PROFESSOR RONGQIAN YAO (Orcid ID : 0000-0003-4923-3625)

MR. SIQI LAN (Orcid ID : 0000-0002-0701-1912)

Article type : Article

Developing and validating a clinical warfarin dose-initiation model for Black-African patients in South Africa and Uganda

Innocent G. Asiimwe[1], Catriona Waitt[1,2], Christine Sekaggya-Wiltshire[2], Claire Hutchinson[1], Emmy Okello[3], Eunice J. Zhang[1], Jerome R. Semakula[2], Johannes P. Mouton[4], Karen Cohen[4], Marc Blockman[4], Mohammed Lamorde[2], Andrea L. Jorgensen[5,6], and Munir Pirmohamed[1,6]

[1]The Wolfson Centre for Personalized Medicine, MRC Centre for Drug Safety Science, Department of Pharmacology and Therapeutics, Institute of Systems, Molecular and Integrative Biology, University of Liverpool, United Kingdom. [2]Infectious Diseases Institute, Makerere University College of Health Sciences, Kampala, Uganda. [3]Uganda Heart Institute, Kampala, Uganda. [4]Division of Clinical Pharmacology, Department of Medicine, University of Cape Town, Cape Town, South Africa. [5]Department of Biostatistics, Institute of Population Health Sciences, University of Liverpool, United Kingdom. [6]These authors contributed equally.

Correspondence: Innocent G. Asiimwe (i.asiimwe@liverpool.ac.uk, +441517955387); Munir Pirmohamed (munirp@liverpool.ac.uk, +44 151 794 5549).

CONFLICT OF INTEREST

MP receives research funding from various organisations including governmental bodies such as the MRC, EU Commission and Health Education England. MP also received partnership funding for the following: MRC Clinical Pharmacology Training Scheme (co-funded by MRC and Roche, UCB, Eli Lilly and Novartis); a PhD studentship jointly funded by EPSRC and Astra Zeneca; and

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/CPT.2128

grant funding from Vistagen Therapeutics. MP also received unrestricted educational grant support for the UK Pharmacogenetics and Stratified Medicine Network from Bristol-Myers Squibb and UCB. He has developed an HLA genotyping panel with MC Diagnostics, but does not benefit financially from this. None of the funding MP received is related to the current paper. All other authors declared no competing interests for this work.

FUNDING

This research was commissioned by the National Institute for Health Research (NIHR) Global Health Research Group on Warfarin anticoagulation in patients with cardiovascular disease in sub-Saharan Africa [ref: 16/137/101] using UK aid from the UK Government. The views expressed in this publication are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

KEY WORDS: model; dosing; development; personalized medicine; warfarin.

ABSTRACT

Warfarin remains the oral anticoagulant of choice in sub-Saharan Africa. However, dosing is challenging due to a highly variable clinical response for a given dose. This study aimed to develop and validate a clinical warfarin dose-initiation model in sub-Saharan Black-African patients. For the development cohort, we used data from 364 patients who were recruited from 8 outpatient clinics and hospital departments in Uganda and South Africa (June 2018–July 2019). Validation was undertaken using the International Warfarin Pharmacogenetics Consortium (IWPC) dataset (690 Black patients). Four predictors (age, weight, target International Normalized Ratio range and HIV status) were included in the final model which achieved mean absolute errors (MAEs, mean of absolute differences between true dose and dose predicted by the model) of 11.6 (95% CI 10.4 to 12.8) and 12.5 (11.6 to 13.4) mg/week in the development and validation cohorts respectively. Two other clinical models, IWPC and Gage, respectively obtained MAEs of 12.5 (11.3 to 13.7) and 12.7 (11.5 to 13.8) mg/week in the development cohort, and 12.1 (11.2 to 13.0) and 12.2 (11.4 to 13.1) mg/week in the validation cohort. Compared to fixed dose-initiation, our model decreased the percentage of patients at high risk of suboptimal anticoagulation by 7.5% (1.5% to 13.7%) and 11.9% (7.1% to 16.8%) in the development and validation cohorts, respectively. The clinical utility of this model will be tested in a prospective study.

INTRODUCTION

Warfarin remains the most widely used oral anticoagulant worldwide, even after the introduction of new oral anticoagulants.[1] In sub-Saharan Africa and other low-income countries, it remains preferred due to its significantly lower cost.[2] However, warfarin dosing is challenging because of its narrow therapeutic window and large intra- and interpatient variability in clinical

response for a given dose. Moreover, poor anticoagulation can lead to thrombotic or bleeding events if the International Normalized Ratio (INR), a measure of blood coagulation ability, is below or above therapeutic range, respectively. It is therefore unsurprising that warfarin is an important cause of preventable adverse drug reaction (ADR)-related hospitalizations in South Africa. For example, in a study by Mouton et al, 164 (8%) of 1951 hospitalizations were ADR-related, and of these, 11 were due to warfarin, which was the most commonly implicated drug.[3]

To improve the accuracy of warfarin dosing, several dose-prediction models that include clinical, demographic and environmental factors such as age, height, weight and interacting drugs as well as genetic factors such as polymorphisms in the genes *CYP2C9* (cytochrome P450, family 2, subfamily C, polypeptide 9), and *VKORC1* (vitamin K epoxide reductase complex, subunit1) have been developed.[4] Clinical dosing models include only clinical, demographic and environmental factors whereas pharmacogenetic models additionally incorporate genetic factors. To date, most models have been developed in Whites and these may not be applicable to sub-Saharan African populations.

The need for dose-prediction models that are applicable to sub-Saharan African patients is emphasized by the poor quality of warfarin anticoagulation (defined as having a time in therapeutic INR range [TTR] of less than 65%[5]) in this region. In an earlier study in South Africa and Uganda, we observed that the median TTR was 41% (range 35% to 48%),[6] similar to previous reports in this region.[7-10] This is quite low when compared to our previous experience with an European cohort (mean TTRs of 60% [fixed-dose initiation] and 67% [genotype-guided dosing]).[11] To improve warfarin anticoagulation through optimizing dosing, we therefore aimed to develop and validate a clinical warfarin dose-initiation model for sub-Saharan African patients.

METHODS

This reporting follows the TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) statement[12] (checklist in Table S1). Although the TRIPOD statement focuses on prediction models for individual prognosis or diagnosis, most

recommendations and reporting items equally apply to the development and validation of dose prediction models.[12]

Source of data

The dataset used for developing the dose-initiation model, the so-called 'development cohort' comprised 364 warfarin-treated patients. They were recruited between June 2018 and July 2019 as part of an observational study being conducted in Uganda and South Africa (ClinicalTrials.gov Identifier: NCT03512080) by the National Institute of Health Research (NIHR) Global Health Research Group on WARfarin anticoagulation in PATients in Sub-SaHaran Africa (War-PATH; http://warpath.info/). This is a collaboration between the University of Liverpool, Infectious Diseases Institute (Makerere University, Uganda) and the University of Cape Town (South Africa). The study complies with the Declaration of Helsinki[13] and was approved by Institutional Review Boards (IRB) of the University of Liverpool (UK; ref: 2934), University of Cape Town (South Africa; ref: 672/2017), and Joint Clinical Research Centre (Uganda; ref: JC3017). Additionally, work in Uganda was approved by the Uganda National Council for Science and Technology (ref: HS164ES). IRB approval and written individual-patient informed consent were obtained before patient enrolment.

The dataset used for validating the dose-initiation model, the so-called 'external validation cohort', comprised data from the International Warfarin Pharmacogenetics Consortium (IWPC) study which collated individual patient data from participants in 9 countries (four continents).[14] The IWPC ethnicity dataset containing detailed de-identified curated demographic, clinical and genetic data from 6922 multiethnic chronic warfarin users was downloaded from the PharmacoGenomics KnowledgeBase (PharmGKB) website (https://www.pharmgkb.org/downloads, under the sub-heading "International Warfarin Pharmacogenetics Consortium (IWPC)").

Participants

The War-PATH study population consisted of warfarin-treated patients of self-reported Black-African ethnicity who were recruited from 8 outpatient clinics and hospital departments (Table S2). Inclusion criteria were consenting adult patients (≥18 years) treated with warfarin for (a)

venous thromboembolism (VTE) or atrial fibrillation (AF) with a target International Normalised Ratio (INR) range 2.0–3.0; or (b) valvular heart disease with a target INR range 2.5–3.5. Patients were only included if they had attained the outcome of interest (the stable warfarin dose, as defined below). As the purpose of the model was to predict what a patient's stable dose was going to be, patients having never achieved stable warfarin dose as defined below were excluded. Patients who were unwilling to take part, pregnant women or patients with any other contraindications based on clinician judgement were also excluded.

The study design and eligibility criteria for the IWPC study have previously been reported.[14] For the external validation cohort, we only included patients who were Black or African American and had a target INR of either 2.5 or 3.0 (where not available, target INR was estimated based on treatment indication).

Outcome

The outcome of interest was stable warfarin dose, defined for the development cohort as the same dose for two consecutive clinic visits in the 12 months preceding recruitment, with the INR being in therapeutic range at each of those visits. War-PATH study case report forms were used to capture weekly stable dose as well as other variables described below. The IWPC study sites used different definitions for warfarin stable dose, most requiring a dose that produced stable anticoagulation levels (measured using INR) over a defined time period.[14]

Predictors

As part of the observational study (details above), the following data were captured for each War-PATH patient during enrolment: country of recruitment, age, weight, height, gender, employment status, annual individual/household income, education status, housing type, distance between residence and health centre, time taken to travel from home to health centre, indication for anticoagulation (used to infer the target INR range), smoking status, alcohol intake per week, renal biomarker data (serum creatinine and estimated glomerular filtration rate), liver biomarker data (aspartate aminotransferase, alanine aminotransferase, gamma-glutamyl transferase, total bilirubin and albumin), HIV status, TB status, other comorbidities and use of

concomitant medications including herbal/complementary, traditional or "over-the counter" medications.

Sample size

During model development, all available data (*n* = 364) was used to maximise the power and generalisability of the results. To avoid the risks of over- and under-fitting, the study participant-per-candidate predictor parameter was set at 20.[15] Consequently, we could only consider a maximum of 18 candidate predictor parameters. For model validation, again all eligible patients were included.

Missing data

Missing data was defined as the absence of data values within a specific variable category in the case report form. Variables coded as 'unknown' (e.g. for HIV status) were therefore not considered as missing. Again, to maximize power and generalizability of the results, we did not exclude cases with missing data for any of the included predictor variables. Rather we assumed that missing data occurred at random and performed Multivariate Imputation by Chained Equations (MICE) within R.[16] The included predictor variables and outcome variable[17] were used to impute missing data using predictive mean matching for continuous variables and logistic regression for binary/categorical variables. Following the recommendation to have the imputed datasets to at least equal the percentage of incomplete cases,[18] we created multiple datasets which were identical except for the imputed values to reflect the uncertainty associated with each of the imputations. The estimates derived from these datasets were combined using Rubin's rules.[19]

Statistical analysis methods

Multivariable linear regression models were fitted using four regression approaches (ordinary least squares, quantile regression, weighted quantile regression and non-linear least squares) on the development cohort. Starting with a list of potential variables, all possible linear models were fitted, with the optimal models chosen with reference to mean absolute error (MAE, defined as

the mean of the absolute values for the difference between predicted and stable doses), mean absolute percentage error (MAPE) and logarithmic ratio of estimate to actual value. Bootstrap validation was applied to correct overfitting and the final models were externally validated in the IWPC cohort. In both development and external validation cohorts, we compared our models with current warfarin initiation practice in sub-Saharan Africa (fixed dose of 35 mg/week) and two widely known dose prediction models (the IWPC[14] and Gage[20] models). Other performance indices were also employed for this purpose, including the percentage of patients with ideal dose (defined as predicted dose within 20% of actual dose) and the coefficient of determination (R^2) to respectively enable us to assess the clinical relevance and fit accuracy of the model. Bias was assessed using a measure derived from the mean of the logarithm of the accuracy ratio while clinical safety was computed as the percentage of patients at risk of under- or over-dosing (defined as having an actual dose at least 40% lower or higher than the predicted dose respectively – see Text S3 for justification in choosing 40%). Where we computed the difference between the mean performance of our model and that of a comparator model, statistical significance was considered as 95% confidence intervals that did not contain zero. Texts S1-S3 provide a detailed explanation of the analysis methods and metrics employed. All analyses were conducted in R version 3.6.1[21] (R code used is available in Text S4).

Sensitivity analysis

Because our model and the Gage model required age as a continuous variable and yet the IWPC dataset had categorized it (age recorded in decades), we made an assumption that the age corresponded to the lower-class boundary (e.g. for the decade of 20 to 29 years, the age was considered to be 20). To test the accuracy of this assumption, we conducted sensitivity analyses in which the age was taken as the midpoint value (e.g. 24.5 years) or upper-class boundary (e.g. 29 years). For the first decade (10 to 19 years), the lower-class boundary and midpoint values were considered to be 18 and 18.5 years respectively since only adults were included in the IWPC dataset.

RESULTS

Participants

We included the 364 patients that had been recruited by 31st July 2019, and characteristics of all patients are presented in Table 1. The flow diagram for the IWPC patients included in the external validation cohort is shown in Figure S1.

Model development and specification

Figure 1 summarizes the predictor selection process. Based on expert guidance and literature review, we considered three predictors (age, weight, and target INR range) to have higher clinical relevance and therefore these were not subjected to selection during the modeling stage. Four other predictors including country of recruitment, gender, HIV status and simvastatin/amiodarone status were also selected for use during the modeling process. HIV-status had three factor levels, and this translated into a study participant-per-candidate predictor parameter of 45.6. Amongst these seven predictor variables, only 3.0% (*n* = 11) cases were missing data with all cases missing weight (Figure S2–S3).

The performances of the best 3-, 4-, 5-, 6- and 7-variable models for each of the four regression approaches are shown in Figure S4. For ordinary least squares, quantile regression and non-linear least squares, the 4-variable models (predictors: age, weight, target INR, and HIV status) were preferable in terms of the mean absolute error (MAE) or the mean absolute logarithm of the accuracy ratio (MALAR) and model parsimony. For weighted quantile regression, the 3-variable models (predictors: age, weight, and target INR) were chosen based on the mean absolute percentage error (MAPE). Since each regression approach fitted non-transformed, square-root and logarithmic doses, 12 models in total were selected. The performances of these models in terms of the MAE, MAPE and the logarithm of the accuracy ratio-derived measures are shown in Table S3. We chose the 4-variable model (original coefficients in Table S4) fitted using non-linear least squares and non-transformed dose (details in Table S3). Applying a shrinkage factor of 0.8892 to the coefficients of this model produced a shrunken model (*Weekly dose in mg = 20.2832 – 0.0656 x Age [years] + 0.2178 x Weight [kg] + 7.3190 [if Target INR range is 2.5 to 3.5] + 8.7973 [if HIV positive] + 3.4054 [if HIV unknown]*, Table 2), henceforth called the War-PATH model.

Prediction accuracy

The performances of the War-PATH, IWPC, Gage and fixed-dose (35 mg/wk) models in the development and validation cohorts are shown in Table 3 (the performance of the War-PATH model using its original coefficients is shown in Table S5). The War-PATH model, compared to fixed dose-initiation, provided dose estimates that were closer to the actual doses (as shown by a lower MAE) in both the development (11.6 [95% CI 10.4; 12.8] vs 12.3 [10.9; 13.7] mg/week) and external validation (12.5 [11.6; 13.4] vs 13.8 [12.7; 14.9] mg/week) cohorts. However, these reductions (respectively 0.7 [-1.1; 2.5] and 1.3 [-0.1; 2.8] mg/wk) were not statistically significant. A similar trend was observed with the 'unbiased' MAPE: War-PATH vs fixed dose-initiation predicted doses were on average within 34.0% (31.0%; 37.2%) vs 36.5% (32.7%; 40.5%) of the actual doses in the development cohort and within 35.5% (33.0%; 38.1%) vs 40.1% (37.1%; 43.1%) of the actual doses in the validation cohort. Only the difference in the validation cohort (a MAPE reduction of 4.6% [0.7%; 8.5%]) was statistically significant. In terms of biasness of predictions, the War-PATH model (bias of 0.3% [-3.5%; 4.2%] above the actual dose) was unbiased (95% CIs contained zero) while fixed dose-initiation systematically underpredicted (5.8% [1.8%; 9.6%] below the actual dose) in the development cohort. In the validation cohort, our model systematically overpredicted (4.3% [1.4%; 7.3%] above the actual dose) while fixed dose-initiation underpredicted (9.5% [6.6%; 12.3%] below the actual dose). The bias of our model in the validation cohort was, however, eliminated when the age was taken as the upper-class boundary (predicted doses 2.8% [-0.1%; 5.7%] above the actual) (Table S6).

Regarding the MAE and unbiased MAPE, our model performed slightly better than the IWPC and Gage models in the development cohort and similarly in the external validation cohort as summarized in Table 3. The IWPC and Gage models systematically over-predicted in the development cohort (predicted doses were, respectively, on average 7.5% [3.2%; 12.0%] and 13.3% [8.9%; 17.9%] above the actual doses). In the IWPC validation cohort; however, only the Gage model retained this bias (predicted doses 9.7% [6.7%; 12.9%] above the actual) and this bias was eliminated when age was taken as the upper-class boundary (predicted doses 2.6% [-0.2%; 5.5%] above the actual) (Table S6). Our model's fit accuracy (R^2) was comparable to these two models in the development cohort (War-PATH vs IWPC vs Gage: 15.1% [11.6; 18.6] vs 12.8

[7.2; 18.5] vs 15.6 [9.0; 22.3]) but the worst in the validation cohort (War-PATH vs IWPC vs Gage: 12.2 [9.9; 14.5] vs 23.5 [19.1; 27.9] vs 24.2 [18.9; 29.5]).

Clinical relevance and safety

Our model (41.3% [95% CI 36.2%; 46.3%] performed slightly better than the IWPC (37.5% [32.4%; 42.6%]) and Gage (37.3% [32.4%; 42.4%]) models in terms of the percentage of patients with ideal dose in the development cohort (performance was similar in the validation cohort, Table 3). Consequently, it was the preferred clinical model to consider for implementation in South Africa and Uganda.

Compared to existing clinical practice (fixed dose-initiation with 35 mg/wk), our model performed slightly worse in the development cohort (2.7% [-4.4%; 9.8%] less patients with ideal dose) but slightly better in the validation cohort (2.0% [-3.3%; 7.3%] more patients with ideal dose). The differences in the performance of the War-PATH model and fixed dose-initiation in the low (≤21 mg/wk), intermediate (>21 and <49 mg/wk), and high (≥49 mg/wk) dose groups are shown in Figure 2 (percentage of patients with ideal dose) and Table S7 (percentage of patients with ideal, underestimated or overestimated doses). Neither of these approaches were able to predict ideal dose in the low dose groups, in either of the cohorts. In the development cohort, our model was better than fixed dose-initiation in the high dose group (19.7% [11.5%; 27.9%] more patients with ideal dose) but this came at the expense of the performance in the intermediate dose group (12.3% [3.6%; 20.7%] less patients with ideal dose). A similar trend was observed in the validation cohort (respective differences of 29.5% [23.3%; 35.8%] and 10.7% [4.2%; 17.3%]).

With the IWPC dose thresholds, 32 and 97 patients in the development cohort respectively required ≤21 mg/wk and ≥49 mg/wk (Figure 2). This translates into 8.8% and 26.6% of the patients in the development cohort being at high risks of over- and under-anticoagulation respectively (or a total of 35.4% of the patients being at risk of sub-optimal anticoagulation). On the other hand, only 27.9% of the development cohort was computed as being at risk of sub-optimal anticoagulation with our model. This implies that using our model would reduce the percentage of patients at risk of sub-optimal anticoagulation by 7.5% (1.5%; 13.7%) (this figure is

the same as the increase in the number of patients in the 'low risk' dose group, Figure 3). In the validation cohort, fixed dosing would place 37.6% of the patients at a risk of sub-optimal anticoagulation and as for the development cohort, our model would decrease this figure by 11.9% (7.1%; 16.8%).

DISCUSSION

Warfarin dosing remains challenging due to a narrow therapeutic window and large intra- and interpatient variability in dose requirements due to clinical and genetic factors. Initial dosing (e.g. 5 mg/day) is often empirical with dose adjustments made until the patient is within therapeutic range. However, during this time, patients are at an increased risk of bleeding or thromboembolic events if too much or too little warfarin is prescribed, respectively.[4] To facilitate warfarin dose-initiation in two of the underrepresented sub-Saharan African populations, we have developed a dose-initiation model that includes four clinical factors namely age, weight, target INR range and HIV status. Three of these (age, weight and target INR) have established relationships with warfarin dose and have been included in many other models[1,22-24] and so were not subjected to selection during the modeling process. Of the variables subjected to selection, HIV status was the most important predictor as evidenced by its inclusion in the final model. Of note, HIV status was represented in the model as a categorical variable with three levels – negative, positive and HIV status unknown. Including the third category allows for the possibility of using the model even where a patient refuses an HIV test. Given the HIV infection prevalence, albeit highly variable,[25] in African countries (10% in Ugandans and 22% in South Africans in this study), it was important to include this in our model. HIV infection per se can lead to a hypercoagulable state[26-30] resulting in increased dose requirements. However, doses may also change because of interactions with antiretroviral drugs.

Compared to current practice in these two countries (fixed dose-initiation), this clinical model decreased the percentage of patients at high risk of suboptimal anticoagulation by 7.5% and 11.9% in the development and validation cohorts, respectively. The greatest benefits were seen in the patients at high risk of under-anticoagulation – patients who are predisposed to thromboembolic events even when warfarin therapy has been initiated. For the patients at high

risk of over-anticoagulation, this clinical model performed similarly to fixed dose-initiation which implies it needs further development to improve prediction.

Genetic variants in the *CYP2C9* and *VKORC1* genes account for about 40% of the variance in daily warfarin dose requirement.[31] In a systematic review, we have recently quantified the effect of genetic variants in the *CYP2C9* and *VKORC1* genes that are more prevalent in Black-African populations,[32] and it would be expected that inclusion of these variants to the model should improve prediction for all patients. In order to test this, we will be developing a pharmacogenetic model during the ongoing collaborative project in Uganda and South Africa (WARfarin anticoagulation in PATients in Sub-SaHaran Africa; http://warpath.info/).

Based on the mean absolute error (MAE), unbiased mean absolute percentage error, and percentage of patients with ideal dose, our model performed better than two of the most popular clinical models (the Gage and IWPC models) in the development cohort and comparably in the external validation cohort. Although the IWPC model was developed in a subgroup of patients with a target INR range of 2.0–3.0, we also tested it in patients with a target INR range of 2.5–3.5 since testing a prediction model in a completely different clinical setting (domain validation) is an accepted form of external validation.[33] We did not compare the performance of our model with any of the existing pharmacogenetic models but based on previous studies, current pharmacogenetic models may not work better than this clinical model in this patient group. For example, a race-based evaluation of the IWPC dataset using the IWPC, Gage and eleven other pharmacogenetic models revealed that the MAE was 12.0–21.0 mg/wk in Blacks[34] (estimates of MAE in this study were 12.5 mg/wk in the external validation cohort). Liu et al.[35] made similar observations when they reanalyzed the IWPC dataset using nine different machine learning techniques (MAEs for the pharmacogenetic models were 12.2–13.8 mg/wk for Blacks). In terms of the coefficient of determination (R^2), our model (15%) performed similarly to the IWPC (13%) and Gage (17%) models in the development cohort, but worse in the validation cohort (12% vs 24% and 24%, respectively). Importantly, the R^2 which relates to the proportion of the variance explained by a model was not our primary metric. This is because the value of a model is best depicted by its prediction accuracy, not by its fit accuracy.[36-38] For instance, even

fixed dose-initiation with an R^2 that is close to zero is considered useful when the MAE is considered.

There are limitations to our study. We excluded unstable patients from our study which may limit the generalizability of the model. However, patients with unstable dose do not have an outcome variable value and therefore could not be included in the prediction model we developed. Second, our dose-initiation model cannot be used during the dose-revision period. For this, national treatment guidelines[39] or published guidelines[40,41] are available. Third, we excluded other relevant well-known variables such as body mass index, adherence and vitamin K status that have previously been found to be important in influencing stable dose mostly because they were unavailable, expensive to capture or missing in a large proportion of patients – which would have affected implementation. Fourth, we did not study children, where to date, no dosing model with clinical utility has been developed in any ethnic group. Lastly, our analysis did not address whether precise initial dosing would result in reduction in the time to attain and/or time in stable therapeutic INR.

We will be implementing this model in a prospective cohort of Ugandan and South African patients where a clinical decision has been made to start warfarin. We anticipate that implementing a consistent dose-initiation model will improve time in therapeutic range (TTR) and other clinical outcomes as demonstrated by studies in Western nations.[42] It is important to note that even modest TTR improvements can have significant benefits on clinical outcomes and cost as demonstrated by Rose et al.[43] who showed that in a population of 67,077 atrial fibrillation patients, a 5% improvement will prevent nearly 200 ischemic strokes, over 600 deaths and close to $16 million per year.

To our knowledge, we have developed the first warfarin dose-initiation clinical model for sub-Saharan African patients. It performs better than fixed dosing in terms of more accurate dosing and puts less patients at risk of suboptimal anticoagulation. We hope that its implementation and validation in a prospective cohort will inform future large-scale implementation. Our long-term aim is to also evaluate the importance of genetic variants in improving warfarin dosing and anticoagulation in patients in Uganda and South Africa. Given that genetic testing is not readily available in African countries because of cost and lack of facilities, it could be argued that a

clinical dosing model is more important in the near-term. However, it is also important to ensure that genomic medicine does not bypass developing countries as this will exacerbate health inequalities.

STUDY HIGHLIGHTS

What is the current knowledge on the topic?

Warfarin dosing remains challenging due to a highly variable clinical response for a given dose.

What question did this study address?

Can a clinical dose-initiation model be developed and validated for sub-Saharan Black-African patients?

What does this study add to our knowledge?

We have developed the first warfarin dose-initiation clinical model for Black-African patients in Uganda and South Africa.

How might this change clinical pharmacology or translational science?

We will be implementing and validating this model in a prospective cohort to inform future large-scale implementation. More optimized dosing should improve the quality of warfarin anticoagulation in these two developing countries.

DATA AVAILABILITY

The IWPC dataset is available at the PharmGKB website (https://www.pharmgkb.org/downloads) whereas the War-PATH dataset is available from the corresponding author on reasonable request.

ACKNOWLEDGEMENTS

We thank staff of Uganda Heart Institute, Mulago National Referral Hospital, Uganda Heart Institute, St Francis Hospital Nsambya, Infectious Diseases Institute Adult HIV clinic, Groote Schuur Hospital, Gugulethu Community Health Centre, Khayelitsha Site B Community Health Centre, and Mfuleni Community Health Centre for their support during data collection. I.G.A. thanks the University of Liverpool for studentship funding support. I.G.A also thanks Ronald Bisaso Kuteesa (Department of Pharmacology and Therapeutics, Makerere University College of Health Sciences) for the insightful discussions on the SWEEPER operator.

AUTHOR CONTRIBUTIONS

I.G.A, C.W, C.S.W, C.H, E.J.Z, J.R.S, J.P.M, K.C, M.B, M.L, A.L.J, and MP wrote the manuscript. I.G.A, C.W, C.S.W, C.H, E.J.Z, J.R.S, J.P.M, K.C, M.B, M.L, A.L.J, and MP designed the research. C.W, C.S.W, C.H, E.O, J.R.S, J.P.M, K.C, M.B, M.L, A.L.J, and MP performed the research. I.G.A. and A.L.J. analyzed the data. C.W, C.S.W, C.H, E.O, J.R.S, J.P.M, K.C, M.B and M.L contributed new reagents/analytical tools.

REFERENCES

1. Shendre A, Dillon C, Limdi NA. Pharmacogenetics of warfarin dosing in patients of African and European ancestry. *Pharmacogenomics.* 2018;19(17):1357-1371.
2. Laas DJ, Naidoo M. Oral anticoagulants and atrial fibrillation: A South African perspective. *S Afr Med J.* 2018;108(8):640-646.
3. Mouton JP, Njuguna C, Kramer N, et al. Adverse Drug Reactions Causing Admission to Medical Wards: A Cross-Sectional Survey at 4 Hospitals in South Africa. *Medicine (Baltimore).* 2016;95(19):e3437.
4. Lee MT, Klein TE. Pharmacogenetics of warfarin: challenges and opportunities. *J Hum Genet.* 2013;58(6):334-338.
5. The National Institute for Health and Care Excellence. Atrial fibrillation: management (Clinical guideline CG180). In. Online2014.
6. Semakula JR, Mouton JP, Jorgensen A, et al. A cross-sectional evaluation of five warfarin anticoagulation services in Uganda and South Africa. *PLoS ONE.* 2020; 15(1):e0227458.
7. Ebrahim I, Bryer A, Cohen K, Mouton JP, Msemburi W, Blockman M. Poor anticoagulation control in patients taking warfarin at a tertiary and district-level prothrombin clinic in Cape Town, South Africa. *S Afr Med J.* 2018;108(6):490-494.
8. Sekaggya C, Nalwanga D, Von Braun A, et al. Challenges in achieving a target international normalized ratio for deep vein thrombosis among HIV-infected patients with tuberculosis: a case series. *BMC Hematol.* 2016;16:16.
9. Sonuga BO, Hellenberg DA, Cupido CS, Jaeger C. Profile and anticoagulation outcomes of patients on warfarin therapy in an urban hospital in Cape Town, South Africa. *Afr J Prim Health Care Fam Med.* 2016;8(1):e1-8.
10. Mouton J, Blockman M, Sekaggya-Wiltshire C, et al. Improving anticoagulation in sub-Saharan Africa -- what are the challenges, and how can we overcome them? *Authorea.* 2020.
11. Pirmohamed M, Burnside G, Eriksson N, et al. A randomized trial of genotype-guided dosing of warfarin. *N Engl J Med.* 2013;369(24):2294-2303.
12. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med.* 2015;162(1):W1-73.
13. World Medical A. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA.* 2013;310(20):2191-2194.

14. International Warfarin Pharmacogenetics Consortium, Klein TE, Altman RB, et al. Estimation of the warfarin dose with clinical and pharmacogenetic data. *N Engl J Med.* 2009;360(8):753-764.
15. Moons KGM, Wolff RF, Riley RD, et al. PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. *Ann Intern Med.* 2019;170(1):W1-W33.
16. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. . *Journal of Statistical Software.* 2011;45(3):1-67.
17. Moons KG, Donders RA, Stijnen T, Harrell FE, Jr. Using the outcome for imputation of missing predictor values was preferred. *J Clin Epidemiol.* 2006;59(10):1092-1101.
18. White IR, Royston P, Wood AM. Multiple imputation using chained equations: Issues and guidance for practice. *Stat Med.* 2011;30(4):377-399.
19. Rubin DB. *Multiple Imputation for Nonresponse in Surveys.* New York: J. Wiley & Sons; 1987.
20. Gage BF, Eby C, Johnson JA, et al. Use of pharmacogenetic and clinical factors to predict the therapeutic dose of warfarin. *Clin Pharmacol Ther.* 2008;84(3):326-331.
21. *R: A language and environment for statistical computing.* [computer program]. Vienna: R Foundation for Statistical Computing; 2019.
22. Pirmohamed M, Kamali F, Daly AK, Wadelius M. Oral anticoagulation: a critique of recent advances and controversies. *Trends Pharmacol Sci.* 2015;36(3):153-163.
23. Verhoef TI, Redekop WK, Daly AK, van Schie RM, de Boer A, Maitland-van der Zee AH. Pharmacogenetic-guided dosing of coumarin anticoagulants: algorithms for warfarin, acenocoumarol and phenprocoumon. *Br J Clin Pharmacol.* 2014;77(4):626-641.
24. Asiimwe IG, Zhang EJ, Osanlou R, Jorgensen AL, Pirmohamed M. Warfarin dosing algorithms: a systematic review. *Br J Clin Pharmacol.* 2020.
25. Dwyer-Lindgren L, Cork MA, Sligar A, et al. Mapping HIV prevalence in sub-Saharan Africa between 2000 and 2017. *Nature.* 2019;570(7760):189-193.
26. Baker JV, Brummel-Ziedins K, Neuhaus J, et al. HIV replication alters the composition of extrinsic pathway coagulation factors and increases thrombin generation. *J Am Heart Assoc.* 2013;2(4):e000264.
27. Bibas M, Biava G, Antinori A. HIV-Associated Venous Thromboembolism. *Mediterr J Hematol Infect Dis.* 2011;3(1):e2011030.
28. Anderson AM, Chane T, Patel M, Chen S, Xue W, Easley KA. Warfarin therapy in the HIV medical home model: low rates of therapeutic anticoagulation despite adherence and differences in dosing based on specific antiretrovirals. *AIDS Patient Care STDS.* 2012;26(8):454-462.

29. Durand M, Sinyavskaya L, Jin YL, Tremblay CL, Ducruet T, Laskine M. Incidence of Venous Thromboembolism in Patients Living with HIV: A Cohort Study. *AIDS Patient Care STDS.* 2019;33(11):455-458.
30. Vululi ST, Bugeza S, Zeridah M, et al. Prevalence of lower limb deep venous thrombosis among adult HIV positive patients attending an outpatient clinic at Mulago Hospital. *AIDS Res Ther.* 2018;15(1):3.
31. Bourgeois S, Jorgensen A, Zhang EJ, et al. A multi-factorial analysis of response to warfarin in a UK prospective cohort. *Genome Med.* 2016;8(1):2.
32. Asiimwe IG, Zhang EJ, Osanlou R, et al. Genetic factors influencing warfarin dose in Black-African patients: a systematic review and meta-analysis. *Clin Pharmacol Ther.* 2019.
33. Hendriksen JM, Geersing GJ, Moons KG, de Groot JA. Diagnostic and prognostic prediction models. *J Thromb Haemost.* 2013;11 Suppl 1:129-141.
34. Shin J, Cao D. Comparison of warfarin pharmacogenetic dosing algorithms in a racially diverse large cohort. *Pharmacogenomics.* 2011;12(1):125-134.
35. Liu R, Li X, Zhang W, Zhou HH. Comparison of Nine Statistical Model Based Warfarin Pharmacogenetic Dosing Algorithms Using the Racially Diverse International Warfarin Pharmacogenetic Consortium Cohort Database. *PLoS One.* 2015;10(8):e0135784.
36. Alexander DL, Tropsha A, Winkler DA. Beware of R(2): Simple, Unambiguous Assessment of the Prediction Accuracy of QSAR and QSPR Models. *J Chem Inf Model.* 2015;55(7):1316-1322.
37. Lo B, Gao X. Assessing software cost estimation models: Criteria for accuracy, consistency and regression. . *Australasian Journal of Information Systems.* 1997;5(1):30-44.
38. Sheiner LB, Beal SL. Some suggestions for measuring predictive performance. *J Pharmacokinet Biopharm.* 1981;9(4):503-512.
39. Republic of South Africa. Essential Drugs Programme. Hospital level (Adults) Standard Treatment Guidelines and Essential Medicines List. . In: Health NDo, ed. 4th ed. ed. Pretoria: The National Department of Health; 2015.
40. Kim YK, Nieuwlaat R, Connolly SJ, et al. Effect of a simple two-step warfarin dosing algorithm on anticoagulant control as measured by time in therapeutic range: a pilot study. *J Thromb Haemost.* 2010;8(1):101-106.
41. Jacobson BF, Schapkaitz E. Maintenance of Warfarin Therapy at an Anticoagulation Clinic. *South African Medical Journal* 2007;97(12):1259.
42. Van Spall HG, Wallentin L, Yusuf S, et al. Variation in warfarin dose adjustment practice is responsible for differences in the quality of anticoagulation control between centers and

countries: an analysis of patients receiving warfarin in the randomized evaluation of long-term anticoagulation therapy (RE-LY) trial. *Circulation.* 2012;126(19):2309-2316.

43. Rose AJ, Berlowitz DR, Ash AS, Ozonoff A, Hylek EM, Goldhaber-Fiebert JD. The business case for quality improvement: oral anticoagulation for atrial fibrillation. *Circ Cardiovasc Qual Outcomes.* 2011;4(4):416-424.

FIGURE LEGENDS

Figure 1. Selection of predictors. [a]The exception was simvastatin (a known warfarin potentiating drug) being taken by 9.3% of the patients which was combined with amiodarone (another warfarin potentiating drug) to meet this 10% requirement.

Figure 2. Percentage of patients with ideal dose. Error bars represent 95% confidence intervals. The ideal dose was defined as the predicted dose within 20% of the actual dose while the fixed dose was 35 mg of warfarin per week.

Figure 3. Percentage of patients put at risk of either under- or over-anticoagulation by the different dosing strategies.

SUPPLEMENTARY FILES

1. Supplementary Methods, Figures and Tables

Table 1. Patient characteristics.

Variables		Development cohort (*n* = 364)	External validation cohort (*n* = 690)
Country of recruitment, *n* (%)	South Africa	193 (53.0%)	-
	Uganda	171 (47.0%)	-
Age (years)	Mean (SD)	46.2 (15.3)	-
	Median (IQR)	46.0 (34.0–56.7)	-
	Range	18.1–87.0	-
Age (years) in categories, *n* (%)	10-19	12 (3.3%)	3 (0.4%)
	20-29	46 (12.6%)	24 (3.5%)
	30-39	63 (17.3%)	55 (8.0%)
	40-49	96 (26.4%)	114 (16.5%)
	50-59	72 (19.8%)	172 (24.9%)
	60-69	51 (14.0%)	154 (22.3%)
	70-79	18 (4.9%)	119 (17.2%)
	80-89	6 (1.6%)	43 (6.2%)
	≥90	-	6 (0.9%)
Gender, *n* (%)	Female	266 (73.1%)	393 (57%)
	Male	98 (26.9%)	297 (43%)
Weight (kg)	Mean (SD)	75.0 (20.8)	91.2 (27.4)
	Median (IQR)	71.0 (59.4–85.0)	86.4 (72.7–104.1)
	Range	36.5–150.0	41.2–237.7
	Missing	11 (3.0%)	7 (1.0%)
Height (cm)	Mean (SD)	163 (9)	170 (10)
	Median (IQR)	163 (158–-169)	170 (163–178)
	Range	142–183	127–201
	Missing	163 (44.8%)	15 (2.2%)
INR Target range[a], *n* (%)	2.0-3.0	237 (65.1%)	622 (90.1%)
	2.5-3.5	127 (34.9%)	68 (9.9%)
Venous Thromboembolism, *n* (%)	Yes	105 (28.8%)	189 (27.4%)
	No	259 (71.2%)	474 (68.7%)
	Missing	-	27 (3.9%)
HIV status, *n* (%)	Negative	282 (77.5%)	-
	Positive	59 (16.2%)	-
	Unknown	23 (6.3%)	690 (100.0%)
Amiodarone, *n* (%)	Yes	2 (0.5%)	36 (5.2%)
	No	362 (99.5%)	625 (90.6%)
	Missing	-	29 (4.2%)
Simvastatin, *n* (%)	Yes	34 (9.3%)	120 (17.4%)

	No	330 (90.7%)	569 (82.5%)
	Missing	-	1 (0.1%)
Enzyme Inducers[b], *n* (%)	Yes	5 (1.4%)	15 (2.2%)
	No	359 (98.6%)	516 (74.8%)
	Missing	-	159 (23%)
Current smoker, *n* (%)	Yes	17 (4.7%)	127 (18.4%)
	No	345 (94.8%)	539 (78.1%)
	Missing	2 (0.5%)	24 (3.5%)
Stable warfarin dose (mg/wk)	Mean (SD)	40.5 (17.7)	42.3 (18.7)
	Median (IQR)	35.0 (30.0–52.5)	38.0 (30.0–52.5)
	Range	8.8–137.5	10.0–126.0

[a]Those with heart valve disorders have a higher target range (2.5–3.5) than the rest (2.0–3.0) who include those with atrial fibrillation and venous thromboembolism. [b]Includes those taking carbamazepine, phenytoin, rifampicin, or rifampin. IQR = interquartile range; SD = standard deviation.

Table 2. War-PATH clinical dosing model

Sign	Estimate[a]	SD[b]	Covariates[c]
	20.2832	0.7918	Intercept
-	0.0656	0.0761	Age
+	0.2178	0.0708	Weight
+	7.3190	1.6369	Target INR range
+	8.7973	2.5944	HIV Positive
+	3.4054	3.4353	HIV Unknown
=	**Weekly warfarin dose[d]**		

[a]A shrinkage factor of 0.8892 (SD 0.0578) was applied to the coefficients of Age, Weight, Target INR range, HIV Positive and HIV Unknown, and the intercept re-estimated. [b]SDs were computed using bootstrapping (at least 1000 replicates). Except for the intercept (re-estimation SD used), SDs incorporate the SD of the shrinkage factor. [c]For Age, input age in years; Weight, input weight in kg; Target INR range, input 0 if target INR is 2.0–3.0 and 1 if target INR is 2.5–3.5; HIV Positive, input 0 if HIV negative or HIV unknown and 1 if HIV positive; and for HIV Unknown, input 0 if HIV negative and HIV positive and 1 if HIV unknown. [d]Predicted doses are rounded off to the nearest 2.5 mg/week. HIV = human immunodeficiency virus; INR = international normalized ratio; SD = standard deviation. Based on 10 imputed datasets.

Table 3. Performance of the clinical models and fixed dose-initiation[a]

Model	Development cohort (N = 364)[b]					External validation cohort (N = 690)				
	MAE (95% CI), mg/wk	Unbiased MAPE[c] (95% CI), %	Bias[d], (95% CI), %	% of patients with ideal[e] dose (95% CI)	R^2 (95% CI), %	MAE (95% CI), mg/wk	Unbiased MAPE[c] (95% CI), %	Bias[d], (95% CI), %	% of patients with ideal[e] dose (95% CI)	R^2 (95% CI), %
War-PATH	11.6 (10.4; 12.8)	34.0 (31.0; 37.2)	0.3 (-3.5; 4.2)	41.3 (36.2; 46.3)	15.1 (11.6; 18.6)	12.5 (11.6; 13.4)	35.5 (33.0; 38.1)	4.3 (1.4; 7.3)	42.0 (38.3; 45.8)	12.2 (9.9; 14.5)
IWPC	12.5 (11.3; 13.7)	37.2 (33.6; 40.7)	7.5 (3.2; 12.0)	37.5 (32.4; 42.6)	12.8 (7.2; 18.5)	12.1 (11.2; 13.0)	34.4 (31.8; 37)	2.5 (-0.4; 5.4)	45.2 (41.3; 49.1)	23.5 (19.1; 27.9)
Gage	12.7 (11.5; 13.8)	37.6 (34.1; 41.1)	13.3 (8.9; 17.9)	37.3 (32.4; 42.4)	15.6 (9.0; 22.3)	12.2 (11.4; 13.1)	34.8 (32.3; 37.4)	9.7 (6.7; 12.9)	42.3 (38.6; 46.0)	24.2 (18.9; 29.5)
Fixed dose[f]	12.3 (10.9; 13.7)	36.5 (32.7; 40.5)	-5.8 (-9.6; -1.8)	44.0 (38.8; 49.1)	0.0 (0.0; 0.0)	13.8 (12.7; 14.9)	40.1 (37.1; 43.1)	-9.5 (-12.3; -6.6)	40.0 (36.2; 43.8)	0.0 (0.0; 0.0)

[a]When referring to development and validation cohorts, this is in relation to the War-PATH model e.g. since part of the external validation cohort is IWPC's and Gage's development cohorts. [b]Computed doses rounded to the nearest 2.5mg (except for R^2). [c]Unbiased MAPE = (exp(mean(absolute(log(predicted dose/actual dose)))) − 1) × 100. [d]Bias = (exp(mean(log(predicted dose/actual dose))) − 1) × 100 (negative and positive values imply under- and over-estimation respectively). [e]The ideal dose was defined as the predicted dose within 20% of the actual dose. [f]The fixed dose was 35mg of warfarin per week. CI = confidence intervals; IWPC = International Warfarin Pharmacogenetics Consortium; MAE = mean absolute error; MAPE = mean absolute percentage error; R^2 = coefficient of determination. Based on 100 imputed datasets.