Original Article

Improving Anticoagulation Measurement

Novel Warfarin Composite Measure

Zayd Razouki, MD, MSc; James F. Burgess, Jr, PhD; Al Ozonoff, PhD; Shibei Zhao, MPH; Dan Berlowitz, MD, MPH; Adam J. Rose, MD, MSc

Background—Percent time in therapeutic range (TTR) and international normalized ratio (INR) variability both measure warfarin control and are associated with outcomes independently. Here, we examine the advantages of a warfarin composite measure (WCM), which summarizes the 2 when measuring patient outcomes. We also examine how the measure chosen would affect anticoagulation clinic performance rankings.

Methods and Results—We constructed WCM using an equally weighted method, adding standardized TTR to standardized log-transformed INR variability using 103,897 warfarin-experienced patients from 100 anticoagulation clinics. We examined the association of WCM with ischemic stroke, major bleeding, and fatal bleeding, using a subset of patients with atrial fibrillation (n=40,404). We divided patients into quintiles based on their level of control for TTR, log INR variability, and WCM. We calculated the hazard ratios for ischemic stroke, major bleeding, and fatal bleeding stratified by these quintiles. WCM hazard ratios for stroke and fatal bleeding showed the largest difference between excellent control and poorest control quintile compared with TTR and log INR variability, but not for major bleeding. In addition, we compared site rankings obtained using each of our 3 performance measures. Kappa scores for identifying outlier and nonoutlier clinics between WCM and its components were moderate (κ=0.56 for TTR and κ=0.62 for log INR variability) but was weak between TTR and log INR variability (κ=0.13).

Conclusions—WCM produces the largest range of risk for warfarin complications, widening the floor ceiling effects that limit the use of TTR and INR variability as separate measures. Anticoagulation clinics ranking changed considerably according to the anticoagulation measure that was selected. (Circ Cardiovasc Qual Outcomes. 2015;8:600-607. DOI: 10.1161/CIRCOUTCOMES.115.001789.)

Key Words: atrial fibrillation ■ blood coagulation ■ international normalized ratio ■ quality of care ■ stroke ■ warfarin
WHAT IS KNOWN

- Percent time in therapeutic range (TTR, a measure of anticoagulation intensity) and international normalized ratio (INR) variability (a measure of anticoagulation stability) are 2 different independent measures that are associated with adverse events with warfarin therapy.
- TTR and INR variability together provide more information about risk of warfarin adverse events than each individually.

WHAT THE STUDY ADDS

- Warfarin composite measure is a novel and valid summary score that combines TTR and INR variability, thus accounting for both the intensity and stability of warfarin therapy.
- Warfarin composite measure produces the largest range of risk for warfarin complications, reducing the measurement constrains that limit the use of TTR and INR variability as separate measures.
- Warfarin composite measure, as a performance measure, reconciles the discordance in ranking anticoagulation clinics that occurs when using TTR or INR variability separately.

WHAT THE STUDY ADDS

- Warfarin composite measure is a novel and valid summary score that combines TTR and INR variability, thus accounting for both the intensity and stability of warfarin therapy.
- Warfarin composite measure produces the largest range of risk for warfarin complications, reducing the measurement constrains that limit the use of TTR and INR variability as separate measures.
- Warfarin composite measure, as a performance measure, reconciles the discordance in ranking anticoagulation clinics that occurs when using TTR or INR variability separately.

Measure in large health integrated care systems. INR variability measures stability, but not appropriate intensity, of anticoagulation. Similar to TTR, high INR variability (ie, unstable anticoagulation) has been shown to be associated with warfarin-associated complications, independently from TTR. Despite its linkage to important clinical outcomes, INR variability has not been used as a performance measure to the best of our knowledge.

Previous studies have suggested that there is an added value when using both measures. The 2 measures together provide more information about the risk of adverse events than each one separately. However, there is no single measure that summarizes the quality of anticoagulation control using both TTR and INR variability. Therefore, in this study, we had 2 aims. Our first aim was to construct a composite measure that encompasses both TTR and INR variability. We aimed to illustrate the added benefit of using this intermediate composite measure outcome (warfarin composite measure [WCM]) by comparing the association of this measure to each of its components when assessing the risk of warfarin-associated complications. Our second aim was to examine the level of agreement between different anticoagulation measures in identifying different levels of control and examining the feasibility of using WCM to profile anticoagulation clinics (ACC) in the Veterans Health Administration (VHA). We demonstrate how the definition of what might constitute good or poor anticoagulation control changes, and how ACC ranking differs, depending on the choice of the 3 anticoagulation measures (ie, TTR, INR variability, and WCM).

A composite measure that provides a global measure of anticoagulation quality that is tightly linked to warfarin complications would encourage providers to focus on all components of anticoagulation control, not just those measured by the current standard of care, TTR. A comprehensive performance measure would also facilitate quality improvement efforts that can succeed in improving both dimensions of good control.

Methods

Data Collection

The database for this study has been described elsewhere. In brief, the Veterans Affairs Study to Improve Anticoagulation (VARIA) included all patients considered to be receiving oral anticoagulation therapy from the VA between October 1, 2006 and September 30, 2008. For this study, we only included those patients who had previously received warfarin for at least 6 months within the VHA system (ie, experienced warfarin users). Patients who completed their 6-month initiation phase before October 1, 2006 were included from the beginning of the period; other patients entered the database after completing 6 months of therapy. Patients who are prescribed warfarin from the VA are fully managed in special ACCs, whereas patients who are prescribed warfarin outside the VA are not included in this data set.

The study was approved by the Institutional Review Board of the Bedford VA Medical Center.

Patients

Our full sample included all patients (n=103,897) who already received warfarin for at least 6 months from the VA during the FY07 to FY08 (ie, at least 30 days’ worth dispensed by the pharmacy). We included INR tests within the VA when patients were on warfarin, that is, when a patient was either (1) in possession of warfarin or (2) having INR tests every 42 days. A similar approach was used to define time on warfarin. We defined the period of warfarin possession as the duration of the most recent VA prescription for warfarin, plus 30 days. Because patients may be instructed to take half doses of warfarin, we recognize that going >30 days beyond the end of a prescription does not necessarily indicate that warfarin therapy has stopped. We therefore also allowed a consistent pattern of INR measurements (ie, every 42 days or less) to indicate that a patient was still being managed. Patients with chronic liver disease were only included if they received warfarin; frequent INR tests for such patients might have been performed to monitor liver function.

INR values during hospitalization and 14 days after discharge were excluded in this study and were not used to calculate the 3 anticoagulation measures, reasoning that warfarin management might be held for clinically important scenarios, including procedures and bleeding events. We excluded patients whose primary indication to receive warfarin was valvular heart disease, because their target INR range might be 2.5 to 3.5 rather than the conventional target of 2 to 3, which would change the calculation of TTR. We also excluded patients who only recorded INR values 1.3 and lower, reasoning that most such patients received INR tests for reasons unrelated to warfarin management (eg, frequent emergency department visits).

For our first aim, comparing the risk of warfarin-associated complications for each anticoagulation measure, we limited our sample to patients who were receiving anticoagulation for atrial fibrillation, as opposed to other indications, such as venous thromboembolism. This was done to create a sample of patients with relatively uniform risk for the outcomes of interest. We also limited this portion of our study to patients >65 years, who were enrolled in fee-for-service Medicare. Patients with Medicare Advantage plans were excluded because that program does not produce itemized claims data, which would lead to undercounting of events that occurred outside the VHA. We used a merged VHA-Medicare database to maximize the possibility of capturing all relevant events and complications related to warfarin therapy. After these exclusions, we were left with (n=40,404) patients for the study of clinical outcomes.

For the second aim of this study (ie, examining the impact of the selected measure on site performance profiling), we used our full sample (n=103,897), as described above.

Laboratory Values

To calculate measures of INR control, we included INR values obtained within the VA system when patients were on warfarin, that is,
when a patient was either (1) in possession of warfarin or (2) having INR tests at least every 42 days. We defined the period of warfarin possession as the duration of the most recent VA prescription for warfarin, plus 30 days. Because patients may be instructed to take half doses of warfarin, we recognize that going >30 days beyond the end of prescription does not necessarily indicate that warfarin therapy has stopped. We therefore allowed a consistent pattern of INR measurements (ie, every 42 days or less) to indicate that a patient was still being managed.

We excluded INR tests measured while the patient was hospitalized within the VA system. Patients who are hospitalized may receive temporary parental anticoagulation (eg, with heparin) or no anticoagulation, so out-of-range INR values while hospitalized may be intentional and do not necessarily reflect poor quality care.

Calculating Percent Time in TTR
We calculated TTR using the method by Rosendaal, which uses linear interpolation to assign an INR value to each day between successive observed INR values. Gaps of 56 days or more between INR values are not interpolated. After interpolation, we calculated the percentage of time during which the interpolated INR values lie between 2.0 and 3.0 (from 0% to 100%).

Calculating INR Variability
INR variability was calculated using the method of Fihn, which reflects the degree to which each patient's INR varies around his or her previous INR. This formula does not take into account the intensity of warfarin therapy; a patient is most stable if his or her INRs are around the same level even if the INR is constantly above or below the limits of the target range. We log-transformed INR variability to reduce the leverage and undue influence of extreme observations on our estimates. More negative INR variability values are actually consistent with better control. For example, a patient with log INR variability of ~5.0 has a much more stable anticoagulation regimen compared with another patient with a log INR variability of ~0.75.

Constructing the Composite Measure (WCM)
We constructed WCM using an equally weighted approach, where each individual measure is simply added to the other. An equally weighted approach is the default approach when constructing a composite measure and when there is no clear rationale that a measure should be weighted different from the other. Furthermore, we justify weighting TTR and log INR variability equally based on the absolute values of the coefficients of variation that were found to be comparable in magnitude (35.8 for TTR and 42.5 for log INR Variability). In addition, both TTR and log INR variability have similar c-statistics when predicting our study outcomes. For predicting major bleeding, including covariates, TTR had a c-statistic of 0.62, whereas log INR variability had a c-statistic of 0.61. Similarly, for predicting fatal bleeding, including covariates, TTR had a c-statistic of 0.70, whereas log INR variability had a c-statistic of 0.69. Finally, for predicting ischemic stroke, including covariates, TTR had a c-statistic of 0.67, whereas log INR variability had a c-statistic of 0.67. The combination of similar coefficients of variation and similar goodness of fit statistic when predicting outcomes suggests that the 2 measures can be combined with equal weighting into a composite measure without unbalancing the effect of each measure.

Because the unit measurement for TTR and log INR variability are different, we converted both TTR and log INR variability values into a standardized Z score. Standardizing the measures into a unit-less measure allows summing the 2 measures despite differences in their unit measurements. TTR and log INR variability have opposite polarity; high TTR represents good anticoagulation, whereas high log INR variability represents poor anticoagulation. Therefore, we reversed the polarity of log INR variability to achieve a unidirectional relationship between the 2 measures so that they could be combined:

\[
WCM = \left[\text{Standardized TTR} + (\text{Standardized log INR variability} * -1)\right]
\]

Primary Outcomes
To examine the first study aim, we compared 3 main outcomes: ischemic stroke, major bleeding, and fatal bleeding, which is a subset of major bleeding where death occurred within 30 days. We compared these outcomes across our 3 anticoagulation measures: TTR, log INR variability, and WCM.

The outcome of death was assessed using the VA Vital Status Mini-File, which is known to be an accurate source for date of death. We identified major bleeding using a method we have previously described. Briefly, we began with all International Classification of Diseases-Ninth Revision codes that might describe a hemorrhagic event and then retained the events that fulfilled 1 of 4 criteria: fatal hemorrhage, bleeding into a critical anatomic site, bleeding associated with transfusion, or bleeding which was identified as the main reason for a hospital stay. This is an adaptation of the definition of major bleeding of the International Society of Thrombosis and Haemostasis to a large, automated data set. Ischemic strokes were also defined using International Classification of Diseases-Ninth Revision codes. We used a restrictive set of codes to define our stroke outcome, emphasizing specificity over sensitivity. Namely, we used 433.x1 and 434.x1, Occlusion of... artery... with cerebral infarction. We only used inpatient codes for this outcome, as per our earlier work, but did allow codes in either the primary or secondary position to define a stroke.

Covariates
We adjusted for various clinical variables that are known to be associated with outcomes. For ischemic strokes, we adjusted for all of the variables included in the CHADS2 score (congestive heart failure, hypertension, age, diabetes mellitus, and previous stroke/transient ischemic attack). For major and fatal bleeding, we adjusted for many of the variables contained within the HAS-BLED bleeding risk score, including age, hypertension, chronic kidney disease, liver disease, prior stroke, alcohol abuse, and drug abuse. We did not adjust for other components of HAS-BLED, such as history of major bleeding and antiplatelet therapy, because we could not identify those elements with confidence in our data set.

Anticoagulation Clinics
Patients receiving warfarin are managed by specialized ACCs in the VA healthcare system where each clinic is attached to a medical care site. There are 128 ACC of care in the VHA system. We excluded 28 of these clinics from our study because our data-checking procedures revealed possible problems with data completeness at those clinics that have been fully described in a published article. We also excluded several months of data from an additional 14 clinics for the same reason.

Risk Adjustment Model
We used our full sample (n=103,897) to examine how using different anticoagulation measure as performance measures could affect ACC ranking. A risk adjustment model that our group previously used to adjust TTR was also used to risk-adjust log INR variability. The derivation and validation of this model, including the definitions of our clinical variables, using International Classification of Diseases-Ninth Revision codes, is fully described in a previously published article. Table I in the Data Supplement shows covariates included in our model, with applicable β-coefficients for TTR and log INR variability.

Variables contained within the model include sociodemographic factors (age, sex, race, area-level poverty, and distance from care), factors characterizing warfarin therapy (indication for therapy, how long ago therapy began), and factors characterizing illness burden (physical and mental comorbid health conditions, number of medications, and number of hospitalizations). We computed new coefficients for these variables, first to examine the association with log INR variability, then to examine their associations with WCM.

Statistical Analyses
To compare the association of WCM and its components (TTR and log INR variability) with clinical outcomes, we divided each measure into
quintiles and assigned patients into these quintiles based on their degree of control with each measure. Patients were often assigned to different quintiles using different measures. First, we calculated the crude percentage of our outcomes (ischemic stroke, major bleeding, and fatal bleeding) within each quintile. Then, we calculated hazard ratios (HRs) for each of our outcomes (ischemic stroke, major bleeding, and fatal bleeding) using Cox regression analyses. The quintile excellent control was used as a reference category. We visually compared the HRs for outcomes across quintiles. We were not able to perform statistical comparison testing to query whether differences in HRs were statistically significant, because individual patients were not distinctly separated across the 3 studied measures, and there is no statistical test that can be used to compare across the 3 independent variables. Therefore, estimates are to be compared visually.

To examine whether WCM better discriminates patient outcomes compared with TTR and log INR variability, we calculated the c-statistic for each measure examining major bleeding, fatal bleeding, and stroke. We performed a full sample bootstrap with 300 iterations of resampling. For each iteration, we calculated the c-statistic and calculated the 2.5% and 97.5% quintiles of the resulting distribution as the lower and upper limits of the empirical confidence interval.

To illustrate the distribution and frequency of our sample across various levels of anticoagulation control according to the 3 anticoagulation measures, we created 3 5×5 tables, dividing each measure into quintiles based on the level of control.

To profile ACCs according to the 3 anticoagulation measures, we calculated descriptive statistics for TTR, log INR variability, and WCM for all clinics included in our study. For each clinic, we calculated mean observed (O) clinic values for TTR, log INR variability, and WCM. We then applied a linear regression model to risk-adjust log INR variability, calculating the expected (E) value for each clinic, and then we calculated observed minus expected value for each clinic. Adjusted TTR was calculated in a previous published study using a similar risk-adjustment model.28 The effect sizes that are associated with both measures were used to calculate the adjusted values for WCM. Clinics were marked as outliers in their performance, if the absolute difference between an observed value and an expected value exceeded 1 SD, with a SD measured by the entire sample of patients. Using SD-based yardsticks to define important effect sizes is common when no consensus exists about what would be an important difference:26 a full SD is generally considered to be a large difference.

We also examined the correlation among the ranking order of ACCs (1) before and after adjustment and (2) among the 3 anticoagulation measures. Finally, we calculated a chance-corrected weighted (1) before and after adjustment and (2) among the 3 anticoagulation measures. We also examined the correlation among the ranking order of ACCs (1) before and after adjustment and (2) among the 3 anticoagulation measures. Finally, we calculated a chance-corrected weighted k score to assess the level of agreement among our measures in defining high and low performance outlier clinics.

Results
Comparing the Association and the Discriminative Ability for Each Anticoagulation Measure (TTR, INR Variability, and WCM) With Warfarin-Associated Complications
For this analysis, we included a subset of patients (n=40404) above the age of 65 years with atrial fibrillation. Our sample was (98%) male with a mean age of 76 years. Patients were observed for ≤2 years, with an average observation time of 14 months. The mean values for TTR and log INR variability are reported for each quintile of anticoagulation control for this cohort in Table 1.

Time interpolated to measure TTR was, 44610.22 person-years, representing 80% of total time between INR intervals during our study period. The remaining time was excluded because time intervals longer than 56 days could not be interpolated according to the methods by Rosendaal. Such values were excluded from all calculations, not just those for TTR.

Overall, 3.1% of patients experienced an ischemic stroke event, 6.4% experienced a major bleeding, and 0.9% experienced fatal bleeding, a subset of major bleeding. Tables II to IV in the Data Supplement show the frequency of major bleeding, fatal bleeding, and ischemic stroke within each quintile for each anticoagulation measure.

Compared with log INR variability, WCM HRs’ for major bleeding showed the largest difference in outcomes between excellent control and other levels of anticoagulation control. Specifically, compared with excellent control, the HR for major bleeding was greater at levels of very poor control (HR: WCM=1.95 versus log INR variability=1.58) and poor control (HR: WCM=1.57 versus log INR variability=1.25). Compared with TTR, WCM had a similar HR between excellent control and other levels of control. Specifically, compared with excellent control, the HR for major bleeding was similar at levels of very poor control (HR: WCM=1.95 versus TTR=1.99). One exception was a slightly higher HR for major bleeding at the level of poor control (HR: WCM=1.57 versus TTR=1.35). Table 2 shows the HRs for major bleeding according to quintiles of anticoagulation.

Compared with TTR and log INR variability, WCM HRs’ for fatal bleeding showed the largest difference in outcomes between excellent control and other levels of control. Specifically, compared with excellent control, the HR for fatal bleeding was greater with very poor control (HR: WCM=3.04 versus TTR=2.18 versus log INR variability=2.43) or poor control (HR: WCM=1.69 versus TTR=1.46 versus log INR variability=1.29). Table 3 shows the HRs for fatal bleeding according to quintiles of anticoagulation control.

Similarly, compared with TTR and log INR variability, WCM HRs’ for ischemic stroke showed the largest difference in outcomes between excellent control and other levels of control. Specifically, compared with excellent control, the HR for

Table 1: Illustrating Mean Levels of Anticoagulation Control According to TTR and Log INR Variability for (n=40404) Patients With Atrial Fibrillation, Divided into Quintiles

<table>
<thead>
<tr>
<th>Variables</th>
<th>n</th>
<th>Mean Age</th>
<th>Mean and (Range) TTR</th>
<th>Mean and (Range) Log INR Variability</th>
<th>Mean and (Range) WCM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall mean (SD)</td>
<td>40404</td>
<td>76.78</td>
<td>63.56 (21.2)</td>
<td>−3.41 (1.35)</td>
<td>0 (1.0)</td>
</tr>
<tr>
<td>Quintile 1: poorest</td>
<td>8080</td>
<td>77.02</td>
<td>37.61 (0 to 47.11)</td>
<td>−1.78 (−11.51 to 4.48)</td>
<td>−2.16 (−7.67 to 1.13)</td>
</tr>
<tr>
<td>Quintile 2: poor control</td>
<td>8081</td>
<td>76.81</td>
<td>54.14 (47.12 to 60.89)</td>
<td>−2.93 (−4.48 to 3.82)</td>
<td>−0.6 (−1.13 to 0.13)</td>
</tr>
<tr>
<td>Quintile 3: fair control</td>
<td>8081</td>
<td>76.68</td>
<td>64.58 (60.89 to 71.15)</td>
<td>−3.48 (−3.82 to 3.21)</td>
<td>0.28 (−0.13 to 0.67)</td>
</tr>
<tr>
<td>Quintile 4: good control</td>
<td>8081</td>
<td>76.71</td>
<td>74.24 (71.15 to 81.76)</td>
<td>−3.99 (−3.21 to 2.40)</td>
<td>1.07 (0.67 to 1.52)</td>
</tr>
<tr>
<td>Quintile 5: excellent control</td>
<td>8081</td>
<td>76.71</td>
<td>87.22 (81.76 to 100.0)</td>
<td>−4.86 (−2.40 to 3.58)</td>
<td>2.29 (1.52 to 7.63)</td>
</tr>
</tbody>
</table>

INR indicates international normalized ratio; and TTR, therapeutic range.
stroke was greater with very poor control (HR: WCM=2.43 versus TTR=2.10 versus log INR variability=1.74) or poor control (HR: WCM=1.69 versus TTR=1.46 versus log INR variability=1.29). Table 4 shows the HRs for stroke according to quintiles of anticoagulation control.

WCM did not better discriminate patient level outcomes compared with TTR and INR variability. There was no significant improvement in the c-statistics across the 3 models when examining major bleeding, fatal bleeding, and ischemic stroke (Table V in the Data Supplement).

### Comparing 3 Anticoagulation Measures at Various Control Levels and as Performance Measures When Ranking ACC

#### Patient Population

For these analyses, we included our full sample of 103,897 patients who received warfarin anticoagulation from 100 VA ACCs. Baseline characteristics are described in Table I in the Data Supplement. The majority of our sample were men (98%) with an average age of 72 years. The main indication for warfarin anticoagulation was atrial fibrillation (64%), with 29% for venous thromboembolic disease and 9% for other indications. Our sample had a substantial burden of physical illness. For example, 39% had diabetes mellitus and 31% had heart failure. The median number of patients managed at each clinic was 862 (interquartile range, 559–1301.8). The mean heart failure. The median number of patients managed at each clinic was 862 (interquartile range, 559–1301.8). The mean heart failure.

#### Table 2. Hazard Ratios for Major Bleeding Comparing 3 Anticoagulation Measures, TTR, Log INR Variability, and WCM

<table>
<thead>
<tr>
<th>Quintiles Measurements</th>
<th>TTR</th>
<th>Log INR Variability</th>
<th>WCM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poorest control</td>
<td>2.18 (1.57–3.03)</td>
<td>2.49 (1.79–3.47)</td>
<td>3.04 (2.16–4.28)</td>
</tr>
<tr>
<td>Poor control</td>
<td>1.49 (1.06–2.10)</td>
<td>1.45 (1.02–2.08)</td>
<td>1.67 (1.16–2.42)</td>
</tr>
<tr>
<td>Fair control</td>
<td>1.19 (0.84–1.70)</td>
<td>1.18 (0.81–1.72)</td>
<td>1.45 (0.99–2.11)</td>
</tr>
<tr>
<td>Good control</td>
<td>0.92 (0.68–1.23)</td>
<td>1.18 (0.81–1.71)</td>
<td>1.18 (0.79–1.75)</td>
</tr>
<tr>
<td>Excellent control</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
</tbody>
</table>

Effect across each column was statistically significant at the $P<0.001$ level.

#### Table 3. Hazard Ratios for Fatal Bleeding Comparing 3 Anticoagulation Measures, TTR, Log INR Variability, and WCM

<table>
<thead>
<tr>
<th>Quintiles Measurements</th>
<th>TTR</th>
<th>Log INR Variability</th>
<th>WCM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poorest control</td>
<td>2.18 (1.57–2.53)</td>
<td>1.58 (1.40–1.79)</td>
<td>1.95 (1.72–2.21)</td>
</tr>
<tr>
<td>Poor control</td>
<td>1.35 (1.19–1.54)</td>
<td>1.25 (1.10–1.42)</td>
<td>1.57 (1.38–1.79)</td>
</tr>
<tr>
<td>Fair control</td>
<td>1.19 (1.04–1.35)</td>
<td>1.15 (1.01–1.31)</td>
<td>1.18 (1.03–1.35)</td>
</tr>
<tr>
<td>Good control</td>
<td>0.95 (0.83–1.09)</td>
<td>1.02 (0.89–1.17)</td>
<td>1.05 (0.92–1.21)</td>
</tr>
<tr>
<td>Excellent control</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
</tbody>
</table>

Effect across each column was statistically significant at the $P<0.001$ level.

#### Table 4. Hazard Ratios for Ischemic Stroke Comparing 3 Anticoagulation Measures, TTR, Log INR Variability, and WCM

<table>
<thead>
<tr>
<th>Quintiles Measurements</th>
<th>TTR</th>
<th>Log INR Variability</th>
<th>WCM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poorest control</td>
<td>2.10 (1.76–2.51)</td>
<td>1.74 (1.46–2.08)</td>
<td>2.43 (2.01–2.92)</td>
</tr>
<tr>
<td>Poor control</td>
<td>1.46 (1.22–1.76)</td>
<td>1.29 (1.07–1.55)</td>
<td>1.69 (1.39–2.05)</td>
</tr>
<tr>
<td>Moderate control</td>
<td>1.08 (0.89–1.31)</td>
<td>1.16 (0.96–1.40)</td>
<td>1.30 (1.06–1.59)</td>
</tr>
<tr>
<td>Good control</td>
<td>0.93 (0.76–1.14)</td>
<td>0.96 (0.79–1.17)</td>
<td>1.18 (0.96–1.46)</td>
</tr>
<tr>
<td>Excellent control</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
</tbody>
</table>

Effect across each column was statistically significant at the $P<0.001$ level.

There was considerable discordance among our measures in classifying patients’ anticoagulation control. For example, 2.73% patients were classified as having excellent control according to log INR variability (ie, in the quintile of excellent control) but poorest control according to TTR (ie, in the quintile of worse control). Also, 0.71% of patients were classified to have excellent control according to TTR but poorest control according to log INR variability. Such instances of relatively extreme discordance remained but were less frequent between WCM and its components, where only 0.05% of patients were classified to have excellent control according to TTR and poorest control according to WCM. Similarly, 0.46% of patients were classified to have the poorest control according to WCM but excellent control according to log INR variability. No patient classified to have excellent control according to WCM was classified to have poorest control according to log INR variability (Tables VII–IX in the Data Supplement).

As would be expected, the risk-adjusted clinic rankings were strongly correlated between TTR and WCM ($r=0.93$) and between log INR variability and WCM ($r=0.90$), but were less correlated between TTR and log INR variability ($r=0.70$). The correlation also appeared modest when examining the magnitude of observed minus expected value (rather than clinic rankings) between TTR and log INR variability ($r=0.51$, $P<0.001$; Figure 1).

A total of 40 ACCs were identified as performance outliers by at least 1 of our 3 measures. Profiling ACCs based on TTR identified 10 high and 10 low performing outliers using our definition of 1 SD difference from the mean (Figure I in the Data Supplement); profiling ACCs according to log INR variability identified 14 high and 12 low performing outliers (Figure II in the Data Supplement); and profiling ACCs according to WCM identified 13 high and 16 low outliers (Figure 2). Although there was some overlap regarding which clinics were identified as performance outliers, it was only limited overlap. One clinic (Site 74) was actually identified as a high performing outlier by WCM and a low performing outlier by TTR. No site was identified as a performing outlier based on WCM only; all those sites were also a performing outlier on at least one of the component measures.

Our results indicate that the degree of chance-corrected agreement in identifying sites that are considered outliers (both high and low) and nonoutlier clinics, measured by weighted κ score ($\kappa$) between WCM and its components...
is moderate ($\kappa = 0.56$ for TTR and $\kappa = 0.62$ for log INR variability), whereas the level of agreement between TTR and log INR variability was weak ($\kappa = 0.13$).

**Discussion**

We found that a novel composite measure (ie, WCM) produces the largest range of risk for adverse clinical events when benchmarked against excellent anticoagulation control when compared with its components (ie, TTR and log INR variability). This study demonstrates the usefulness of a single composite measure, as opposed to using 2 measures at once. This approach allows providers to address warfarin management in a comprehensive way, addressing the 2 crucial dimensions of anticoagulation control simultaneously. Our results concord with findings of previous studies that demonstrated an advantage in measuring anticoagulation quality using both its important dimensions, intensity and stability.\textsuperscript{14,15}

Measuring anticoagulation using WCM, as a summary score, widens the ceiling and floor measurement effects that limit the use of single measures. We observed that both TTR and INR variability may experience floor or ceiling effects with flat risk associations at either end of the INR control scales. In contrast, WCM does not demonstrate similar floor or ceiling effects, with roughly linear difference in risk across our sample.

The modest agreement between TTR and log INR variability suggests that it is inherently challenging to define what constitutes good or poor anticoagulation control according to a single anticoagulation measure. There are a significant number of patients that could be classified as having poor control according to one measure, whereas having good control according to the other. WCM has the advantage of including both dimensions of what constitutes good anticoagulation control, namely appropriate intensity and stability.

![Figure 1](image1.png)

**Figure 1.** Demonstrate the correlation between site therapeutic range (TTR) and site log international normalized ratio (INR) variability for 100 anticoagulation clinics, measured by (observed minus expected value [O–E]). For outlier sites, observed TTR differs from expected by $\geq 4.9\%$. For outlier sites, observed log INR variability differs from expected by $\geq 0.29$.

![Figure 2](image2.png)

**Figure 2.** Overall anticoagulation clinic performance as measured by observed mean site log warfarin composite measure (WCM) minus expected mean site WCM. The horizontal line indicates the overall mean WCM ($-0.018$). Each vertical bar indicates the 95% confidence interval (CI) around a site’s expected mean WCM, whereas each open dot indicates the observed mean site WCM. For outlier sites (where observed WCM differs from expected by $\geq 0.037$), the dot is replaced by an asterisk.
Defining the level of anticoagulation control according to WCM is closely linked to the patient outcomes that inherently matter; it is a truism that intermediate outcome measures should be as tightly linked to definitive outcomes as possible to help ensure that we can work to improve the dimensions that are most related to important clinical outcomes that matter most.\(^5\)

Furthermore, we illustrate the feasibility of profiling ACCs according to 3 different anticoagulation measures (TTR, INR variability, and WCM), and how ACCs performance and ranking would change depending on the measure chosen. Despite the fact that all 3 measures examine anticoagulation control, we illustrate how the choice of one measure can significantly change what might constitute a performance outlier. The concordance between TTR and INR variability in identifying high and low performing clinics was relatively slight. This is not surprising because these 2 measures assess different domains of anticoagulation control. Our results illustrate the general point that ranking sites can markedly change if different measures are adopted. Therefore, the development of a credible composite measure, which combines both TTR and INR variability, is an important step toward reconciling the discordance of using a single measure.

Establishing criterion validity (ie, the ability to predict clinical events) is only part of making a case for a new performance measure. Health systems will also need to convince stakeholders, such as clinical providers, that the new measure is not only valid, but also feasible and easily understood. It is a concern that the new composite measure is even more complex than either of the measures it would replace; however, both the component measures also require a compute to calculate them, so it may be that a certain point of complexity may have been passed already. Nevertheless, it should be recognized that the factual basis for any measure is not sufficient to ensure that it will be embraced by end users.

We acknowledge several limitations in our study. First, our patient-level outcomes were detected using automated data and were not validated by chart review. However, we used thoughtful approaches to using these automated data to maximize the accuracy of our findings. Second, ideally we should have defined levels of anticoagulation control according to clinically meaningful cutoff points. We used quintiles, a generic approach, largely because of a lack of accepted standards regarding what would be a clinically meaningful difference. Third, our sample was overwhelmingly men (98%) and reflects the population of VA patients, limiting to the generalizability of our results to other populations. Fourth, we were not able to generate statistical comparisons across our 3 different measures, because our patients’ groups were not distinctly separated. Fifth, we adjusted for many confounding factors for major bleeding and fatal bleeding; however, we were not able to adjust for all factors, including history of prior bleeding and the use of antiplatelets. In addition our administrative data could not account for other important factors that are related to warfarin complications, including adherence to warfarin therapy. Finally, in addition to our observed minus expected value approach, there are other ways to profile sites based on Bayesian methods, which can be especially important when dealing with small sample size.\(^6\) However, because of our large sample size at even the smallest sites of care, this was not necessary here.

In conclusion, our intermediate composite measure (WCM) summarizes 2 aspects of anticoagulation control, namely intensity and stability. Our findings suggest that WCM has validity in measuring anticoagulation control and is closely linked to adverse events. WCM produced the largest range of risk for ischemic stroke and fatal bleeding compared with its components; this was only true when comparing WCM with log INR variability for major bleeding but not TTR.

We would recommend the use of WCM as a performance measure for clinic profiling, as it seems to identify different clinics as outliers. Both on an empirical and a theoretical basis, WCM captures more completely relevant dimensions of anticoagulation control than its component measures.

**Acknowledgments**

The opinions expressed are those of the authors and do not necessarily represent the official views or policies of the US Department of Veterans Affairs.

**Sources of Funding**

Dr Razouki is an awardee of the Office of Academic Affiliations for a Postdoctoral Training Program in Health Services Research, Durham VA Medical Center (TPM 21–026). This study was supported by the Health Service Research & Development IIR-10 to 374 (PI: Dr Rose).

**Disclosures**

None.

**References**


Improving Anticoagulation Measurement: Novel Warfarin Composite Measure
Zayd Razouki, James F. Burgess, Jr, Al Ozonoff, Shibei Zhao, Dan Berlowitz and Adam J. Rose

Circ Cardiovasc Qual Outcomes. 2015;8:600-607; originally published online September 29, 2015;
doi: 10.1161/CIRCOUTCOMES.115.001789
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/8/6/600

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/