to risk over a lifetime when they are equivalent values • Select a larger risk estimate for a 20-year than 10-year risk
2. Framing:	
How is the risk framed?	• Give equal weight to an absolute risk reduction, a number needed to treat, and a relative risk reduction when they are equivalent • Give equal weight to positively framed risk and negatively framed risk when they are the same
3. Testing:	
How does testing modify the risk?	• Understand that disease detection is not the same as disease prevention • For a screening test, give more weight to mortality reduction than improvement in 5-year survival rates or increased detection rates. • Understand that prevalence affects the predictive value of a test • Understand that changing a test-cutoff affects the number of false-positives and false negatives

3. Fuzzy Trace Theory – a guide for how to test clinical numeracy

Fuzzy Trace Theory is an evidence-based theory of how we make day-to-day medical decisions.[34] A major component of this theory is that decision-makers rarely use precise numbers in decision-making. If numbers affect medical decisions at all, it is because we encode our interpretation of the numbers into a gist – a bottom-line fuzzy sort of meaning a clinician has in his or her head about the nature and magnitude of a health risk. This is because our memory of the literal facts such as an exact number or date tends to slip away with time. What remains for a longer time is our gist of the number. And this is what we retrieve when we have to make a decision.

For example, the exact number of deaths from lung cancer every year - 160,000 – is unlikely to be used directly in a decision to screen for lung cancer. A clinician may only remember something fuzzy such as "a lot of lung cancer" or "an important amount of lung cancer". Based on such a gist a clinician may be more likely to recommend a screening test for lung cancer.

In creating the CRIT, our goal was to evaluate at a level where mistakes would have the biggest impact on day-to-day medical decision making. Thus, our test attempts to evaluate the gist that a clinician encodes from looking at a particular health risk statistic. We call this gist the 'risk gist.' The risk gist can be influenced by a number of factors related to a number as well as person-related and context-related factors.[10,35] A detailed analysis of this literature is beyond the scope of the current paper. However, factors influencing the gist of a health risk statistic specific to how the statistic is defined and framed – and the subsequent effect these factors have on a clinician's interpretation – are precisely what we have sought to measure with the CRIT. In other words, the CRIT was intentionally created to exploit some of the known irrational influences affecting a clinician's risk gist.

For example, the framing of a treatment benefit in terms of a relative risk reduction versus an absolute risk reduction is known to affect the way a clinician (and everyone else) interprets that statistic. Instead of testing whether clinicians can calculate relative and absolute risks or convert between the two we did something different. We sought to assess the gist that the clinician assigns to a treatment benefit framed in terms of relative risk reduction(e.g. 'important'). Later in the test – separated by several unrelated questions – we assess the gist that same clinician assigned to a numerically equivalent treatment benefit framed in terms of an absolute risk reduction. In this way, we attempt to evaluate the accuracy of a clinician's risk gist. If the gist from seeing the statistic in terms of relative risk reduction is the same as the gist from seeing the statistic framed in terms of a numerically equivalent absolute risk, then the gist can be said to be accurate. The more a clinician's gist is modified based on the framing, the more inaccurate the gist.

The ability to calculate relative and absolute risks and convert between the two is a prerequisite for an accurate risk gist but it is not sufficient. The clinician must also know that the way a statistic is framed can affect interpretation and must be in the habit of re-framing the statistic so their gist is the same no matter which way the statistic is framed initially. In other words, the clinician must be a critical reader of the statistics they come across. The more knowledgeable a clinician is about the factors that can influence interpretation of a risk statistic (Table 2) and the more critical a clinician is of the statistics they consume, the better they will score on the CRIT.

The CRIT is designed to identify clinicians who recognize these facts and have developed the necessary skills and habits to overcome the influences of how a risk is defined and framed. A clinician that scores highly on our test can be said to have a numerate risk gist – a gist that is immune to framing effects, appropriately modified based on the type of risk, and which is Bayesian. If the CRIT is a valid and reliable test, we can then evaluate whether clinicians with a numerate risk gist make different types of decisions and communicate differently to patients about specific health risks.

4. Model of how CN might affect decisions and health outcomes

In our model (Figure 1), primary numeric or basic math skills are the foundation for interpretive numeracy. Interpretive numeracy involves an ability to understand the strengths and limitations of numbers within a given context such as a journal article or a news report. Many problems arise from misinterpretation of the evidence. Nonetheless, interpretive numeracy undergirds applied numeracy tasks – like risk communication and the balancing of risks and benefits. This applied numeracy then influences the decisions made and ultimately health outcomes.

For most decisions, fuzzy-trace theory tells us that there is often no direct link between analysis (supported by primary numeric skills and interpretive skills) and everyday applied tasks such as risk communication and decision making. However, basic and interpretive skills can inform the 'risk gist.' In this way, evidence and data might 'color' the decision making. In other words, evidence-based decision making depends on having a numerate risk gist. Thus, improved numeracy can lead to improved risk communication and decision making via a numerate risk gist.

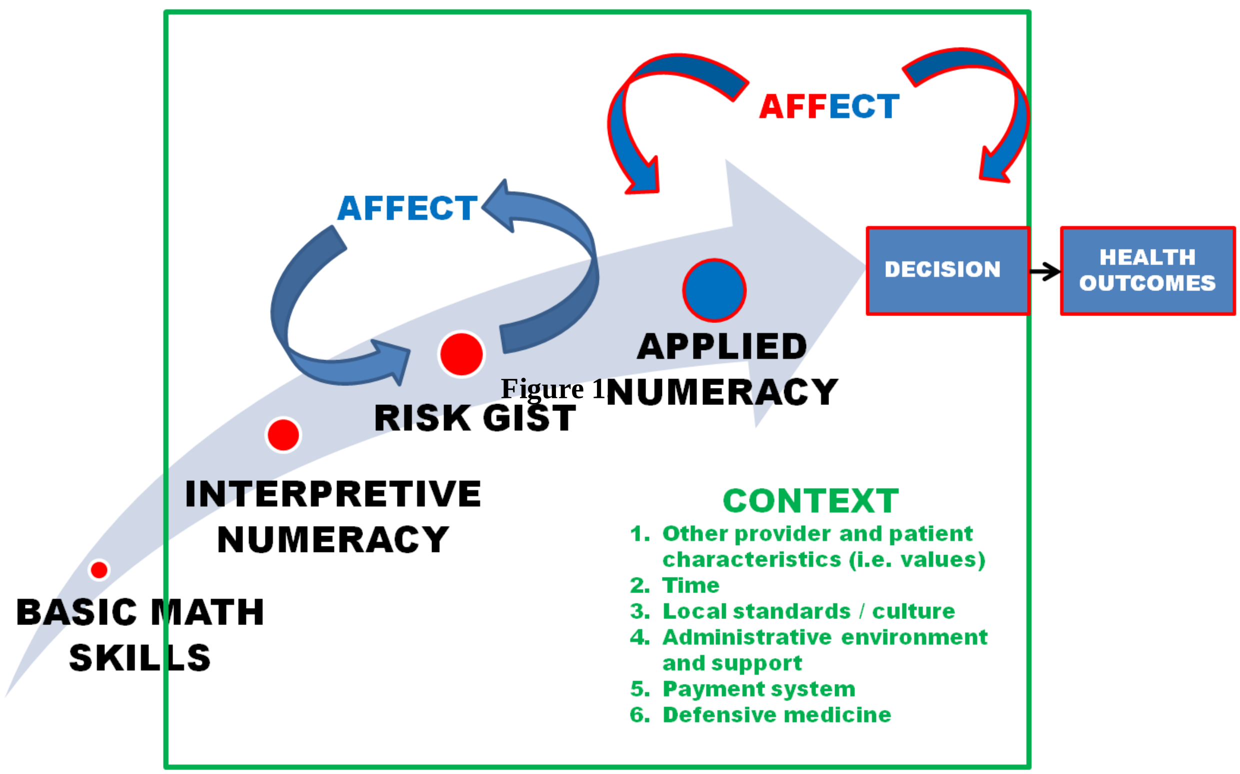

Figure 1. Model of how numeracy might affect decisions and health outcomes

Affect is a term in cognitive psychology referring to the experience of feeling or emotion.[35] It has been shown to have a powerful influence on decision making. In our model, affect can strongly influence the risk gist. For example, simply reading the word ‘cancer’ as part of a health statistic – ‘there are 160,000 cancer deaths each year’ – can generate emotional cues that greatly impact the risk gist. Similarly, affect as well as other contextual factors can influence the way a clinician communicates to patients and the way decisions are made. On the other hand, numeracy has the potential to impact a person’s affect. For instance, knowing that the probability of a fatal airplane crash is highly unlikely can calm the nerves of a flight passenger who otherwise might be anxious about flying.

This model can be a guide to studying the influence that clinical numeracy has on risk communication and day-to-day medical decisions. The model also hypothesizes that certain basic math and interpretive numeracy skills can influence risk gist and even a clinician’s affect. These are testable hypotheses and are part of our research agenda.

5. Future directions

We have already undertaken a cross-sectional validation study to achieve the following specific aims:

1. To formally establish content validity through expert consensus around our framework of critical risk interpretation and the CRIT we have developed.
2. To seek additional evidence of construct validity by assessing the basic attributes, reliability, and criterion validity of the CRIT.
3. To test the hypothesis that more training will lead to higher scores on the CRIT.
4. To test the hypothesis that improved primary numeric and interpretive numeracy skills lead to higher scores on the CRIT.

The CRIT, if it is a valid and reliable test, will allow implementation of a long-term agenda of planned projects to evaluate the effect of clinical numeracy on informed patient decision making, evidence based practice, and high-value, cost-conscious care.

References:

1. Qaseem, Amir, Patrick Alguire, Paul Dallas, Lawrence E Feinberg, Faith T Fitzgerald, Carrie Horwitch, Linda Humphrey, et al. “Appropriate Use of Screening and Diagnostic Tests to Foster High-Value, Cost-Conscious Care.” *Annals of Internal Medicine* 156, no. 2 (January 17, 2012): 147–149.
2. Berwick, D. M, H. V Fineberg, and M. C Weinstein. “When Doctors Meet Numbers.” *The American Journal of Medicine* 71, no. 6 (1981): 991–998.
3. Wulff, H R, B Andersen, P Brandenhoff, and F Guttler. “What Do Doctors Know About Statistics?” *Statistics in Medicine* 6, no. 1 (February 1987): 3–10.
4. Fahey, T., S. Griffiths, and T. J. Peters. “Evidence Based Purchasing: Understanding Results of Clinical Trials and Systematic Reviews.” *BMJ : British Medical Journal* 311, no. 7012 (October 21, 1995): 1056–1060.
5. Hoffrage, U, and G Gigerenzer. “Using Natural Frequencies to Improve Diagnostic Inferences.” *Academic Medicine: Journal of the Association of American Medical Colleges* 73, no. 5 (May 1998): 538–540.
6. Gigerenzer, G., U. Hoffrage, and A. Ebert. “AIDS Counselling for Low-risk Clients.” *AIDS Care* 10, no. 2 (April 1998): 197–211.
7. Sheridan, Stacey L, and Michael Pignone. “Numeracy and the Medical Student’s Ability to Interpret Data.” *Effective Clinical Practice: ECP* 5, no. 1 (February 2002): 35–40.
8. Windish, Donna M., Stephen J. Huot, and Michael L. Green. “Medicine Residents’ Understanding of the Biostatistics and Results in the Medical Literature.” *JAMA: The Journal of the American Medical Association* 298, no. 9 (2007): 1010 –1022.
9. Perneger, Thomas V., and Thomas Agoritsas. “Doctors and Patients’ Susceptibility to Framing Bias: A Randomized Trial.” *Journal of General Internal Medicine* (July 2011).
10. Gigerenzer, G., W. Gaissmaier, E. Kurz-Milcke, L.M. Schwartz, and S. Woloshin. “Helping Doctors and Patients Make Sense of Health Statistics.” *Psychological Science in the Public Interest* 8, no. 2 (2007): 53.
11. Wegwarth, Odette, Lisa M Schwartz, Steven Woloshin, Wolfgang Gaissmaier, and Gerd Gigerenzer. “Do Physicians Understand Cancer Screening Statistics? A National Survey of Primary Care Physicians in the United States.” *Annals of Internal Medicine* 156, no. 5 (March 6, 2012): 340–349.
12. Shaneyfelt, Terrence, Karyn D Baum, Douglas Bell, David Feldstein, Thomas K Houston, Scott Kaatz, Chad Whelan, and Michael Green. “Instruments for Evaluating Education in Evidence-Based Practice.” *JAMA: The Journal of the American Medical Association* 296, no. 9 (September 6, 2006): 1116–1127.
13. Schwartz, L. M, S. Woloshin, W. C Black, and H. G Welch. “The Role of Numeracy in Understanding the Benefit of Screening Mammography.” *Annals of Internal Medicine* 127, no. 11 (1997): 966.
14. Lipkus, Isaac M., Greg Samsa, and Barbara K. Rimer. “General Performance on a Numeracy Scale Among Highly Educated Samples.” *Medical Decision Making* 21, no. 1 (February 1, 2001): 37 –44.
15. Fagerlin, Angela, Brian J. Zikmund-Fisher, Peter A. Ubel, Aleksandra Jankovic, Holly A. Derry, and Dylan M. Smith. “Measuring Numeracy Without a Math Test: Development of the Subjective Numeracy Scale.” *Medical Decision Making* 27, no. 5: 672 –680.
16. Schwartz, Lisa M., Steven Woloshin, and H. Gilbert Welch. “Can Patients Interpret Health Information? An Assessment of the Medical Data Interpretation Test.” *Medical Decision Making* 25, no. 3: 290 –300.
17. Peters, E., J. Hibbard, P. Slovic, and N. Dieckmann. “Numeracy Skill And The Communication, Comprehension, And Use Of Risk-Benefit Information.” *Health Affairs* 26, no. 3 (May 2007): 741–748.

18. Reyna, Valerie F., Wendy L. Nelson, Paul K. Han, and Nathan F. Dieckmann. "How Numeracy Influences Risk Comprehension and Medical Decision Making." *Psychological Bulletin* 135, no. 6 (November 2009): 943–973.
19. Peters, Ellen, Daniel Västfjäll, Paul Slovic, C.K. Mertz, Ketti Mazzocco, and Stephan Dickert. "Numeracy and Decision Making." *Psychological Science (Wiley-Blackwell)* 17, no. 5 (May 2006): 407–413.
20. Nelson, Wendy, Valerie F. Reyna, Angela Fagerlin, Isaac Lipkus, and Ellen Peters. "Clinical Implications of Numeracy: Theory and Practice." *Annals of Behavioral Medicine* 35, no. 3 (May 2008): 261–274.
21. Gazmararian, Julie A., James W. Curran, Ruth M. Parker, Jay M. Bernhardt, and Barbara A. DeBuono. "Public Health Literacy in America: An Ethical Imperative." *American Journal of Preventive Medicine* 28, no. 3 (April 2005): 317–322.
22. Apter, Andrea J., Michael K. Paasche-Orlow, Janine T. Remillard, Ian M. Bennett, Elana Pearl Ben-Joseph, Rosanna M. Batista, James Hyde, and Rima E. Rudd. "Numeracy and Communication with Patients: They Are Counting on Us." *Journal of General Internal Medicine* 23, no. 12 (December 2008): 2117–2124.
23. Say, Rebecca, Madeleine Murtagh, and Richard Thomson. "Patients' Preference for Involvement in Medical Decision Making: A Narrative Review." *Patient Education and Counseling* 60, no. 2 (February 2006): 102–114.
24. Auerbach, Stephen M. "Do Patients Want Control over Their Own Health Care? A Review of Measures, Findings, and Research Issues." *Journal of Health Psychology* 6, no. 2 (March 1, 2001): 191–203.
25. Coulter, A. "Do Patients Want a Choice and Does It Work?" *BMJ* 341, no. oct14 2 (October 14, 2010): c4989–c4989.
26. Deber, Raisa B, Nancy Kraetschmer, Sara Urowitz, and Natasha Sharpe. "Do People Want to Be Autonomous Patients? Preferred Roles in Treatment Decision-making in Several Patient Populations." *Health Expectations* 10, no. 3 (September 1, 2007): 248–258.
27. Sackett, D. L., W. M. Rosenberg, J. A. Gray, R. B. Haynes, and W. S. Richardson. "Evidence Based Medicine: What It Is and What It Isn't." *BMJ : British Medical Journal* 312, no. 7023 (January 13, 1996): 71–72.
28. Brody, Howard. "Medicine's Ethical Responsibility for Health Care Reform — The Top Five List." *New England Journal of Medicine* 362, no. 4 (January 28, 2010): 283–285.
29. Fisher, Elliott S., Julie P. Bynum, and Jonathan S. Skinner. "Slowing the Growth of Health Care Costs — Lessons from Regional Variation." *The New England Journal of Medicine* 360, no. 9 (February 26, 2009): 849–852.
30. Sirovich, Brenda, Patricia M. Gallagher, David E. Wennberg, and Elliott S. Fisher. "Discretionary Decision Making By Primary Care Physicians And The Cost Of U.S. Health Care." *Health Affairs (Project Hope)* 27, no. 3 (2008): 813–823.
31. Snyder, Lois. "American College of Physicians Ethics Manual." *Annals of Internal Medicine* 156, no. 1 Part 2 (January 3, 2012): 73–104.
32. Schapira, Marilyn M., Kathlyn E. Fletcher, Mary Ann Gilligan, Toni K. King, Purushottam W. Laud, B. Alexendra Matthews, Joan M. Neuner, and Elisabeth Hayes. "A Framework for Health Numeracy: How Patients Use Quantitative Skills in Health Care." *Journal of Health Communication* 13, no. 5 (July 23, 2008): 501–517.
33. Gigerenzer, Gerd, and J.A. Muir Gray, eds. *Better Doctors, Better Patients, Better Decisions: Envisioning Health Care 2020*. 1st ed. The MIT Press, 2011.
34. Reyna, V. F. "A Theory of Medical Decision Making and Health: Fuzzy Trace Theory." *Medical Decision Making* 28, no. 6 (November 17, 2008): 850–865.
35. Kahneman, Daniel. *Thinking, Fast and Slow*. 1st ed. Farrar, Straus and Giroux, 2011.