

Seduction and Insight from Cross-National Comparisons

In this issue of the *Journal* (pp. 875–880), Wunsch and colleagues provide a thought-provoking paired comparison of intensive care services during terminal hospitalizations in the United States and England (1). They demonstrate that the English are more likely to die in a hospital than are Americans, but that far fewer terminal hospitalizations in England involve the intensive care unit (ICU). Older Americans and those with nonsurgical conditions are particularly more likely to be provided with intensive care during their terminal hospitalization than are their English counterparts.

Cross-national comparisons like that of Wunsch and colleagues are among the most seductive of study designs. They seem to offer a chance to bring rigorous data to bear on the big issues of social justice, institutional priorities, and how we ought to organize society. We want such studies to tell us why our health care systems are in the mess that they are in, and how to fix them. Yet although cross-national comparisons can offer important insights, even the most rigorous one-off comparisons generally cannot deliver the answers to our pressing questions about health care systems.

Despite analyzing over 14 million terminal hospitalizations, Wunsch and colleagues cannot truly answer (and to their credit, do not claim to answer) the “why” question. Their comparison is cross-sectional and between two nations. This methodology limits inference about the causes of the observed cross-national differences. It may be that their finding of stark differences in use of intensive care reflects differences between American and English norms surrounding patient decision making, autonomy, and entitlement. Or it may be that, unlike in the United States, centralized budgeting and allocation of high-dependency units in the United Kingdom precludes filling beds simply because they are there or turning to ICUs to monitor even low-risk conditions. With an n of 2, Wunsch and colleagues do not hope to arbitrate between these—and other—possibilities.

Do these concerns mean that we should not attempt cross-national comparisons of critical care? We hesitate to draw such a conclusion. Despite the apparent limitations imposed by an n of 2, firm answers to the “why” questions *are* possible using comparative case studies. Indeed, important questions have been addressed in such a context: Do democracies do better than authoritarian regimes at protecting their citizens from death due to malnutrition or famine (2)? What role do employers play in determining the shape of welfare states (3)? Generating valid causal inferences from paired case studies requires engaging with social science methodologies that may be unfamiliar to many readers of this journal. Carefully researched, historically grounded case studies can be leveraged to produce compelling results by testing each step in a proposed causal pathway (4–6). An exploration of the evolution of critical care, taking into account diverse national histories and healthcare ecosystems, would help contextualize what we are learning about contemporary divergences in practice (7, 8). Although we currently know very little about the implications of differences in critical care organization and delivery for outcomes at the patient level, these questions are simply unanswered, not unanswerable.

Even if we understood why our national systems function as they do, and the consequences for patient outcomes, such knowledge might not add up to a recipe for national-level interventions to improve care. A whole range of interlinked political and policy institutions, not just the health care system, but also the tax system, the electoral system, the territorial organization of the state, and the organization of the medical profession, affect the kinds of health policies that can actually be enacted, even during periods of seeming policy openness. Health policy in the United States and the United Kingdom is powerfully limited by path dependence and other forms of institutional inertia (9, 10). Hence, the policy lessons that we might draw from a cross-national comparison, such as that by Wunsch and colleagues, must take into account both the potentially limited institutional room for maneuver and the opportunities for real change that do exist.

Unfortunately, policy lessons from cross-national studies can be powerful, but they will not tell us what is “right.” Data alone cannot provide the normative answer about overuse and underuse. Our beliefs about the appropriateness of current levels of health care expenditures, and the relative value and opportunity cost of those expenditures, are derived from normative judgments, not empirical science. Comparisons illustrate the divergent consequences of past policy choices, but deciding what to do next depends on our values and ethics.

Dr. Wunsch and colleagues are conscious of such limitations, as well as of pedestrian challenges, such as data equivalence and decedent bias. Navigating these hazards is worthwhile, as it opens the possibility of interesting and important new questions. If we believe that English and American hospitals provide more or less equivalent standards of care, then Wunsch and colleagues’ results suggest that there is substantial room for organizational difference without sacrificing quality. Cross-national comparisons can help us think broadly about alternative approaches to care for the critically ill, as we learn more about how conscientious physicians practice quite differently in different settings.

In the context of growing health budget crises, organizational innovation might focus on maintaining clinical excellence while reducing the costs associated with the care of patients. Although national health policy contexts may be different, both English and American patients are cared for in hospitals that have found ways of making do with limited resources. This variation can be a rich source of hypotheses about alternative ways to organize hospital care for the critically ill. Are there conditions with equivalent mortality regardless of the ICU? What are the costs and benefits of providing ICU-like services on hospital floors? We cannot assume that the results of comparisons of policy innovations within one country will generalize to another context with a different constellation of policies. But, done properly, such comparisons could provide a reasonable foundation for carefully conducted interventional studies, particularly at the hospital level.

Faced with a provocative finding of cross-national difference, the scientific community faces a choice between at least two paths. One path leads to carefully unpacking the origins of this difference and teaching us something generally true about how critical care systems develop. The other path leads into the hospitals, using observational data to imagine new ways to organize care and generate the equipoise necessary for careful interventional studies of such innovations. The first path helps us

shape national policy levers. The latter path helps us redesign care organizations to bring change to patients. Both are necessary.

Conflict of Interest Statement: Neither author has a financial relationship with a commercial entity that has an interest in the subject of this manuscript.

THEODORE J. IWASHYNA, M.D., PH.D.
University of Michigan Medical School
Ann Arbor, Michigan

JULIA LYNCH, PH.D.
University of Pennsylvania
Philadelphia, Pennsylvania

References

1. Wunsch H, Linde-Zwirble WT, Harrison DA, Barnato AE, Rowan KM, Angus DC. Use of intensive care services during terminal hospitalizations in England and the United States. *Am J Respir Crit Care Med* 2009;180:875–880.
2. Drèze J, Sen A. Hunger and public action. Oxford: Clarendon Press; 1989.
3. Mares I. The politics of social risk. New York: Cambridge University Press; 2003.

4. Campbell D. Degrees of freedom and the case study. *Comp Polit Stud* 1975;8:178–193.
5. King G, Keohane R, Verba S. Designing social inquiry: scientific inference in qualitative research. Princeton, NJ: Princeton University Press; 1994.
6. Collier D, Brady H, Seawright J. Sources of leverage in causal inference: toward an alternative view of methodology. In: Brady H, Collier D, eds. Rethinking social inquiry: diverse tools, shared standards. Lanham, MD: Rowman & Littlefield; 2004.
7. Wunsch H, Rowan K, Angus DC. International comparisons in critical care: a necessity and challenge. *Curr Opin Crit Care* 2007;13:725–731.
8. Wunsch H, Angus DC, Harrison DA, Collange O, Fowler R, Hoste EAJ, de Keizer NF, Kersten A, Linde-Zwirble WT, Sandiumenge A, et al. Variation in critical care services across North America and Western Europe. *Crit Care Med* 2008;36:2787–2793.
9. Pierson P. The new politics of the welfare state. *World Polit* 1996;48:143–179.
10. Hacker J. The divided welfare state: the battle over public and private social benefits in the United States. New York: Cambridge University Press; 2002.

DOI: 10.1164/rccm.200907-1013ED

Diagnosing Tuberculosis in Patients with HIV Do We Know Enough?

The convergence of the HIV and tuberculosis (TB) epidemics has complicated control of both conditions (1). With shared adverse social factors and greatly enhanced risk for TB disease among patients with coinfection, there have been stunning rises in TB prevalence, drug resistance, and mortality in many of the worst-affected areas. The high fatality of the recently reported extensively drug-resistant TB (XDR-TB) series in Africa bears witness to this ongoing crisis (2).

In many developing areas, patients with TB are being diagnosed largely by sputum microscopy. Unfortunately, HIV-related TB is often sputum smear-negative. In this issue of the *Journal* (pp. 903–908), Monkongdee and colleagues evaluated the performance of acid-fast bacilli (AFB) smears and mycobacterial cultures in diagnosing TB among persons infected with HIV in Thailand and Vietnam, using culture-confirmed TB as the gold standard (3). Two culture methods, solid medium (Lowenstein Jensen [LJ]) and liquid medium (Mycobacteria Growth Indicator Tube [MGIT] and MycoF-Lytic bottle, using BACTEC systems) (BD, Franklin Lakes, NJ), were included. AFB smears of the first collected sputum diagnosed only 29% of culture-positive TB cases. Adding a second and third sputum specimen for AFB smear diagnosed only an additional 7 and 2% of cases, respectively. The authors infer that the minimal incremental yield of the third smear in this study makes it not worth doing and supports WHO's recommendation to use two, rather than three, sputum smears in a person infected with HIV who is suspected of having TB (4).

Although these findings are useful in optimizing resource use, it is by no means the whole story. In fact, the more critical finding of this study is that, all told, reliance on either two or even three sputum AFB smears failed to diagnose approximately two-thirds of patients with pulmonary TB who were infected with HIV and who would have been diagnosed if cultures had been done. Overlooking this killer disease in people living with HIV could have grave implications. Unlike HIV, TB spreads by air. Smear-negative but culture-positive cases remain an important infectious source (5, 6). With a delay in diagnosis, both drug-susceptible and drug-resistant bacillary strains can spread easily among subjects

infected with HIV and into the general community. Emphasis has therefore been rightly put on the diagnosis of smear-negative and extrapulmonary TB cases among persons infected with HIV in the revised Stop TB Strategy (7) with the incorporation of clinical evaluation, chest radiography, empirical antibiotic trial, and sputum culture in the diagnostic workup. The current study by Monkongdee and colleagues thus provides much-needed evidence in support of such recommendations.

As previously demonstrated, MGIT cultures had a higher diagnostic yield than LJ solid medium cultures (8). In the study by Monkongdee and colleagues, one, two, and three LJ cultures respectively identified 48, 62, and 68% of pulmonary cases in contrast to 71, 88, and 98% for MGIT cultures. Therefore, neither three sputum smears, nor three LJ cultures guaranteed a 70% case detection rate of pulmonary TB. Furthermore, enlarged peripheral lymph nodes were found in 12% of enrolled subjects, and 42% of lymph node aspirates were culture positive. To diagnose smear-negative and extrapulmonary tuberculosis, the authors therefore recommend adding broth-based cultures and lymph node aspirate (if applicable) on top of two sputum smears.

The study by Monkongdee and colleagues does not address the issue of drug resistance. Globally, rates of drug resistance have risen critically, and in some areas of the world, rates of multidrug-resistant TB (MDR-TB) and XDR-TB are shockingly high (9). Although two (or three) sputum smears are useful in diagnosing approximately one-third of cases of TB, in areas where there is significant prevalence of drug resistance, starting patients who were smear-diagnosed on standard TB regimens without obtaining cultures runs a clear risk of treatment failure, not only undermining the ability to achieve the World Health Organization's target of 85% treatment success, but also risking the amplification of further drug resistance. Indeed, this has contributed to the rise of higher-order resistance (MDR and XDR) in some geographic settings (2).

Regular use of chest radiographs and empirical trial of antibiotics alone, although significantly increasing the numbers of smear-negative patients with HIV treated for TB, would not