Methods Paper

Comparison of the Abilities of Summary Measures of International Normalized Ratio Control to Predict Clinically Relevant Bleeding

Adam J. Rose, MD, MSc; Thomas Delate, PhD, MS; Al Ozonoff, PhD; Daniel M. Witt, PharmD

Background—Limited research has compared the measures of summarizing international normalized ratio (INR) control over time. Measures that are more predictive of patient outcomes would be preferred as those that are easier to calculate and understand.

Methods and Results—We examined 676 patients who received long-term warfarin therapy to treat atrial fibrillation: 125 patients who experienced major hemorrhage and 551 matched controls who did not. Patient INR control was characterized using various measures, from simple (proportion of INR values in range) to complex (eg, area under the curve above target range, squared) measures. Conditional logistic regression was used to examine the ability of each measure to predict the outcome of clinically relevant bleeding across quintiles of control. All measures were associated with clinically relevant bleeding to some extent: patients with the poorest control had significantly more bleeding events compared with patients with the best control. The measure most strongly associated with bleeding was a combination of percent time in therapeutic range and INR variability (odds ratio of 4.34, comparing the lowest to the highest quintiles of control). The strongest single predictor was INR variability, followed closely by time in therapeutic range. More computationally complex measures, which had been expected to perform better, were not so strongly associated with bleeding.

Conclusions—INR variability was the most strongly associated predictor of clinically relevant bleeding followed closely by time in therapeutic range. Using both measures together had an even stronger association. These findings support continued use of INR variability, time in therapeutic range, or both for research and quality assurance efforts. (Circ Cardiovasc Qual Outcomes. 2015;8:00-00. DOI: 10.1161/CIRCOUTCOMES.115.001768.)

Key Words: anticoagulants ■ case–control studies ■ hemorrhage ■ quality of health care ■ research design ■ warfarin

Warfarin is the most commonly used oral anticoagulant for treatment and prevention of thromboembolic disease and is expected to remain so for the foreseeable future. Although presently there is enthusiasm about the availability of non–vitamin K antagonist oral anticoagulants, working to improve the management of warfarin remains highly relevant because this effort has great potential to benefit many patients.1–3 One important issue in this is the question of how best to measure anticoagulation control achieved with warfarin as valid control measures are a necessary precondition to quality improvement activity.4

Warfarin control is best understood as an intermediate outcome of care, in that control has been correlated with definitive outcomes, such as hemorrhage and stroke.5–11 Intermediate outcomes of care are useful for quality assurance and research efforts because they can be assessed with smaller samples and briefer follow-up than are often needed to confidently assess rates of definitive outcomes.12,13 Using the most predictive intermediate outcome will ensure that efforts to improve performance on intermediate outcomes can achieve the maximum possible benefit for patients. Secondarily, if ≥2 intermediate outcome measures have a similar ability to predict definitive outcomes, the one that is easiest to calculate and understand would be preferred.12,13

Presently, the most commonly used summary measure of international normalized ratio (INR) control for research and quality assurance is percent time in therapeutic range (TTR).5–11,14 In particular, we are referring to individual TTR, which is more tightly linked to outcomes than center-level TTR, as we have shown.15 In addition to TTR, at least 1 other measure of control has been linked to definitive outcomes: INR variability.16,17 Studies on which of these 2 measures is
more predictive of complications are divided with one study suggesting that INR variability is more predictive of complications, another suggesting that TTR is more predictive, and still another suggesting that they are similarly predictive. Nevertheless, studies that have examined predictive ability have found that TTR and INR variability provide complementary information, such that their use together predicts complication events better than either one alone.

There are other measures to summarize INR control that have not been formally compared with the measures in current use. Some of these measures, such as the proportion of INR values in range, have the advantage of being simpler to compute and understand; however, no study to our knowledge has evaluated whether simpler measures are as strongly associated with outcomes as TTR or other more commonly used measures. Conversely, measures of INR control more complex than those currently in use (eg, area under the curve [AUC] above target range, squared) may have better predictive ability, but those comparisons have not been examined either. Nevertheless, if a simpler measure could provide similar predictive ability, its use would be preferred. Therefore, the objective of this study is to compare systematically a roster of widely used and novel summary INR control measures across a range of computational complexity. Our goal was to identify the measure that is most closely correlated with a definitive outcome (clinically relevant bleeding) and, in the event that 2 measures are similarly correlated, to recommend the use of the one that is the simplest to compute and most straightforward to interpret.

Methods

Study Design and Setting

This was an observational, methodological study that used a nested case-control study design. It incorporated data from Kaiser Permanente Colorado (KPCO) administrative databases, including demographic, membership, inpatient and outpatient diagnoses and procedures, outpatient pharmacy, and laboratory measurement data. Manual electronic health record review was used to validate warfarin-related complications as described below.

KPCO provides integrated healthcare services to covered enrollees in the Denver-Boulder metropolitan area. All KPCO patients receiving warfarin therapy are managed by a comprehensive centralized Clinical Pharmacy Anticoagulation and Anemia Service. Details of Clinical Pharmacy Anticoagulation and Anemia Service activities are recorded in the patient’s electronic health record and in a state-of-the-art electronic anticoagulant database, DAWN-AC (4S-Systems Ltd, Cumbria, United Kingdom).

Data extraction and creation of the deidentified study data set was performed at KPCO, whereas data analyses were performed at Boston Children’s Hospital. All aspects of the study were reviewed and approved by the KPCO institutional review board. Subject informed consent was waived as this was a data-only, retrospective study. The
Two expert reviewers using a standardized abstraction form confirmed visit or documentation of a fatal event directly attributable to bleeding. A bleeding event resulting in hospitalization or emergency department accuracy of summary INR control measures (mainly TTR).6,20

(3) no <3 but no >56 days between successive INR measurements; and (4) at least 1 INR measured during the 56 days after the control index date (postindex anchor). Cases were matched ≤5 controls on CHADS, score21 and sex using the greedy method.22 Briefly, the greedy method determines a distance value between all cases and potential controls based on the matching variables, randomly sorts cases and controls, matches the first case to the closest (ie, shortest distance) control, and then moves to the next case and matches it to the closest control.21 A control could be matched to only 1 case.

Study Variables
Along with INR measurements, the following information was compiled for all study patients: (1) age as of respective index date, (2) sex, (3) length of warfarin therapy before respective index date, (4) concomitant nonsteroidal anti-inflammatory drug or antiplatelet use, and (5) history of the following during the 180 days before respective index date: (a) hypertension, (b) diabetes mellitus; (c) stroke/transient ischemic attack, (d) heart failure, (e) renal dysfunction, (f) hepatic dysfunction, (g) alcohol abuse, and (h) previous bleeding/anemia.

The dependent variable was warfarin-related, clinically relevant bleeding. The independent variables were summary 180-day INR control measures (with an in-range INR value between 2.0 and 3.0). Measures are listed below from the simplest to most complex in calculation and interpretation. Each measure is listed with a citation for its previous use (when it exists) plus a brief rationale for why it would be expected to predict complications. We do not discuss measures specifically designed to predict thrombosis here as bleeding is our outcome; however, any measure of time spent above the therapeutic range would have an analogous measure for low INR.

1. Proportion of INR values in range: Calculated by dividing the count of in-therapeutic range INRs by the total count of INRs measured. This measure has been used by sites with limited resources for basic quality assurance purposes12 but has not been tested in regard to its ability to predict definitive outcomes.12 Rationale: Patients with more INR values in range, and thus, fewer values out-of-range, are less likely to experience a warfarin-related complication.

2. INR variability. Calculated as the SD around the mean INR value. There are several subtly different versions of INR variability in use.9 We used the method described by Fihn et al.16 Rationale: Patients whose INR results have the greatest variability will experience rapid swings in their degree of anticoagulation, which may predispose to complications.

3. Percent TTR. Calculated using linear interpolation to assign an INR value to all time points between 2 INR measurements and computing the proportion of time that fell inside the therapeutic range (often referred to as the Rosendaal method).20 This is a widely used summary measure of control.5,6,10 Rationale: Similar rationale to proportion of values in range but corrects for an artifact, namely that INR values may be checked more often when the INR is out of range, thus reducing the proportion of values in range. By using a denominator of time, this artifact does not corrupt the measurement of anticoagulation control.

4. TTR and INR variability. Included as simultaneous predictors to provide complementary information with 2 predictors in the model, rather than just 1.5,6,10
Rationale: Because TTR and INR variability capture separate dimensions of good control, it is likely that using the 2 together captures more information than either alone.

5. Percent time above range. Calculated using similar methods to TTR except that the measurement is assigned according to the proportion of time with INR >3.0. This measure would be expected to be more closely related to risk of bleeding than TTR because although time spent below range affects TTR, it would not be expected to impart a higher risk for bleeding. Another study has examined this measure.

Rationale: Although time spent in range is related to complications, specifically quantifying time spent above range may be more predictive of the specific outcome of bleeding (conversely, time below range may better predict thrombosis). Time spent in range means that the time is not spent out of range. TTR implicitly assumes that time out of range is evenly balanced between high and low, which may not be the case.

6. Percent time with INR >4.0. Calculated using similar methods to TTR except that it encompasses time spent with an INR >4.0. This measure would be expected to be a better predictor of bleeding still because although INR values between 3 and 4 increase the risk of bleeding slightly, it may be that INRs >4.0 increase the risk of bleeding more substantially.

Variation: We also examined percent time above 3.5.

Rationale: Time spent with INR extremely high may be more predictive of bleeding than time spent with INR only mildly elevated.

7. AUC above target range. Calculated using the standard method for evaluating the area under a curve. This measure closely parallels the general thinking that risk for bleeding is a function not only of how far outside range the patient is but also how long he/she remains in outside range.
Table 2. Spearman Correlations for Agreement Among Measures*

<table>
<thead>
<tr>
<th>Measure</th>
<th>Proportion of INRs in Range</th>
<th>TTR</th>
<th>Log INR Variability</th>
<th>AUC Above Range</th>
<th>AUC Above Range Squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion of INRs in range</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TTR</td>
<td>0.85</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log INR variability</td>
<td>0.53</td>
<td>0.42</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AUC above range</td>
<td>0.47</td>
<td>0.38</td>
<td>0.68</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>AUC above range squared</td>
<td>0.48</td>
<td>0.38</td>
<td>0.72</td>
<td>0.99</td>
<td>1.0</td>
</tr>
</tbody>
</table>

*AUC indicates area under the curve; INR, international normalized ratio; and TTR, percent time in therapeutic range.

Results

Baseline Characteristics

A total of 125 cases were matched to 551 controls (Figures 1 and 2). Unmatched controls had statistically significantly lower mean CHADS, score and age than matched controls. Matched patients were predominantly older, with most over age 70 (Table 1). Approximately one-half of patients were women, and patients had a moderate level of comorbid illness and risk for stroke based on CHADS3 score. Because of the matched design, age, sex, CHADS3 score, and other measures of comorbidity were broadly similar between cases (who experienced clinically relevant bleeding) and controls (who did not experience clinically relevant bleeding), although some differences did attain statistical significance. Control patients had better INR control than case patients on every index we measured. Bleeding events were primarily gastrointestinal and epistaxis bleeds. As noted earlier, no patient experienced a thromboembolic event during the study.

Concordance Among Summmary Measures

Correlation between measures ranged from a low of 0.38 (TTR and AUC above range; TTR and AUC above range squared) to a high of 0.99 (AUC above range and AUC above range squared; Table 2). Log INR variability correlated relatively weakly with raw TTR (r=0.42), suggesting that it somewhat measures separate constructs. AUC measures correlated fairly strongly with log INR variability, suggesting that they may blend aspects of this measure. Proportion of INRs in range was strongly correlated with TTR, indicating similar measurement properties.

Ability of Summary Measures to Predict Clinically Relevant Bleeding

All summary measures were predictive of bleeding events (Table 3). For every measure, the lowest control quintile and, in almost every case, the second-lowest quintile had higher rates of bleeding than the highest quintile (Table 4). Differences between the medium, high, and highest control quintiles were not statistically significant (except for the difference between the medium and highest quintiles of percent time with INR >4.0). The most strongly associated measure (as quantified by the largest odds ratio for the lowest control quintile) was the combination of TTR and INR variability. INR variability alone was the next strongest, followed by TTR. The simple measure of proportion of INR values in range was associated with at least a 2-fold increased risk for bleeding in the low and lowest quintiles of control compared with the highest quintile.

Other more complex measures of control did not have strong associations, including percent time far above range (either >4.0 or >3.5, which is not shown), AUC above range, AUC above range squared, and AUC above 2.0, squared (which is not shown).
bleeding in our sample of patients receiving warfarin for AF, based on the odds ratio gradient between the lowest and highest quintiles of control. The next strongest associated measure was TTR. An approach using both measures together performed slightly better than either measure alone. Other measures generally were not as strongly associated, including measures that would seem to capture more directly phenomena that contribute to bleeding risk. For example, we observed relatively poor performance for percent time with INR >4.0, a measure whose rationale is that INR values between 3 and 4 may impart only a trivial risk of bleeding although having a large influence on the TTR value. Our impression is that these measures performed poorly because so few patients in our sample registered any time with an INR >4.0.

Our study suggests that continued reliance on measures currently in widespread use, namely TTR and INR variability, is warranted. Both measures have a long history of strong and consistent associations with clinical outcomes and numerous applications to research and quality improvement. We have also shown, for the first time to our knowledge, that the easiest-to-calculate measure, proportion of INR values in range, is significantly associated with bleeding risk, as well as being strongly correlated with TTR ($r=0.85$ in our sample). This is an important finding because calculating INR variability and TTR requires statistical software and, thus, may not always be feasible to calculate in most clinical practices. Our results suggest that the proportion of INR values in range is a reasonable INR control summary measure for use in quality improvement efforts in practices unable to calculate INR variability and TTR. In fact, we separately calculated an equation to predict TTR when only the proportion of INR values in range is known, in part to facilitate performance comparisons across sites:

Regression equation: expected $TTR_r=-0.0415+1.0511\times$ (proportion of INR values in range).

Intercept: $-0.0415\,(-0.0712, -0.0118; P=0.006).

Slope: $1.0511\, (1.0026, 1.0997; P<0.0001).

Conversely, measures which introduce additional complexity (eg, AUC above target range and AUC above target range, squared) were no more strongly associated with bleeding outcomes than the proportion of INR values in range; thus, their use would seem to be unwarranted.

Limited previous research has attempted to compare multiple measures, and to our knowledge, no previous study has compared the full array of measures examined here. van Leeuwen et al did report that both TTR and INR variability predict warfarin-related complication outcomes among patients with prosthetic heart valves. In that study, the authors found that INR variability was not a statistically significant predictor of bleeding, whereas TTR was. Also in that study, percent time with INR >4.0 had similar predictive ability to TTR. By comparison, we found that INR variability was a slightly better predictor than TTR. We also found that TTR is more closely associated with bleeding compared with percent time with INR >4.0, in contrast to the study by van Leeuwen et al. Finally, our 2 studies share the finding that INR variability and TTR provide complementary information and together predict events better than either one alone.

A previous study found that TTR was more highly predictive of both bleeding and stroke than was INR variability. In that study, as in ours, the 2 measures together provided additional, complementary information, as evidenced by better prediction than either one alone. Still another study by Lind et al found that INR variability was a much stronger predictor of outcomes than TTR. In that study, the predictive ability of TTR was so slight that it did not meaningfully add to that of INR variability when combined in one model. The results of the study by Lind et al are at odds with those of

### Table 3. ORs of Clinically Relevant Bleeding by Quintiles of Control According to 180-Day Summary Measures of INR Control

<table>
<thead>
<tr>
<th>Summary Measure</th>
<th>OR (1-SD Increase/Decrease)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion of INR values in range</td>
<td>1.47* (1.20–1.80)</td>
</tr>
<tr>
<td>Percent TTR</td>
<td>1.46* (1.20–1.77)</td>
</tr>
<tr>
<td>INR variability</td>
<td>1.67* (1.35–2.08)</td>
</tr>
<tr>
<td>TTR+INR variability</td>
<td>1.39* (1.03–1.62)</td>
</tr>
<tr>
<td>INR</td>
<td>1.51* (1.20–1.90)</td>
</tr>
</tbody>
</table>

ORs are given with 95% confidence intervals. Conditional logistic regression was used because of the matched design. INR indicates international normalized ratio; OR, odds ratio; and TTR, percent time in therapeutic range.

*P<0.001.
†P<0.05.

### Table 4. Rates Per 100 Patient Years of Warfarin Treatment for Clinically Relevant Bleeding by 180-Day Summary Measures of International Normalized Ratio Control

<table>
<thead>
<tr>
<th>Summary Measure</th>
<th>Quintile</th>
<th>Highest/Best</th>
<th>High</th>
<th>Medium</th>
<th>Low</th>
<th>Lowest/Worst</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion of INR values in range</td>
<td></td>
<td>3.25</td>
<td>2.95</td>
<td>5.51</td>
<td>8.71</td>
<td>8.64</td>
</tr>
<tr>
<td>Percent TTR</td>
<td></td>
<td>2.48</td>
<td>2.47</td>
<td>5.87</td>
<td>7.95</td>
<td>7.38</td>
</tr>
<tr>
<td>INR variability</td>
<td></td>
<td>2.62</td>
<td>3.76</td>
<td>4.95</td>
<td>6.53</td>
<td>11.47</td>
</tr>
<tr>
<td>TTR+INR variability</td>
<td></td>
<td>2.55</td>
<td>2.91</td>
<td>3.87</td>
<td>7.51</td>
<td>9.87</td>
</tr>
<tr>
<td>Percent time above target range</td>
<td></td>
<td>3.61</td>
<td>4.06</td>
<td>7.34</td>
<td>6.54</td>
<td>7.74</td>
</tr>
<tr>
<td>Percent time with INR &gt;4.0</td>
<td></td>
<td>4.21</td>
<td>8.82</td>
<td>11.22</td>
<td>7.04</td>
<td>10.34</td>
</tr>
<tr>
<td>Area under curve above target range</td>
<td></td>
<td>3.61</td>
<td>3.69</td>
<td>5.57</td>
<td>7.84</td>
<td>8.49</td>
</tr>
<tr>
<td>Area under curve above target range squared</td>
<td></td>
<td>3.61</td>
<td>4.11</td>
<td>4.87</td>
<td>8.00</td>
<td>8.58</td>
</tr>
</tbody>
</table>

INR indicates international normalized ratio; and TTR, percent time in therapeutic range.
the other past studies,6-10 and our present study and are difficult to explain.

Our results point out a limitation of current clinical guidance in that they do not address measures of INR control other than TTR. For example, the European Society of Cardiology position article recommends that if a TTR of 70% cannot be attained, consideration should be given to using therapies other than a vitamin K antagonist.25 Our results suggest that judging the adequacy of vitamin K antagonist therapy by TTR alone may be insufficient and that other parameters may play a role, such as INR variability.

One noteworthy limitation of our study is that we were unable to examine the ability of any measure to predict thromboembolism. Because thromboembolic complications occurred so infrequently in our study sample, our study focused solely on the use of these measures to predict clinically relevant bleeding. Another limitation is that our study focused specifically on patients treated for AF but not on those treated for other conditions, such as venous thromboembolism or mechanical heart valves. We chose to focus on patients with AF to remove potential bias introduced by combining patients with different indications for warfarin therapy and because patients with AF comprise the majority of patients receiving warfarin in most anticoagulation practices. Finally, our study was conducted in a single health system and may not be generalizable to the entire US population of patients treated for AF.

In conclusion, we examined a full range of summary INR controls in regard to their prediction of clinically relevant bleeding events. We found that 2 widely used measures, TTR and INR variability, performed better than other measures we examined. Combining these 2 measures produced more accurate prediction, although only by a small margin. Although these measures are not the easiest to calculate, our results support the continued use of these metrics, either separately or together, for research and quality assurance efforts because of their predictive abilities and relative ease of interpretation.

Sources of Funding
This study was funded by US Department of Veterans Affairs HSRR&D Investigator-Initiated Research 10-374 (principal investigator [PI]: A.J. Rose) and the Kaiser Permanente Colorado Pharmacy Department (PI: T. Delate).

Disclosures
None.

References
20. Rose AJ, Petrakis BA, Callahan P, Mambourg S, Patel D, Hylek EM, Bokhour BG. Organizational characteristics of high- and low-performing


Comparison of the Abilities of Summary Measures of International Normalized Ratio Control to Predict Clinically Relevant Bleeding
Adam J. Rose, Thomas Delate, Al Ozonoff and Daniel M. Witt

Circ Cardiovasc Qual Outcomes. published online September 1, 2015;
Circulation: Cardiovascular Quality and Outcomes is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2015 American Heart Association, Inc. All rights reserved.
Print ISSN: 1941-7705. Online ISSN: 1941-7713

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circoutcomes.ahajournals.org/content/early/2015/09/01/CIRCOUTCOMES.115.001768

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation: Cardiovascular Quality and Outcomes can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Circulation: Cardiovascular Quality and Outcomes is online at:
http://circoutcomes.ahajournals.org//subscriptions/