

## **Medicine's dangerous optimism – Lessons from Dr. Pangloss**

Benjamin L. Mazer, MD, MBA

Department of Pathology, Yale University School of Medicine, New Haven, CT

Email: [bmazer@gmail.com](mailto:bmazer@gmail.com)

This is an author's accepted manuscript distributed non-commercially on the author's personal website. This version is after peer review but before copyediting and typesetting. There may be changes in the final version.

### Version of Record:

Mazer BL. Medicine's dangerous optimism – Lessons from Dr. Pangloss. *Journal of the Royal Society of Medicine*. 2020. Vol. 113(7): 262-265.

DOI: <https://doi.org/10.1177/0141076820907061>

Copyright © 2020 The Royal Society of Medicine.

Per SAGE reuse guidelines: <https://us.sagepub.com/en-us/nam/journal-author-archiving-policies-and-re-use>.

There is an ongoing struggle in medicine between skepticism and optimism. This debate is moral and philosophical, but, more importantly, it influences the way we gather and interpret evidence. Consider the story of Dr. Pangloss, the fictional “professor of metaphysico-theologico-cosmolo-nigology” in Voltaire’s satirical eighteenth-century novel *Candide*. Dr. Pangloss is remembered for declaring that we live in the “best of all possible worlds.” Pangloss could find logical explanations for the pain and turmoil he saw around him. No one suffered without a good reason. In the face of healthcare’s overwhelming complexity, doctors can also inadvertently resort to assuming our current situation is the best we can hope for.

I am defining Panglossian optimism as the unproven assumption that an observed outcome is the necessary outcome. “All this cannot be otherwise,” as Pangloss puts it. Stephen Jay Gould and Richard Lewontin famously critiqued the evolutionary biology of their day as Panglossian for assuming that all observed animal traits were evolutionary adaptations, rather than some being incidental occurrences.<sup>1</sup> I see this Panglossian error being made in medicine, as well. In presenting the following examples of Panglossian fallacies, I hope to encourage reflection about how an instinct toward rationalizing our traditions and behaviors can limit medicine’s progress.

**Panglossian fallacy #1: Favorable outcomes are attributable to medical care, unfavorable outcomes to a lack of it**

New therapies are continually introduced into medical practice. It is easy to assume that improvements in patient outcomes temporally associated with these practices are caused

by them. Colorectal cancer mortality in the United States, for example, has fallen by more than 50% in recent decades, coincident with the introduction of colon cancer screening.<sup>2</sup> While screening has likely played a significant role in declining mortality, it cannot entirely justify it. The decline in mortality started before screening reached a high penetrance, and before the expected delay between polypectomy and cancer prevention. Yet “these trends are often attributed to screening,” according to H. Gilbert Welch and Douglas Robertson.<sup>2</sup> They suggest over-attribution occurs because “it’s tempting to take credit for good news.” There may also be other beguiling instincts at work: the age-old challenge of differentiating correlation from causation, as well as the difficulty physicians can have reconciling complex epidemiologic trends with their individual clinical experiences.

Inversely, lack of care is now implied to be the cause of bad outcomes. Studies quantifying the extent that inadequate care contributes to mortality have suffered from this fallacy. One popular method uses administrative data to rank hospitals based on mortality and other quality indicators.<sup>3</sup> Inferred from these rankings is that if hospitals with above average ratings are doing something right, then hospitals with below average ratings must be causing deaths due to inadequate practices. But mortality rates and other statistics can be influenced by factors beyond quality of care, such as case mix or socioeconomic differences. These methods have led to preposterous estimates of deaths caused by preventable adverse events, some higher than 400,000 annually in the United States.<sup>4</sup> The estimates’ flawed optimism is that “once the doctor intervenes, death is optional,” radiologist Saurabh Jha has said.<sup>5</sup> Hospitals are the best of all possible worlds, where every poor outcome is preventable with the right care.

## **Panglossian fallacy #2: Arduous training and examination are what produce good doctors**

Since nearly every doctor practicing in the United States today completed a set of standardized licensing exams and an accredited residency program, it is often believed that licensing exams and residency are what it must take to become a doctor. This belief is widespread, even though most educational traditions are not the product of rigorous study.

The USMLE Step 1 exam score is integral to selecting trainees for residency programs. Many doctors feel performance on this basic science exam isn't predictive of clinical ability, and we shouldn't rely on it.<sup>6</sup> But the exam has its ardent defenders. Regardless of whether the test predicts good doctoring, its use in easily winnowing applicants is considered valuable. Officials at the National Board of Medical Examiners and Federation of State Medical Boards recently defended the use of Step 1 scores by stating that "a program director who receives hundreds or thousands of applications every year makes comprehensive review of every application nearly impossible."<sup>7</sup> If evaluating applicants is currently too challenging without a Step 1 score, then a Step 1 score must be necessary to evaluate applicants. This is not a sound assumption. A tool that does not select for the qualities we desire inserts bias and noise into the process, making it less efficient.

We also encounter faulty Panglossian reasoning in debates over whether residency duty hours should be restricted for patient and trainee wellbeing. Many experienced physicians imagine their skills molded in the cauldron of inhumane work hours. It is true that they worked inhumane hours, and that many possess excellent skills. The Panglossian assumption is that the latter derives from the former. This type of cognitive leap is intuitive,

though fallible. The problem with Panglossian medicine is it doesn't stop at our personal thinking; it extends into the design of our science.

To address the duty-hour question, two trials have recently been performed on surgery and internal medicine residents.<sup>8,9</sup> The trials were designed as “noninferiority” studies, a low bar in which longer shifts need only show not substantially worse patient mortality than shorter shifts. “Shorter” shifts (up to 16-28 hours) are still longer than in the majority of professional jobs, yet they were disfavored from the start by the noninferiority design. Since even longer shifts are so unusual, they should have to prove superiority – a decrease in mortality. These studies show that Panglossian optimism can be nostalgic, seeking to justify fading traditions as necessary, even lowering the bar of evidence to do so.

### **Panglossian fallacy #3: Physician outcomes predict patient outcomes**

Some interventions are designed not to modify meaningful patient outcomes but to modify the behavior of physicians. In these situations, the typical choices physicians make are implied to be correct, or at least inevitable. In its extreme form, this study design produces drugs and surgeries that treat doctors instead of patients.

Consider the approval of thrombopoietin receptor agonist Avatrombopag for perioperative use in thrombocytopenia of chronic liver disease.<sup>10</sup> The primary outcome of its pivotal studies was a composite of the number of platelet transfusions given and rescue procedures for bleeding. The Panglossian belief in this trial is that any platelets transfused were needed in the first place, so a decrease in transfusions represents a good clinical outcome. In a prophylactic setting, this is an especially tenuous belief. The study showed

that Avatrombopag increases platelet count, and physicians instinctively transfuse platelets for low counts even in the absence of bleeding. Increased platelet counts will always mean fewer transfusions given our threshold-driven standard of care. Yet guidelines acknowledge that transfusion thresholds in chronic liver disease patients are based on “survey data or low-quality evidence.”<sup>11</sup> Instead of using a patient-centered outcome like bleeding or death, this study design rewards the following sequence of events: A patient is not bleeding; we provide an intervention; and the patient continues not to bleed. It’s easy for doctors to conclude that we live in the best of all possible worlds.

The assumption that physician behaviors are predictive of good outcomes can also be seen in surgical studies. A clinical trial tested a method of removing additional “cavity shave” margins during partial mastectomy to supplement traditional surgical margins.<sup>12</sup> The primary outcome was not cancer recurrence or mortality, but the presence of positive margins and the need for a second surgery to address them. These are fundamentally physician outcomes. A pathologist can declare any tissue free from tumor, though in truth this is always an incomplete sampling. The pathologist’s declaration is not a statement of fact but a test with imperfect sensitivity and specificity. Regardless, surgeons re-operate for positive margins. Any intervention that results in a “negative margin” will result in fewer re-operations. This is tautological medicine. Whether the pathologist’s diagnosis is positive or negative and whether or not the surgeon chooses to operate are just assumed to be a necessary sequence of events. This trial was considered successful, but it risks falsely reassuring doctors and patients that a “negative margin” in this setting predicts fewer recurrences or improved mortality.

#### **Panglossian fallacy #4: A sufficiently popular intervention cannot be tested**

Panglossian optimism can sometimes prevent us from gathering evidence in the first place. Even when attempts are made to rigorously study the standard of care, they face roadblocks. Among the most well-known examples was the popularity of autologous bone marrow transplant for breast cancer in the 1990's.<sup>13</sup> Patient activism, legal rulings, and regulatory decision-making led to the widespread adoption of this treatment despite a lack of definitive evidence. Subsequently, clinical trials attempting to settle lingering questions about the treatment's efficacy had difficulty recruiting patients. If patients and providers were already convinced that bone marrow transplant was the best course of action, and insurers would cover its use, there was no incentive to risk an alternative approach. Clinical trials ultimately showed no benefit for this treatment, resulting in its abandonment.

A similar story has played out with the “paradigm change” in our understanding of chronic cardiac ischemia.<sup>14</sup> Modern cardiac care has been dominated by the belief that stable coronary artery stenosis requires invasive intervention. Hundreds of thousands of percutaneous coronary interventions (PCI) are performed annually in the United States, but definitive trials testing its effectiveness compared with medical therapy have been completed only sporadically and faced significant challenges. Recent trials like ORBITA<sup>15</sup> and ISCHEMIA<sup>16</sup> continue to question the practice. Studying PCI has been slow and expensive because of how entrenched the practice is in the professional and economic lives of cardiologists. Evidence interrogating this practice, rather than being welcomed with open arms, instead produces “shock” and an “instinct to invalidate” it, as the ORBITA trialists observed.<sup>17</sup>

## **How to be Anti-Panglossian**

Voltaire meant for Pangloss to show us the complacency that comes from assuming “all is for the best.” Educational traditions and widely-practiced interventions are intimidating to change. Instead, we find ways to justify not changing them. Voltaire knew that Panglossian optimism ultimately leads to impotence. “It is impossible that things should be other than they are,” Pangloss declares. But things could always be otherwise. When designing clinical studies, we should be explicit about how we justify assumptions, especially of causality. Too often, methods are presented as a natural consequence of the question at the hand, rather than as an accumulation of suppositions. When we have a strong instinct to justify a commercial intervention or academic passion, design assumptions can feel especially self-explanatory. To be fair, it is often challenging to guess which outcomes would benefit patients most or which side effects would be least desirable. We would probably do better by asking patients about this more often.

Panglossian optimism is insidious because it assumes not only that outcomes are for the best, but that they are under our control. We discount competing trends, unmeasured confounders, and the play of chance. Beyond recognizing these influences, we can use them to our advantage. Medicine could benefit at times from the considered embrace of irrationality. How to practically winnow down thousands of residency applications, for instance, is a real concern. But if standardized exam scoring is not well-correlated to performance, and no easy alternative is available, a lottery would actually be more rational.



A lottery would save applicants costs and anxiety associated with trying to achieve superlative scores, and it wouldn't convey a false sense of fairness.

Noninferiority studies are especially vulnerable to a Panglossian sleight of hand, because with both the selection of comparators and in the width of the noninferiority margin, we will inevitably favor one arm over another. It is easy to give a new intervention too much benefit of the doubt. Experts have suggested that before relying on a non-inferiority design, we should ask ourselves whether the comparator is truly “cheaper, more convenient, less invasive, or less toxic” than the standard of care.<sup>18</sup> When it comes to longer medical training shifts, the “flexibility” promoted by studies seems to primarily benefit the administrators doing the scheduling.

The use of surrogate endpoints is not inherently incorrect. Jennifer Gill and Vinay Prasad recently proposed a thought experiment to demonstrate how to identify a Panglossian surrogate measure, although they did not call it this.<sup>19</sup> Consider a situation where an error in the electronic medical record, rather than an intervention, results in a surrogate value like a patient's platelet count appearing to change. This might affect physician behavior in the same manner as a drug that truly changes the value. Since an electronic glitch cannot improve patient outcome, perhaps an actual change in platelet count from a drug is similar. By identifying these feedback loops within our studies, we can avoid relying on tautological outcomes for evidence.

The best teachers and scientists are Anti-Panglossians. They take nothing for granted, questioning even entrenched medical assumptions. We celebrate Dr. Bernard Fisher for pioneering the partial mastectomy against so much opposition.<sup>20</sup> Anti-Panglossians are not always celebrated; we often reward those who justify comforting beliefs. When medical

authorities promote the status quo, it is hard for the average conscientious but overworked physician to know how to act. Following the standard of care implies endorsement, and fear of assuming legal risk by deviating from the standard only further ossifies medical practices. But I feel that doctors have an instinct toward action. We entered the profession to heal. We rationalize the irrational in times of perceived helplessness. By creating an environment more hospitable to questioning and change, we may be less drawn to false comforts.

## **Declarations**

Competing interests: BLM has received payments from Medscape for writing articles. He has received an honorarium from Hillcrest Healthcare System for a continuing medical education talk.

Funding: None.

Ethical approval: Not required.

Guarantor: BLM.

Contributorship: Sole author.

Acknowledgements: The author thanks Joseph Ross, MD and H. Gilbert Welch, MD for their feedback.

## References

1. Gould SJ, Lewontin RC. The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proc R Soc Lond B* 1979; 205: 581–598.
2. Welch HG, Robertson DJ. Colorectal Cancer on the Decline — Why Screening Can't Explain It All. *N Engl J Med* 2016; 374: 1605–1607.
3. Hogan H. The problem with preventable deaths. *BMJ Qual Saf* 2016; 25: 320–323.
4. Mazer BL, Nabhan C. Strengthening the Medical Error “Meme Pool”. *J Gen Intern Med* 2019; 34: 2264–2267.
5. Jha S. Actually, Medical Errors are the Leading Cause of Death. *The Health Care Blog*, <https://thehealthcareblog.com/blog/2016/05/09/actually-medical-errors-are-the-leading-cause-of-death/> (9 May 2016, accessed 20 December 2019).
6. Carmody JB, Sarkany D, Heitkamp DE. The USMLE Step 1 Pass/Fail Reporting Proposal: Another View. *Acad Radiol* 2019; 26: 1403–1406.
7. Katsufakis PJ, Chaudhry HJ. Improving Residency Selection Requires Close Study and Better Understanding of Stakeholder Needs: *Acad Med* 2019; 94: 305–308.
8. Bilimoria KY, Chung JW, Hedges LV, et al. National Cluster-Randomized Trial of Duty-Hour Flexibility in Surgical Training. *N Engl J Med* 2016; 374: 713–727.
9. Silber JH, Bellini LM, Shea JA, et al. Patient Safety Outcomes under Flexible and Standard Resident Duty-Hour Rules. *N Engl J Med* 2019; 380: 905–914.
10. Terrault N, Chen Y-C, Izumi N, et al. Avatrombopag Before Procedures Reduces Need for Platelet Transfusion in Patients With Chronic Liver Disease and Thrombocytopenia. *Gastroenterology* 2018; 155: 705–718.
11. Patel IJ, Rahim S, Davidson JC, et al. Society of Interventional Radiology Consensus Guidelines for the Periprocedural Management of Thrombotic and Bleeding Risk in Patients Undergoing Percutaneous Image-Guided Interventions—Part II: Recommendations. *J Vasc Interv Radiol* 2019; 30: 1168-1184.e1.
12. Chagpar AB, Killelea BK, Tsangaris TN, et al. A Randomized, Controlled Trial of Cavity Shave Margins in Breast Cancer. *N Engl J Med* 2015; 373: 503–510.
13. Mello MM, Brennan TA. The Controversy Over High-Dose Chemotherapy With Autologous Bone Marrow Transplant For Breast Cancer. *Health Aff* 2001; 20: 101–117.
14. Mitchell JD, Brown DL. Harmonizing the Paradigm With the Data in Stable Coronary Artery Disease: A Review and Viewpoint. *J Am Heart Assoc* 2017; 6: e007006.

15. Al-Lamee R, Thompson D, Dehbi H-M, et al. Percutaneous coronary intervention in stable angina (ORBITA): a double-blind, randomised controlled trial. *The Lancet* 2018; 391: 31–40.
16. Johnson CY. Stents and bypass surgery are no more effective than drugs for stable heart disease, highly anticipated trial results show. *Washington Post*. <https://www.washingtonpost.com/health/2019/11/16/embargoed-drugs-are-effective-invasive-procedures-patients-with-stable-heart-disease-major-trial-finds/> (16 November 2019, accessed 20 December 2019).
17. Francis DP, Al-Lamee R. Percutaneous coronary intervention for stable angina in ORBITA – Authors’ reply. *The Lancet* 2018; 392: 28–30.
18. Prasad V. Non-Inferiority Trials in Medicine: Practice Changing or a Self-Fulfilling Prophecy? *J Gen Intern Med* 2018; 33: 3–5.
19. Gill J, Prasad V. A method to determine if more than surrogate outcomes were improved: The EMR glitch experiment. *Res Pract Thromb Haemost* 2020; 4: 19–22.
20. Gellene D. Dr. Bernard Fisher, Who Revolutionized Breast Cancer Treatment, Dies at 101. *The New York Times*. <https://www.nytimes.com/2019/10/19/science/dr-bernard-fisher-dead.html> (19 October 2019, accessed 20 December 2019).