

Public Reporting of Antibiotic Timing in Patients with Pneumonia: Lessons from a Flawed Performance Measure

Robert M. Wachter, MD; Scott A. Flanders, MD; Christopher Fee, MD; and Peter J. Pronovost, MD, PhD

The administration of antibiotics within 4 hours to patients with community-acquired pneumonia has been criticized as a quality standard because it pressures clinicians to rapidly administer antibiotics despite diagnostic uncertainty at the time of patients' initial presentations. The measure was recently revised (to 6 hours) in response to this criticism. On the basis of the experience with the 4-hour rule, the authors make 5 recommendations for the development of future publicly reported quality measures. First, results from samples with known diagnoses should be extrapolated cautiously, if at all, to patients without a diagnosis. Second, for some measures, "bands" of performance may make more sense than

"all-or-nothing" expectations. Third, representative end users of quality measures should participate in measure development. Fourth, quality measurement and reporting programs should build in mechanisms to reassess measures over time. Finally, biases, both financial and intellectual, that may influence quality measure development should be minimized. These steps will increase the probability that future quality measures will improve care without creating negative unintended consequences.

Ann Intern Med. 2008;149:29-32.

www.annals.org

For author affiliations, see end of text.

Improving health care quality depends on having valid ways to measure quality. Unfortunately, there are few validated quality outcome measurements, because valid and feasible case-mix adjusters are lacking and patients are difficult to follow over time for clinically important outcomes, such as death. Processes of care are easier to identify and measure, but some of these measures will be proven invalid or inappropriate because their scientific rationale was flawed from the start, unanticipated consequences emerge after implementation, or later studies undermine them.

We review how these issues played out in the measure of time to first antibiotic dose (TFAD), also called "door-to-needle time," for patients presenting to the hospital with community-acquired pneumonia (CAP). We also propose lessons that can be learned from the experience.

TFAD AS A QUALITY MEASURE

Community-acquired pneumonia is one of the most common admitting diagnoses in U.S. hospitals, accounting for more than 1 million hospitalizations yearly (1), with short-term mortality rates ranging from 0.5% to 27.1% (2). Given its risk, frequency, and perceived outcome variations, CAP was an obvious candidate for quality measurement and improvement initiatives. Because outcome measurement in CAP was problematic for the usual reasons (data collection burden, case-mix adjustment, and need for posthospital follow-up), investigators sought process measures associated with higher quality.

During the 1990s, the notion of time-based quality measures gained favor because evidence emerged that rapid treatment of myocardial infarction, and later trauma, stroke, and sepsis, improved outcomes (3–7). Naturally, investigators began to examine whether rapid administration of antibiotics might improve CAP outcomes.

In 1997, a retrospective study of 14 069 Medicare patients hospitalized for CAP found that, after adjustment for severity (2) and demographic factors, administration of an-

tibiotics within 8 hours was associated with a lower 30-day mortality rate (odds ratio [OR], 0.85 [95% CI, 0.75 to 0.96]) (8). Patients were included if they had chest radiography results within 2 days of admission consistent with pneumonia and an initial "working diagnosis" of pneumonia.

In 2004, a second retrospective study of 13 771 Medicare patients (age ≥ 65 years) hospitalized for CAP (9) also found that, among the 75% of patients without evidence of prehospital receipt of antibiotics, administration of antibiotics within 4 hours was associated with a lower 30-day mortality rate (OR, 0.85 [CI, 0.76 to 0.95]). Extrapolating these data to a hypothetical national Medicare sample, the authors estimated that achieving TFAD by 4 hours after presentation to the hospital would save more than 1200 lives yearly.

The 2 studies reported that patients who received their first dose of antibiotics in the first hour of their emergency department stay had a higher mortality rate than those who received antibiotics later; however, this finding was attributed to incomplete adjustment for severity of CAP and was therefore not felt to challenge the main conclusion about TFAD (8, 9). Two smaller studies of CAP found no association between early antibiotic administration and outcomes (10, 11). Nevertheless, the authors of the 2004 study (9) editorialized that the 4-hour TFAD quality measure was still valid (12, 13).

TRANSLATION INTO A PERFORMANCE STANDARD

Almost exclusively on the basis of results from the 1997 study, the Medicare National Pneumonia Project endorsed first antibiotics within 8 hours of hospital arrival as a CAP quality measure in 1998. The Medicare National Pneumonia Project tightened its TFAD window to 4 hours in 2002 on the basis of the prepublication results of the 2004 study by Houck and colleagues (9, 12). In 2003, the Infectious Diseases Society of America (IDSA) also endorsed a 4-hour timeframe (14). With support from the

Medicare National Pneumonia Project and IDSA and subsequent endorsement by the National Quality Forum, The Joint Commission and The Centers for Medicare & Medicaid Services (CMS) chose the 4-hour TFAD measure as 1 of their initial core measures of quality (measure PN-5b). Since 2002, this measure has been publicly reported for all U.S. hospitals. In 2006, it became part of a measure set tied to additional payments under several pilot pay-for-performance programs (15).

THE RESPONSE FROM EMERGENCY MEDICINE

The emergency medicine community began raising red flags about the TFAD measure soon after its formulation, and complaints from this community markedly increased after TFAD was publicly reported and became the subject of pay-for-performance programs (16). Published studies challenging the measure soon followed. Although some questioned the association itself, most focused on the issue of diagnostic uncertainty. One study found that 22% of 86 randomly selected patients with pneumonia had uncertain presentations and often lacked infiltrates on chest radiography, which could have appropriately led to delayed antibiotic administration (17). Another study documented cases that were labeled “poor-quality care,” in which delayed use of antibiotics was clinically appropriate (18), whereas still another found that maneuvers to improve TFAD were not very cost-effective (19). In fact, many eligible patients with a working diagnosis of CAP who did not receive antibiotics within 4 hours had no radiographic evidence of pneumonia in the emergency department and did not have a final emergency department diagnosis of CAP (20, 21). Moreover, other studies showed that TFAD measurement led to administration of antibiotics in many patients who proved not to have pneumonia or another infectious disease (22, 23). Finally, a recent systematic review concluded that “evidence from observational studies fails to confirm decreased mortality with early administration of antibiotics in stable patients with [CAP]” (24). On the basis of these studies, analyses, and considerable anecdotal evidence, editorials in the emergency medicine literature argued vigorously for relaxing the TFAD standards (25, 26), pointing out that the measure was skewing emergency department triage priorities and promoting unnecessary antibiotic use (18).

THE RESPONSE FROM PAYERS, REGULATORS, AND PROFESSIONAL SOCIETIES

Within months of the critical publications, The Joint Commission and CMS revisited measure PN-5b. In October 2006, patients eligible for the measure had to have a final emergency department diagnosis of pneumonia (rather than an initial working diagnosis) and objective radiographic findings sometime during the hospitalization. Unfortunately, although the revised criteria solved some of the problems associated with PN-5b, they created new ones. For example, Fee and colleagues (27) worried that

the new measures would generate pressure to administer antibiotics before patients were sent for computed tomography to rule out pulmonary embolism (even in the face of nondiagnostic chest radiographs) or to avoid writing “pneumonia” as the final emergency department diagnosis.

In March 2007, IDSA and the American Thoracic Society issued joint guidelines that abolished time-specific goals for CAP treatment, now recommending that patients receive their first dose of antibiotics as soon as possible after a definitive diagnosis of CAP, preferably in the emergency department (28). One month later, The Joint Commission created a test measure (PN-5c) that relaxed the antibiotic administration window to 6 hours. That same month, the National Quality Forum withdrew its endorsement of measure PN-5b and endorsed PN-5c, which became the publicly reported measure in April 2008. In addition, The Joint Commission created a new data element, “diagnostic uncertainty,” which may exclude patients from TFAD measurement (29, 30). Whether all of these revisions will solve the problems associated with measure PN-5b is unknown; no study has yet shown a benefit from a 6-hour rule, and the diagnostic uncertainty construct has not, to our knowledge, been field-tested and validated.

UNANTICIPATED CONSEQUENCES OF TFAD MEASUREMENT AND REPORTING

Prompt administration of antibiotics to patients with documented pneumonia makes sense, and seeking ironclad evidence to prove its value might seem to be analogous to requiring proof of the value of parachutes (31). Moreover, a randomized trial that withheld early antibiotic treatment in some patients with CAP would be unethical. It was therefore inevitable that decisions about the timing of antibiotic administration in CAP would be based on imperfect retrospective studies, out of necessity (8, 9).

However, the TFAD measure was enacted largely on the evidence derived from 2 large studies, in which conditions (retrospective review of patients with working diagnoses of pneumonia) replicate only in part the predicament that busy emergency medicine physicians face daily: evaluating scores of patients with cough, fever, dyspnea, weakness, dizziness, confusion, or abdominal pain. As Pines (26) has written, “Most ED [emergency department] patients do not present at triage with a sign on their forehead that reads, ‘I have pneumonia; give me antibiotics now!’” Unlike myocardial infarction, in which there is palpable clinical urgency to confirm the diagnosis and a series of tests (cardiac biomarker measurement and electrocardiography) available to reliably do so, no gold standard test for pneumonia exists. Although a triage rule of obtaining an electrocardiogram in any patient whose symptoms, signs, or risk factors make myocardial infarction even a remote possibility makes perfect sense, a similar strategy for chest radiography would be resource intensive, often confusing (given the relatively poor sensitivity and specificity of the

test in CAP [32]), impractical, and even potentially harmful (because of radiation exposure).

In the days before measurement of TFAD, patients with uncertain diagnoses would continue to be evaluated until the diagnosis was clarified. However, the TFAD standard completely transformed the dynamic: Faced with a patient who might have pneumonia, the emergency medicine physician now has a strong incentive (almost always buttressed by social pressure and sometimes by financial incentives) to give antibiotics before 4 hours have passed, even when he or she is still unsure of the diagnosis (33).

WHAT CAN WE LEARN?

For a performance measure to be publicly reported or associated with differential payments, it must be valid not only in its science but also in its measurement and application. What lessons can we glean from experience with the TFAD measure that may help in the future development of publicly reported quality measures?

First, great caution is needed in moving from retrospective studies that examine process–outcome links in patients with known diagnoses to prospective application of quality metrics in patients with uncertain diagnoses. We are not saying that such extrapolation should never be done—particularly when the outcome differences are large, therapy is inexpensive and benign, and the diagnosis is straightforward. However, CAP meets none of these criteria, and implementation of a TFAD standard would have benefited from pilot studies in busy emergency departments that assessed the validity and reliability of the measures and the benefits, harms, and costs of implementation.

Second, we should consider using “bands” of performance for certain measures (16, 34). Even if rapid antibiotic administration is clinically important, the risk that clinicians would be pressured to give unnecessary antibiotics to individual patients would be diminished if, for example, hospital performance was reported within a range of 80% to 100% or 60% to 80% adherence. This approach, which CMS has recently embraced (35), recognizes that certain quality measures may benefit the average patient, but that an all-or-none threshold (36) creates undue pressure to treat, even when clinically inappropriate.

Third, key end-users must be “at the table” during development and approval of quality metrics. In the case of CAP measures, most of the initial participants were experts in pneumonia, infectious diseases, and pulmonary disease—individuals who typically see very ill hospitalized patients with pneumonia who have obvious diagnoses. Future panels that examine such issues as these should include emergency medicine physicians and hospitalists, that is, those who make the initial diagnosis and are under pressure to administer or withhold antibiotics.

Fourth, quality measurement and reporting programs should create mechanisms to assess the validity, reliability, impact, and costs of measures within 1 to 2 years after

implementation. In addition to feedback from individual providers, policymakers and payers should seek input from professional societies and health care systems because these organizations will probably hear about problems long before they surface in the published literature. Those promoting the measures should produce periodic reports on this feedback and emerging evidence about the measures (37) and how they are responding to this information. In this case, CMS, the National Quality Forum, and The Joint Commission should be commended for rapidly responding to the chorus of concerns about the 4-hour rule, even though the responses to date may not completely solve the problems.

Finally, in addition to financial conflicts of interest, caution should be exercised when individuals are both key investigators and policymakers, particularly when the stakes are high. In the case of TFAD, several key researchers had positions with CMS and IDSA and helped drive the conversion of their own studies into performance standards (8, 9). None of us can be entirely impartial when judging the merits of our own research.

CONCLUSION

The story of early antibiotic administration for patients with CAP is not one of malfeasance or incompetence, but rather of differing perspectives and the unintended consequences of imperfect science. As we enter this new world of quality measurement, public reporting, and pay-for-performance, missteps will happen. This should not dissuade us from promoting interventions that can improve patient care nor lead us to insist on unattainable levels of evidence before proceeding.

The case of TFAD as a quality measure in pneumonia suggests the importance of taking actions, such as the 5 steps we described earlier, that increase the probability that future quality measures will improve care without creating unintended and potentially negative consequences. Similar issues, in which aggressive efforts to meet standards may create more harm than good in some patients, are likely to come into play in other clinical areas, such as with performance standards to achieve tight control of glucose, hypertension, and pain (38).

Finally, the TFAD experience demonstrates that flawed measures may periodically pit physicians’ desire to do the right thing (in this case, temporarily withhold antibiotics in patients who might or might not have an infection) against understandable concerns about the public’s view of their quality of care or even against their own financial interests. In these unusual circumstances, professionalism dictates that physicians focus on their patients’ best interest above all else.

From the University of California, San Francisco, San Francisco, California; University of Michigan School of Medicine, Ann Arbor, Michigan; and Johns Hopkins School of Medicine and School of Public Health, Baltimore, Maryland.

Potential Financial Conflicts of Interest: *Consultancies:* R.M. Wachter (American Board of Internal Medicine, Healthcare Advisory Board, Google). *Honoraria:* R.M. Wachter (Dr. Wachter has done more than 100 talks on issues of quality and patients safety to several health care organizations.). *Grants received:* R.M. Wachter (Agency for Healthcare Research and Quality).

Requests for Single Reprints: Robert M. Wachter, MD, Room M-994, Department of Medicine, 505 Parnassus Avenue, University of California, San Francisco, San Francisco, CA 94143-0120; e-mail, bobw@medicine.ucsf.edu.

Current author addresses and author contributions are available at www.annals.org.

References

1. Fine MJ, Stone RA, Singer DE, Coley CM, Marrie TJ, Lave JR, et al. Processes and outcomes of care for patients with community-acquired pneumonia: results from the Pneumonia Patient Outcomes Research Team (PORT) cohort study. *Arch Intern Med.* 1999;159:970-80. [PMID: 10326939]
2. Fine MJ, Auble TE, Yealy DM, Hanusa BH, Weissfeld LA, Singer DE, et al. A prediction rule to identify low-risk patients with community-acquired pneumonia. *N Engl J Med.* 1997;336:243-50. [PMID: 8995086]
3. Morrison LJ, Verbeek PR, McDonald AC, Sawadsky BV, Cook DJ. Mortality and prehospital thrombolysis for acute myocardial infarction: A meta-analysis. *JAMA.* 2000;283:2686-92. [PMID: 10819952]
4. Cannon CP, Gibson CM, Lambrew CT, Shoultz DA, Levy D, French WJ, et al. Relationship of symptom-onset-to-balloon time and door-to-balloon time with mortality in patients undergoing angioplasty for acute myocardial infarction. *JAMA.* 2000;283:2941-7. [PMID: 10865271]
5. Vernon DD, Furnival RA, Hansen KW, Diller EM, Bolte RG, Johnson DG, et al. Effect of a pediatric trauma response team on emergency department treatment time and mortality of pediatric trauma victims. *Pediatrics.* 1999;103:20-4. [PMID: 9917434]
6. Caplan LR. Tissue plasminogen activator for acute ischemic stroke. *N Engl J Med* 1999; 341:1240-1.
7. Rivers E, Nguyen B, Havstad S, Ressler J, Muzzin A, Knoblich B, et al. Early Goal-Directed Therapy Collaborative Group. Early goal-directed therapy in the treatment of severe sepsis and septic shock. *N Engl J Med.* 2001;345:1368-77. [PMID: 11794169]
8. Meehan TP, Fine MJ, Krumholz HM, Scinto JD, Galusha DH, Mockalis JT, et al. Quality of care, process, and outcomes in elderly patients with pneumonia. *JAMA.* 1997;278:2080-4. [PMID: 9403422]
9. Houck PM, Bratzler DW, Nsa W, Ma A, Bartlett JG. Timing of antibiotic administration and outcomes for Medicare patients hospitalized with community-acquired pneumonia. *Arch Intern Med.* 2004;164:637-44. [PMID: 15037492]
10. Dedier J, Singer DE, Chang Y, Moore M, Atlas SJ. Processes of care, illness severity, and outcomes in the management of community-acquired pneumonia at academic hospitals. *Arch Intern Med.* 2001;161:2099-104. [PMID: 11570938]
11. Silber SH, Garrett C, Singh R, Sweeney A, Rosenberg C, Parachiv D, et al. Early administration of antibiotics does not shorten time to clinical stability in patients with moderate-to-severe community-acquired pneumonia. *Chest.* 2003; 124:1798-804. [PMID: 14605051]
12. Houck PM, Bratzler DW, Nsa W, Ma A, Bartlett JG. Antibiotic administration in community-acquired pneumonia [Letter]. *Chest.* 2004;126:320-1. [PMID: 15249486]
13. Houck PM, Bratzler DW. Administration of first hospital antibiotics for community-acquired pneumonia: does timeliness affect outcomes? *Curr Opin Infect Dis.* 2005;18:151-6. [PMID: 15735420]
14. Mandell LA, Bartlett JG, Dowell SF, File TM Jr, Musher DM, Whitney C, et al. Infectious Diseases Society of America. Update of practice guidelines for the management of community-acquired pneumonia in immunocompetent adults. *Clin Infect Dis.* 2003;37:1405-33. [PMID: 14614663]
15. Lindenauer PK, Remus D, Roman S, Rothberg MB, Benjamin EM, Ma A, et al. Public reporting and pay for performance in hospital quality improvement. *N Engl J Med.* 2007;356:486-96. [PMID: 17259444]
16. Houck PM. Antibiotics and pneumonia: is timing everything or just a cause of more problems? [Editorial]. *Chest.* 2006;130:1-3. [PMID: 16840371]
17. Metersky ML, Sweeney TA, Getzow MB, Siddiqui F, Nsa W, Bratzler DW. Antibiotic timing and diagnostic uncertainty in Medicare patients with pneumonia: is it reasonable to expect all patients to receive antibiotics within 4 hours? *Chest* 2006; 130:16-21. [PMID: 16840377]
18. Pines JM. Profiles in patient safety: Antibiotic timing in pneumonia and pay-for-performance. *Acad Emerg Med.* 2006;13:787-90. [PMID: 16766737]
19. Barlow G, Nathwani D, Williams F, Ogston S, Winter J, Jones M, et al. Reducing door-to-antibiotic time in community-acquired pneumonia: Controlled before-and-after evaluation and cost-effectiveness analysis. *Thorax.* 2007; 62:67-74. [PMID: 16928714]
20. Fee C, Weber E, Sharpe BA, Nguy M, Quon T, Bookwalter T. JCAHO/CMS core measures for community-acquired pneumonia [Letter]. *Ann Emerg Med.* 2006;47:505; author reply 506. [PMID: 16631994]
21. Fee C, Weber EJ. Identification of 90% of patients ultimately diagnosed with community-acquired pneumonia within four hours of emergency department arrival may not be feasible. *Ann Emerg Med.* 2007;49:553-9. [PMID: 17210202]
22. Kanwar M, Brar N, Khatib R, Fakh MG. Misdiagnosis of community-acquired pneumonia and inappropriate utilization of antibiotics: side effects of the 4-h antibiotic administration rule. *Chest.* 2007;131:1865-9. [PMID: 17400668]
23. Welker JA, Huston M, McCue JD. Antibiotic timing and errors in diagnosing pneumonia. *Arch Intern Med.* 2008;168:351-6. [PMID: 18299488]
24. Yu KT, Wyr PC. Evidence-based emergency medicine/critically appraised topic. Evidence behind the 4-hour rule for initiation of antibiotic therapy in community-acquired pneumonia. *Ann Emerg Med.* 2008;51:651-62, 662.e1-2. [PMID: 18272253]
25. Rothman RE, Quianzon CC, Kelen GD. Narrowing in on JCAHO recommendations for community-acquired pneumonia [Editorial]. *Acad Emerg Med.* 2006;13:983-5. [PMID: 16946291]
26. Pines JM. Measuring antibiotic timing for pneumonia in the emergency department: another nail in the coffin [Editorial]. *Ann Emerg Med.* 2007;49: 561-3. [PMID: 17350732]
27. Fee C, Weber EJ, Sharpe BA, Quon T. When is a scarlet letter really a red badge of courage? The paradox of percentage of pneumonia patients receiving antibiotics within 4 hours in accordance with JCAHO and CMS core measures [Letter]. *Ann Emerg Med.* 2007;50:205-6. [PMID: 17643863]
28. Mandell LA, Wunderlink RG, Anzueto A, Bartlett JG, Campbell GD, Dean NC, et al. Infectious Diseases Society of America. Infectious Diseases Society of America/American Thoracic Society consensus guidelines on the management of community-acquired pneumonia in adults. *Clin Infect Dis.* 2007;44 Suppl 2:S27-72. [PMID: 17278083]
29. Mitka M. JCAHO tweaks emergency departments' pneumonia treatment standards. *JAMA.* 2007;297:1758-9. [PMID: 17456810]
30. Specifications Manual for National Hospital Quality Measures, version 2.4. Accessed at www.jointcommission.org/NR/rdonlyres/5E08778B-2103-4F88-A950-8EE01609940F/0/nhqm_release_notes24.pdf on 11 May 2008.
31. Smith GC, Pell JP. Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials. *BMJ.* 2003;327:1459-61. [PMID: 14684649]
32. Novack V, Avnon LS, Smolyakov A, Barnea R, Jotkowitz A, Schlaeffer F. Disagreement in the interpretation of chest radiographs among specialists and clinical outcomes of patients hospitalized with suspected pneumonia. *Eur J Intern Med.* 2006;17:43-7. [PMID: 16378885]
33. Wachter RM. Expected and unanticipated consequences of the quality and information technology revolutions. *JAMA.* 2006;295:2780-3. [PMID: 16788133]
34. Hayward RA. All-or-nothing treatment targets make bad performance measures [Editorial]. *Am J Manag Care.* 2007;13:126-8. [PMID: 17335355]
35. U.S. Department of Health and Human Services. Medicare Hospital Value-Based Purchasing Options Paper. Public Listening Session, 12 April 2007. Accessed at www.cms.hhs.gov/AcuteInpatientPPS/downloads/HospitalVBPOptions.pdf on 11 May 2008.
36. Nolan T, Berwick DM. All-or-none measurement raises the bar on performance. *JAMA.* 2006;295:1168-70. [PMID: 16522838]
37. Shojania KG, Sampson M, Ansari MT, Ji J, Doucette S, Moher D. How quickly do systematic reviews go out of date? A survival analysis. *Ann Intern Med.* 2007;147:224-33. [PMID: 17638714]
38. Hayward RA. Performance measurement in search of a path [Editorial]. *N Engl J Med.* 2007;356:951-3. [PMID: 17329704]

Current Author Addresses: Dr. Wachter: Room M-994, Department of Medicine, 505 Parnassus, University of California, San Francisco, San Francisco, CA 94143-0120.

Dr. Flanders: University of Michigan, 3119F Taubman Center, 1500 East Medical Center Drive, Ann Arbor, MI 48109-03766.

Dr. Fee: University of California, Box 0208, 505 Parnassus, San Francisco, CA 94143-0208.

Dr. Pronovost: Departments of Anesthesiology and Critical Care, Surgery, and Health Policy and Management, The Johns Hopkins University School of Medicine, 1909 Thames Street, 2nd Floor, Baltimore, MD 21231.

Author Contributions: Conception and design: R.M. Wachter, S.A. Flanders, P.J. Pronovost.

Drafting of the article: R.M. Wachter, S.A. Flanders, C. Fee, P.J. Pronovost.

Critical revision of the article for important intellectual content: S.A. Flanders, C. Fee, P.J. Pronovost.

Final approval of the article: R.M. Wachter, S.A. Flanders, C. Fee, P.J. Pronovost.

Administrative, technical, or logistic support: R.M. Wachter.