Published in final edited form as:

Ann Surg. 2015 March ; 261(3): 443–444. doi:10.1097/SLA.0000000000000850.

Postoperative Venous Thromboembolism Outcomes Measure – Analytic Exploration of Potential Misclassification of Hospital Quality Due to Surveillance Bias

Jeanette W. Chung, PhD¹, Mila H. Ju, MD, MS¹, Christine V. Kinnier, MD^{1,2}, Min-Woong Sohn, PhD³, and Karl Y. Bilimoria, MD, MS

¹Surgical Outcomes and Quality Improvement Center, Department of Surgery, Feinberg School of Medicine, Northwestern University and Northwestern Memorial Hospital, Chicago, Illinois. 676 St. Clair Street, Arkes Pavilion Suite 6-650, Chicago, Illinois 60611

²Department of Surgery, Massachusetts General Hospital, Boston, MA

³Center for Healthcare Studies – Institute for Public Health and Medicine. Northwestern University. Rubloff Building 10th Floor, 750 North Lake Shore Drive, Chicago, Illinois 60611

Several studies have demonstrated the presence of surveillance bias in the Agency for Healthcare Research and Quality Patient Safety Indicator #12 (PSI12), *Postoperative Venous Thromboembolism (VTE)*, in which hospitals with higher rates of VTE-related diagnostic imaging also have disproportionately higher PSI12 rates.¹ Surveillance bias in PSI12 raises a subsequent question that has received less attention: how accurate is PSI12 in (a) identifying truly poor-quality hospitals (“true positives” for poor VTE outcomes) versus those that only appear to be poor-quality hospitals due to high VTE-imaging rates (“false positives”); and (b) identifying truly high-quality hospitals with low VTE rates (“true negatives”) versus those that only appear to be high-quality hospitals because of inappropriately low VTE imaging rates (“false negatives”). Because incentives are tied to PSI12, it is important to understand the PSI12’s potential for hospital misclassification with respect to quality.

In the absence of a universal screening protocol, VTE-imaging rates are generally below 100%. Imaging is not performed at random, but targeted at patients who are symptomatic or at greater risk for VTE such that patients are prioritized for imaging based on observable signs and symptoms of VTE and factors. This implies that there are diminishing returns to VTE-imaging. Because of this targeting/prioritization, imaging will reveal clinically-significant VTE until imaging rates exceed the true incidence of clinically-significant VTE. At that point, additional imaging will detect non-clinically-significant clots² for which treatment may not have a favorable cost-effectiveness profile. Because the PSI12 numerator counts all diagnosed VTE events as identified on the basis of *International Classification of*

Corresponding Author's Name, Address, Fax, Telephone, E-Mail: Karl Y. Bilimoria, MD MS, Surgical Outcomes and Quality Improvement Center, Department of Surgery, Feinberg School of Medicine, Northwestern University, Northwestern Memorial Hospital, 676 St. Clair Street, Arkes Pavilion Suite 6-650, Chicago, IL 60611. Telephone: 312-695-4853. Fax: 312-695-1462. k-bilimoria@northwestern.edu.

Reprints Address: Reprints will not be available from authors.

*Diseases, 9th Revision (ICD-9-CM) codes,*³ higher VTE-imaging rates can inflate the numerator because ICD-9-CM codes do not differentiate between clinically-significant and subclinical VTE. If subclinical VTE could be reliably distinguished from clinically-significant VTE, subclinical events could be excluded from the numerator to eliminate this source of surveillance bias.

There is, however, another problem with PSI12—one which lurks in the denominator: “all surgical discharges age 18 and older defined by specific DRGs or MS-DRGs and an ICD-9-CM code for an operating room procedure.”³ For any rate measure, all observations in the denominator should be at risk for experiencing the numerator event. In PSI12, the numerator counts patients diagnosed with VTE, but VTE diagnosis depends on imaging. Although all surgical patients are at risk for postoperative VTE, not all patients are at equal risk for imaging due to practice variations, differences in organizational/institutional characteristics (e.g. technological capacity and radiology staffing) and/or heterogeneity in hospital culture. The PSI12 denominator represents the actual population at risk of VTE diagnosis *only* under 100% screening.

To illustrate how this denominator problem along with the inability to differentiate and exclude subclinical VTE from clinically-significant events in the numerator can jointly lead to inaccurate conclusions about hospital quality, consider two hypothetical hospitals, Hospital-X and Hospital-Z. The x-axes in Figure 1 show the number of patients in Hospital-X and Hospital-Z that are in the PSI12 denominator. The primary y- axis shows the true, underlying incidence of VTE in each hospital. In practice, this true rate is unobservable, but for the purposes of this hypothetical illustration, we assume it is known. Dark-shaded regions depict the proportion of each hospital's denominator that develops clinically-significant VTE, while light-shaded regions depict the proportion developing subclinical VTE. Hospital-X has better VTE-related quality of care: its underlying clinically-significant VTE incidence is 20%. Hospital-Z has poorer quality-of-care: its clinically-significant VTE incidence is 40%. Both hospitals have a 30% incidence of subclinical VTE.

The secondary y-axis shows the number of patients that receive VTE-imaging. At 10% surveillance (Line A), both hospitals have identical PSI12 rates of 10%, although Hospital-X has higher quality than Hospital-Z. In Hospital-X, 50% of clinically-significant VTE went undetected, compared to 75% in Hospital-Z. Based on PSI12, both hospitals not only appear the same, but they appear to have better VTE outcomes than they actually do.

At 20% surveillance (Line 2), both hospitals have PSI12 rates of 20%. This captures all clinically-significant VTE in Hospital-X, but 50% of clinically-significant VTE remain undetected in Hospital-Z.

At 40% surveillance (Line 3), both hospitals have PSI12 rates of 40%. However, in Hospital-X, this is comprised of 20 clinically-significant VTE plus an additional 20 subclinical VTE. In Hospital-Z all 40 cases were clinically-significant.

At 100% screening (Line 4), the observed PSI12 rates of Hospital X and Z are 50% and 70%, respectively. The relative ordering is accurate (Hospital-X has lower rates of VTE than

Hospital-Z), although VTE rates are inflated due to inclusion of both clinically-significant and subclinical VTE.

By means of stylized construction, Figure 1 reveals that, holding underlying clinical quality constant, PSI12 may not reflect true clinical quality due to variation in VTE imaging rates. Furthermore, holding surveillance rates constant, PSI12 rates can still fail to reflect true levels of clinical quality due to unobserved heterogeneity in underlying VTE incidence across hospitals.

The analysis of surveillance bias in PSI12 is not simply an empirical or theoretical exercise in measurement science. The potential for PSI12 to misclassify hospitals with respect to quality of care can lead to unintended consequences on multiple levels. Consumers may unwittingly choose “the wrong” hospital based on ostensibly low PSI12 rates if those rates are low because of inadequate surveillance. By the same token, consumers may reject higher quality hospitals because of ostensibly higher PSI12 rates that are high because of more intense surveillance. Payers may misdirect financial rewards towards “false negative” hospitals (worse outcomes than reflected in PSI12) away from “false positive” (better outcomes than reflected in PSI12). Finally, inefficiency in hospital resource allocation may ensue if hospital quality improvement (QI) leaders/teams use PSI12 to guide investment in QI activities.

Developing an empirical approach to quantify the magnitude of misclassification in PSI12 due to surveillance bias hinges on the existence of a gold standard for measuring true underlying VTE incidence that is independent of VTE surveillance imaging. To our knowledge, no such gold standard exists. Nevertheless, PSI12 can be improved if administrative codes are developed and implemented that enable reliable identification and exclusion of subclinical VTE from the measure numerator. Regardless, stakeholders should consider the limitations of this measure when using PSI12 as a basis for evaluating hospital quality.

Acknowledgments

The authors would like to acknowledge support for this work from the Agency for Healthcare Research and Quality (AHRQ R21HS021857, Dr. Bilimoria), the National Institutes of Health (NIH 5T32HL094293, Dr. Ju), and a Center Development Award from Northwestern University and Northwestern Memorial Hospital (Dr. Bilimoria).

Support/Funding: This study was supported by AHRQ R21HS021857 and a Center Development Award from Northwestern University and Northwestern Memorial Hospital to Dr. Bilimoria. Dr. Ju is supported by National Institutes of Health 5T32HL094293.

References

1. Bilimoria KY, Chung J, Ju MH, et al. Evaluation of surveillance bias and the validity of the venous thromboembolism quality measure. *JAMA*. 2013; 310:1482–1489. [PubMed: 24100354]
2. Suh JM, Cronan JJ, Healey TT. Dots are not clots: the over-diagnosis and over-treatment of PE. *Emerg Radiol*. 2010; 17:347–352. [PubMed: 20217446]
3. Agency for Healthcare Research and Quality (AHRQ). [Accessed April 28, 2014] Patient Safety Indicators #12, Technical Specifications, Postoperative Pulmonary Embolism or Deep Vein Thrombosis Rate. Available at: www.qualityindicators.ahrq.gov

Figure 1. Hypothetical Illustration of Hospital Misclassification Due to Surveillance Bias in PSI12

Figure 1 illustrates hypothetical PSI12 and VTE imaging rates for two hypothetical hospitals, "X" and "Z." The x-axis shows the number of patients in the denominator. The primary y-axis shows the true underlying incidence of clinically-significant VTE (dark-shaded region) and subclinical VTE (light-shaded region) per 100 patients in the denominator. The secondary y-axis shows the number of patients that receive VTE imaging per 100 patients in the denominator. By construction, Hospital-X has better VTE outcomes (20% incidence of clinically-significant VTE) than Hospital-Z (40% incidence of clinically-significant VTE). Both hospitals have 30% incidence of subclinical VTE. At 10% surveillance rates (LINE A), both hospitals have the same PSI12 rate: 10%. This is misleading because both hospitals are not of equal quality, and because this underestimates true underlying incidence of clinically-significant VTE in both hospitals (Hospital-X: 20%; Hospital-Z: 40%). At 20% surveillance rates (LINE B), both hospitals have PSI12 rates of 20%. Again, this is misleading because both hospitals are not equal, and because this is an underestimate of Hospital-Z's true VTE incidence. At 40% surveillance (LINE C), both hospitals have PSI12 rates of 40%. This is misleading because it inflates the incidence of clinically-significant VTE in Hospital-X by capturing 20 cases of subclinical VTE. At 100% screening (LINE D), Hospital-X has a PSI12 rate of 50% and Hospital-Z has a PSI12 rate of 70%. While this accurately shows Hospital-X to have better VTE outcomes than Hospital-Z, both PSI12 rates are inflated estimates of the true underlying incidence of clinically-significant VTE in each hospital due to surveillance bias and the inability to identify and exclude subclinical VTE from the numerator of PSI12.