Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: The International Warfarin Pharmacogenetics Consortium. Estimation of the warfarin dose with clinical and pharmacogenetic data. N Engl J Med 2009;360:753-64.

Table of Contents of Supplementary Appendix

Section S1. Variables Included in the Regression Analysis and Warfarin Dosing Algorithms

S1a. Demographic Data

Demographic Data	
Gender	Male, Female or Not Known
Race	Self-reported information and Racial categories used as defined by the U.S. Office of Management and Budget
Ethnicity	Self-reported information and Racial categories used as defined by the Office of Management and Budget
Age	Binned age reported in years (0 - 9, 10 - 19, 20 - 29, 30 - 39, 40 - 49, 50 - 59, 60 - 69, 70 - 79, 80 - 89, 90+)

S1b. Background Data

Background Data	
Height	Reported in centimeters
Weight	Reported in kilograms
Indication for Warfarin Treatment	DVT, PE, Afib/flutter, Heart Valve, Cardiomyopathy/LV Dilation, Stroke, Post-Orthopedic, Other or Not Known
Comorbidities	List of diseases co-occurring in the patient
Diabetes	Yes, Not Present, Not Known
Congestive Heart Failure and/or Cardiomopathy	Yes, Not Present, Not Known
Valve Replacement	Yes, Not Present, Not Known
Medications	List of medications taken
Aspirin	Yes, Not Present, Not Known
Acetaminophen	Yes, Not Present, Not Known
Was Dose of Acetaminophen >1300mg/day	Yes or No
Simvastatin	Yes, Not Present, Not Known
Atorvastatin	Yes, Not Present, Not Known
Fluvastatin	Yes, Not Present, Not Known
Pravastatin	Yes, Not Present, Not Known
Rosuvastatin	Yes, Not Present, Not Known

Cerivastatin	Yes, Not Present, Not Known
Amiodarone**	Yes, Not Present, Not Known
Carbamazepine	Yes, Not Present, Not Known
Phenytoin	Yes, Not Present, Not Known
Rifampin	Yes, Not Present, Not Known
Sulfonamide Antibiotics including Septra, Bactrim, Cotrim and Sulfatrim	Yes, Not Present, Not Known
Macrolide Antibiotics including erythromycin, azithromycin, and clarithromycin	Yes, Not Present, Not Known
Herbal Medications, Vitamins and Supplements including garlic, ginseng, danshen, donquai, vitamins, zinc, iron, magnesium, etc.	Yes, Not Present, Not Known
Target INR	Target International Normalized Ratio or Not Known. Reported as single number or average of min, max in case of provided range.
Subject Reached Stable Dose of Warfarin	Yes, No, Not Known
INR on Reported Therapeutic Dose of Warfarin	International Normalized Ratio on the Therapeutic Dose of Warfarin Reported

**When amiodarone use was unknown, it was assumed that it was not used.

S1c. Phenotypic Data

Phenotypic Data	
Therapeutic Dose of Warfarin	Dose given in milligrams per week
Current Smoker	Yes, Not Present, Not Known

S1d. Genotype Data

Genotype Data	
CYP2C9 genotypes	*1, *2, *3, *4, *5, *6, *7, *8, *9, *10, *11, *12, *13 or Not Known
VKORC1 -1639 G>A (3673), rs9923231, G/A	A/A, A/G, G/G or Not Known
VKORC1 497T>G (5808), rs2884737, A/C	G/G, G/T, T/T or Not Known
VKORC1 1173 C>T (6484), rs9934438, A/G	C/C, C/T, T/T or Not Known
VKORC1 1542G>C (6853), rs8050894, C/G	C/C, C/G, G/G or Not Known
VKORC1 3730 G>A (9041), rs7294, A/G	A/A, A/G, G/G or Not Known
VKORC1 2255C>T (7566), rs2359612, A/G	C/C, C/T, T/T or Not Known
VKORC1 -4451 C>A (861), rs17880887, A/C	A/A, A/C, C/C or Not Known

S1e. Pharmacogenetic dosing algorithm

Warfarin pharmacogenetic dosing algorithm			
		5.6044	
-		0.2614 x	Age in decades
+		0.0087 x	Height in cm
+		0.0128 x	Weight in kg
-		0.8677 x	*VKORC1*^ A/G
-		1.6974 x	*VKORC1* A/A
-		0.4854 x	*VKORC1* genotype unknown
-		0.5211 x	*CYP2C9* *1/*2
-		0.9357 x	*CYP2C9* *1/*3
-		1.0616 x	*CYP2C9* *2/*2
-		1.9206 x	*CYP2C9* *2/*3
-		2.3312 x	*CYP2C9* *3/*3
-		0.2188 x	*CYP2C9* genotype unknown
-		0.1092 x	Asian race
-		0.2760 x	Black or African American
-		0.1032 x	Missing or Mixed race
+		1.1816 x	Enzyme inducer status
-		0.5503 x	Amiodarone status
=	**Square root of weekly warfarin dose****		

****The output of this algorithm must be squared to compute weekly dose in mg.**

^All references to *VKORC1* refer to genotype for rs9923231.

S1f. Clinical dosing algorithm

Warfarin clinical dosing algorithm		
	4.0376	
-	0.2546 x	Age in decades
+	0.0118 x	Height in cm
+	0.0134 x	Weight in kg
-	0.6752 x	Asian race
+	0.4060 x	Black or African American
+	0.0443 x	Missing or Mixed race
+	1.2799 x	Enzyme inducer status
-	0.5695 x	Amiodarone status
=	**Square root of weekly warfarin dose****	

****The output of this algorithm must be squared to compute weekly dose in mg. The output of this algorithm must be squared and then divided by 7 to compute the daily dose in mg.**

Legend for use of algorithms:

- Age in decades = 1 for 10-19, 2 for 20-29, etc…
- *VKORC1* G/A = 1 if heterozygous for rs9923231, otherwise zero
- *VKORC1* A/A = 1 if homozygous for A at rs9923231, otherwise zero
- *VKORC1* genotype unknown = 1 if rs9923231 genotype missing or unknown, otherwise zero
- *CYP2C9* *1/*2 = 1 if *CYP2C9* genotype is *1/*2, otherwise zero
- *CYP2C9* *1/*3 = 1 if *CYP2C9* genotype is *1/*3, otherwise zero
- *CYP2C9* *2/*2 = 1 if homozygous for *CYP2C9* *2 allele, otherwise zero
- *CYP2C9* *2/*3 = 1 if *CYP2C9* genotype is *2/*3, otherwise zero
- *CYP2C9* *3/*3 = 1 if homozygous for *CYP2C9* *3 allele, otherwise zero
- *CYP2C9* genotype unknown = 1 if *CYP2C9* genotype unknown, otherwise zero
- Asian Race = 1 if self-reported race is Asian, otherwise zero
- Black/African American = 1 if self-reported race is Black or African American, otherwise zero
- Missing or Mixed race = 1 if self-reported race is unspecified or mixed, otherwise zero
- Enzyme inducer status = 1 if patient taking carbamazepine, phenytoin, rifampin, or rifampicin, otherwise zero
- Amiodarone status = 1 if patient taking amiodarone, otherwise zero

Section S2. Detailed Definitions of Stable Warfarin Definitions for the Collected IWPC Data

Group	*Definition of Stable Dose of Warfarin for Each Research Group*
1	The dose (unchanged for 6 days) that yielded an INR within 0.5 of the target INR.
2	Average weekly dose (irrespective of achieved INR) that the patient received during the observation period, excluding the first 28 days after warfarin initiation.
3	The chronic (> 30 days) warfarin dose that led to an INR in the therapeutic target range on several occasions.
4	Warfarin therapeutic dose was defined as dose given when patients reach stable therapeutic INR. Stable therapeutic INR was defined as at least two consecutive INR measurements between 1.7-3 on the same warfarin daily or weekly dose measured at least one week apart.
5	Warfarin Therapeutic Dose was defined as a single mean weekly warfarin dose that was calculated by averaging the warfarin dose at each of the 3 consecutive clinic visits. A Stable Weekly Maintenance Dose of Warfarin was defined as a dose that did not vary by more than 10% between clinic visits. In addition, the INR at each of the 3 visits had to be in the patient's specific goal INR range.
6	The warfarin dose that led to an INR in the therapeutic range (2-3) on at least 3 consecutive clinic visits over a minimum period of 3 months.
7	Two consecutive INRs in the target range of 2-3.5 while on a constant dose where INR measures are taken at least 3 days but less than 8 days apart.
8	Stable dose in our data set is defined as the dose of 3 consecutive clinic visits, within therapeutic range of INR, the same daily dose over 3 months based on Higashi et al., JAMA 2002 (PMID 11926893).
9	INR between 2 and 3 for a period >1 month.
10	The definition of stable dose reported here is as follows: Dose at which INR was within therapeutic range (+/- 0.2 INR units) on 3 consecutive visits, with <90 days between subsequent visits.
11	The same warfarin dose for at least 3 consecutive clinic visits. No INR criteria to define stable dose was used because it was assumed that either the INR was in range at each visit or was not sufficiently out-of-range to elicit a change in dose.
12	A cross-sectional study of patients treated with warfarin for at least 2 months and with relatively stable anticoagulation.
13	Patients whose warfarin dose requirement was 1.5 mg per day or less and had a stable warfarin dose requirement for at least 3 consecutive clinic visits with a target International Normalised Ratio (INR) of 2.0 to 4.0 and no apparent cause for low dose requirement such as drug interactions or liver disease.
14	Dose that lead to stable INR over 3 visits.
15	Warfarin therapeutic dose was defined as warfarin dose at stable anticoagulation. Stable anticoagulation was defined when 2 consecutive INR measurements done at least 7 days apart were within the desired therapeutic range (i.e. 2-3), while warfarin dose was not changed.
16	Warfarin dose that lead to the target INR, usually 2-3 on 1 or more occasions over a minimum of 30 days.

17	The dose of warfarin that led to an INR in the target range (2-3 for all indications except valve prosthesis; for valve prosthesis the range is 2.5 - 3.5) on 3 consecutive measurements.
18	INR within therapeutic range of between 2-3 months for at least 3 months.
19	Same dose on 3 or more consecutive clinic visits, with INR within therapeutic range.
20	Subjects were required to be on warfarin and managed in the warfarin clinic for a minimum of 3 months before entry into the study, and thus dosing had stabilized by the time data collection for the study began. Values reported for warfarin dose are the average over the entire time of participation in the study (average = 20.6 months).
21	Stable dose defined as the dose that leads to a stable INR over three consecutive visits following initiation of the drug, with these INR measurements encompassing a period of at least 2 weeks, with a maximum difference between the mean daily dosages of 10%.

Section S3. Detailed Strategy for Creation and 10-fold Cross-Validation of Models

S3a. Model Creation

We gathered all data at the PharmGKB from the 21 contributing sites. Submitters used uninformative ID numbers and each study site was coded with uninformative identifiers. Eighty percent of patients with a target International Normalized Ratio (INR) of 2-3 were randomly selected as the model derivation cohort, in which numerous statistical approaches were tested for building a predictive warfarin dosing algorithm. Following ten-fold cross validation, the model with the lowest mean absolute error (MAE) was selected, then tested in a validation cohort, comprised of the remaining 20% of patients. Performance of the pharmacogenetic model was compared to a clinical model and empiric dose.

We created regression equations using ordinary linear and polynomial regression, artificial neural networks, support vector regression with polynomial (including linear) and Gaussian kernels, regression trees, model trees, least angle regression, Lasso and multivariate adaptive regression splines on the derivation set. We also experimented with boosing and bagging to construct ensemble predictors. Because the distribution of warfarin dose is skewed (with a longer tail at high doses), we created prediction algorithms for raw dose, as well as logarithm and square root of the dose. Where algorithms required the user to set parameters, such as the tradeoff between data fit and margin in a support vector machine (SVM) and the degree of polynomial in a polynomial-kernel SVM, parameters were tuned on each fold of cross-validation using only the training data for that fold, via an inner loop of cross-validation. Finally, we then compared all algorithms based on their cross-validated MAE.

The models produced by linear regression and support vector regression (with a linear kernel, i.e., a polynomial kernel with degree one) were statistically indistinguishable and significantly out performed all the other approaches. We selected the linear regression model because it used fewer variables, and is easier to understand and implement in a clinical setting.

S3b. Model Selection

As noted above, we tested a wide range of statistical modeling and machine learning algorithms for numerical real-valued prediction. Because missing values occurred frequently in the data, we attempted each of: imputation of missing values, omission of data points (patients) with missing values, and treatment of "missing" as an additional discrete value. Modeling was attempted on the raw response variable of stable weekly dose, as well as on logarithmic and square root transformations. After a model was derived, all log and square root transformations were converted back to mg/week units, so that the models could be compared directly. The MAE was always evaluated and compared as mg/week for all models and never log(mg/week) or square-root(mg/week).

Measuring values such as mean absolute error (MAE), percent within 1mg/day, and R-squared for each model on the training set from which the model is built can give an overly-optimistic estimate of the performance of the model on future patients. This problem is further exacerbated by

choosing the best-performing approach on the training data from among many different approaches. We solved this problem by having 20% of the data held aside as a final validation set; this validation set was chosen randomly, stratified by site (i.e., 20% of the patients from each of the 21 sites were held aside). Members of the team performing the modeling and analysis did not have access to this validation set until the final model was selected.

Each of the modeling methods considered was run on the training set to produce a predictive model. We then wished to compare these models to choose our one final model to run on the validation set. But methods capable of producing very rich, detailed models – for example, regression trees -- might overfit the training data; that is, they might produce a detailed model that performs well on the training data but poorly on the validation set. Therefore, we wanted a low-variance, unbiased (or nearly so) estimate of the performance of each modeling approach on which to base our selection of a final model, before turning to the validation set. A widely used method for obtaining such estimates is ten-fold cross-validation, which we employed for this purpose. Following standard protocol, we randomly partitioned the original training set into ten parts, 1, 2, …, N. On the ith iteration, for i from 1 to 10, we trained the model on all but the *ith* part of the data, and then we tested the model on the *ith* part by using it to predict the dose of each patient in that part. Each iteration of cross-validation constitutes a test on a held-aside set of data and can be used to give an unbiased estimate of the performance of the modeling approach; by repeating ten times with disjoint test sets, variance in these estimates owing to particular choice of test set is greatly reduced.

Our methodology therefore was to run each of the real-valued prediction methods, with multiple approaches for missing values, on the training set, and to estimate the mean absolute error of each of these methods by ten-fold cross-validation. We then selected the best method. In the end we had roughly a tie (no statistically significant difference) between ordinary linear regression and support vector regression. Because the linear regression model uses fewer variables, and hence is easier to understand and also to implement in a clinical setting, it is this model that we selected.

S3c. Other Evaluation Metrics and Significance Tests Employed

In addition to MAE, which we selected as the criterion for model selection, in this paper we also have reported other measures of model performance and have reported the results of significance tests. This section provides further details of these measures and tests.

R-square, also called the coefficient of determination and written as R^2, has several common definitions, all of which agree when applied to an ordinary linear regression fit on a derivation dataset, or training set, but which can disagree under other circumstances such as testing on a validation set. The two most widely-used definitions of R-square are as the square of the sample correlation coefficient R and the following.

$$\text{R-square} = 1 - \sum_i \frac{(y_i - f(\vec{x}_i))^2}{(y_i - \bar{y})^2}$$

Here y_i is the i-th response, $f(\vec{x}_i)$ is the model's prediction for the i-th data point, and $\bar{y}$ is the mean of the responses. We used the first definition, the square of the sample correlation

coefficient R for correlation between predicted and actual weekly doses. Adjusted R-square values were computed from R-square, based on the number of variables in a model, in the standard manner. Adjusting for R-square on a validation set technically should not be necessary, since during model selection (including selection of the variables to include), the validation set is not available. Hence while adding unnecessary variables can improve a training R-square, leading to overfitting and hence requiring adjustment to the R-square, adding such variables should not improve the validation R-square.

In addition to MAE and R-square, we use also use number needed to genotype (NNG) and percentage of patients whose predicted dose is within a given error tolerance of the actual stable dose. These are straightforward and are already defined in full within the text.

The text also reports the results of McNemar's test of paired proportions and sign tests. Both were employed to test, within various groups of patients, whether a difference between two models was significant. For example, was the difference in errors for the pharmacogenetic and clinical models significant? The following paragraph gives additional details.

For each patient in a group for which the models gave different predictions (a discordant pair), the numbers of times each model "won", or had lower error, were counted. The two-tailed p-value was computed in the standard manner for McNemar's test. For the sign test it was computed in the standard manner as the probability a model would have win at least as many more (or fewer) times than half under a binomial distribution with probability of a win equal to 0.5.

In the special case where the question was whether a model significantly overpredicted (or underpredicted) dose by at least 20%, the definition of a "win" on a discordant pair was changed to be did not overpredict (underpredict) by at least 20% when the other model did.

Section S4. Imputation of VKORC1 SNPs

Different sites had genotyped different *VKORC1* SNPs. Based on its prevalence in our data set and the evidence suggesting this is the likely functional SNP, we chose rs9923231 as our *VKORC1* SNP. Where it was missing, we imputed its value based on the other *VKORC1* SNPs and race, according to the decision list below. In a decision list, the first applicable condition is employed and all others are skipped. This algorithm is based on the linkage disequilibrium between SNPs as demonstrated in Rieder et al, 2005.[11] LD, presented as R^2 value between groups and calculated with data in the IWPC cohort, for the various SNP pairs was: rs9923231 and rs2359612 in Caucasian ($R^2 = 0.981$), Asian (0.995) and African American (0.463) populations, as well as between rs9923231 and rs9934438 (0.967, 0.996, 0.815 respectively) and rs8050894 (0.933, 0.995, and 0.248 respectively). Imputation of rs9923231 genotype was required in 20.2% of blacks, 28.6% of whites, and 15.8% of Asians.

I**f** Race is not "Black or African American" or "Missing or Mixed Race"**and** rs2359612='C/C'
then impute rs9923231='G/G'

If Race is not "Black or African American" or "Missing or Mixed Race"**and** rs2359612='T/T'
then impute rs9923231='A/A'

If Race is not "Black or African American" or "Missing or Mixed Race" **and** rs2359612='C/T'
then impute rs9923231='A/G'

If rs9934438='C/C' **then impute** rs9923231='G/G'

If rs9934438='T/T' **then impute** rs9923231='A/A'

If rs9934438='C/T' **then impute** rs9923231='A/G'

If Race is not "Black or African American" or "Missing or Mixed Race" **and** rs8050894='G/G'
then impute rs9923231='G/G'

If Race is not "Black or African American" or "Missing or Mixed Race" **and** rs8050894='C/C'
then impute rs9923231='A/A'

If Race is not "Black or African American" or "Missing or Mixed Race" **and** rs8050894='C/G'
then impute rs9923231='A/G'

Otherwise keep rs9923231 coded as “Missing”

Section S5. Sensitivity analysis testing performance of various dosing approaches using different cut-points to create low, intermediate and high dose groups (n=5052)

Dosing approach						
	Range	***≤2 mg/d***	***≤3 mg/d***	***≤3 mg/d***	***≤4 mg/d***	***≤4 mg/d***
	n	669	1711	1711	2706	2706
Pharmacogenetic	% ideal*	15.4	35.0	35.0	39.9	39.9
Clinical only	% ideal	5.4	24.0	24.0	35.5	35.5
Fixed 5mg/day	% ideal	0	0	0	0	0
	Range	***>2, <8 mg/d***	***>3, <8 mg/d***	***>3, <7 mg/d***	***>4, <7 mg/d***	***>4, <6 mg/d***
	n	4016	2974	2716	1721	1269
Pharmacogenetic	% ideal	53.4	55.5	55.9	60.2	60.4
Clinical only	% ideal	46.9	50.7	53.3	52.2	57.1
Fixed 5mg/day	% ideal	37.0	49.9	54.6	86.2	98.4
	Range	***≥8 mg/d***	***≥8 mg/d***	***≥7 mg/d***	***≥7 mg/d***	***≥6 mg/d***
	n	367	367	625	625	1077
Pharmacogenetic	% ideal	19.9	19.9	32.8	32.8	44.0
Clinical only	% ideal	6.3	6.3	13.3	13.3	24.0
Fixed 5mg/day	% ideal	0	0	0	0	21.8

*% ideal – percent of patients for whom the predicted dose was between 80-120% of the actual stable warfarin dose required by the patient.

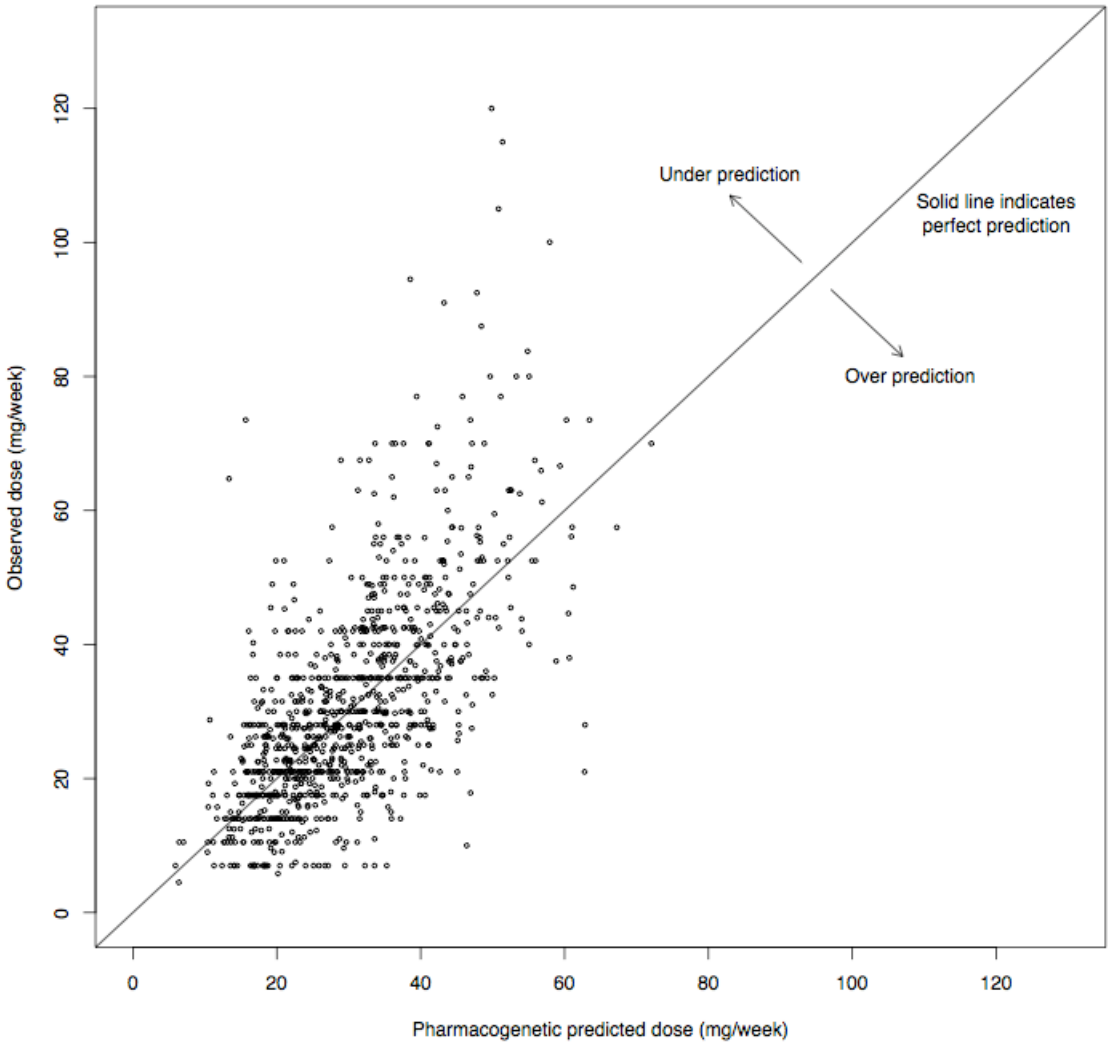
Section S6. IWPC Pharmacogenetic Algorithm in an Excel Workbook (available from NEJM website or at the PharmGKB: accession number PA162372936)

Age (Yrs) although only used at the grain of decades	Height (cm)	Weight (kg)	VKORC1 (Enter A/A; A/G; G/G; U (Unknown))	CYP2C9 (Enter *1/*1; *1/*2; *1/*3; *2/*2; *2/*3; *3/*3; Unknown)	Race (Enter A (Asian); B (Black or African American); C (Caucasian or White); U (Unknown or Mixed))	Taking Enzyme Inducer (Y; N)	Taking Amiodarone (Y; N)	Computed Weekly Starting Dose (mg/Week)

Section S7. **Relative contribution of each parameter to the explained variability (reported as R^2) in warfarin dose requirements.** Shown for the final parameters in the models are the variability explained by that parameter (R^2) in univariate analysis, as the partial R^2 in the pharmacogenetic model, and the partial R^2 from the clinical model.

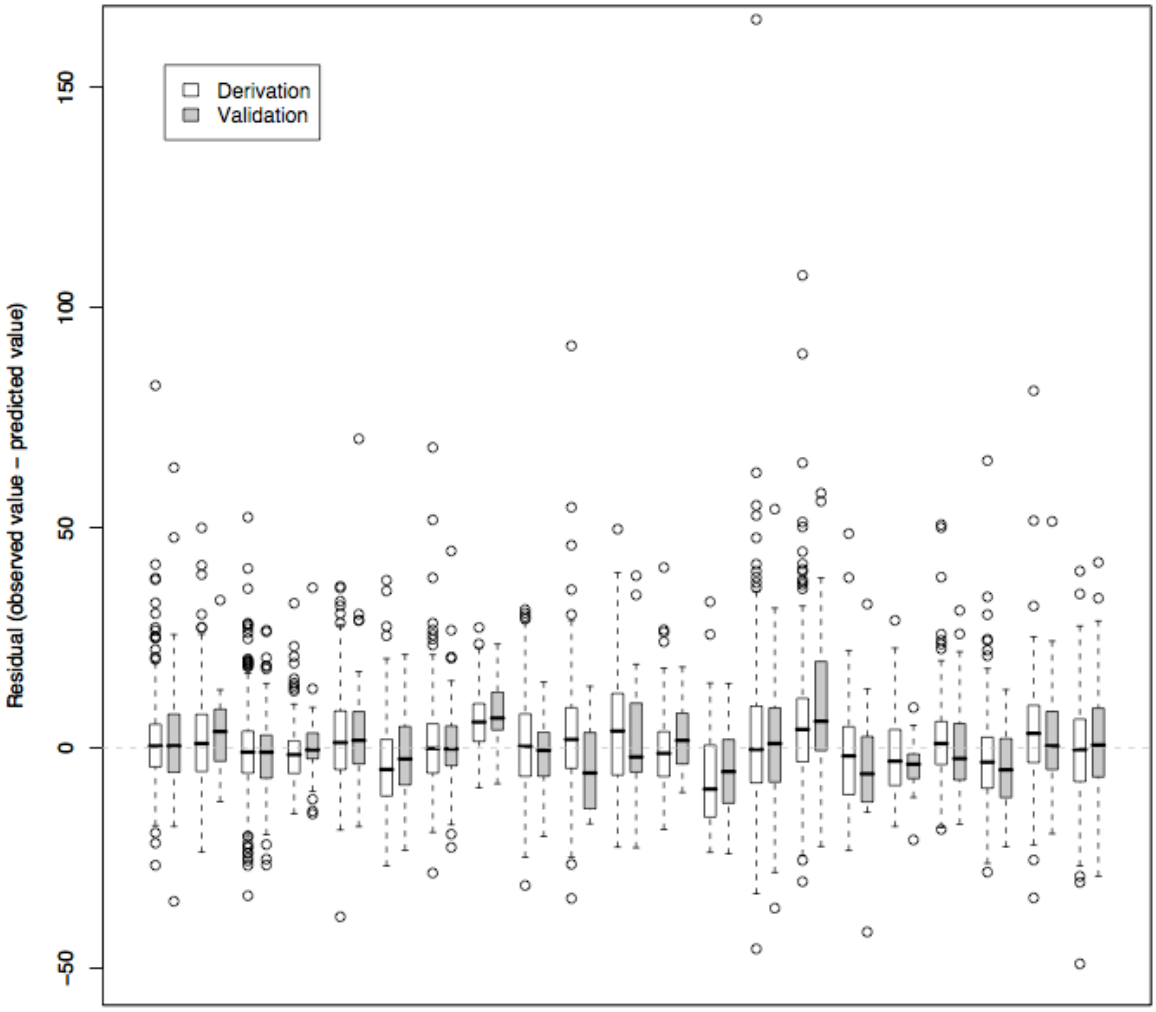
Variable	Univariate R^2	n	*p* value	Partial R^2 from multiple model (pharmacogenetic)	Partial R^2 from multiple model (clinical)
Age in decades	7.8	4043	2.63e-73	6.7	6.3
Height in cm	10.4	4043	3.49e-98	0.3	0.5
Weight in kg	16.9	4043	8.05e-165	2.3	2.5
VKORC1	27.7	4043	3.43e-284	13.3	NA
CYP2C9	5.5	4043	1.35e-46	7.2	NA
Race	14.2	4043	6.78e-134	0.3	4.4
Enzyme inducer status	1.5	4043	6.16e-15	0.7	0.8
Amiodarone status	0.1	4043	1.70e-02	0.6	0.7

Section S8. Predicted vs. Observed Stable Therapeutic Warfarin dose (mg/week) for 1,008 patients in the validation cohort (patient with observed dose of 315 mg/week was excluded).



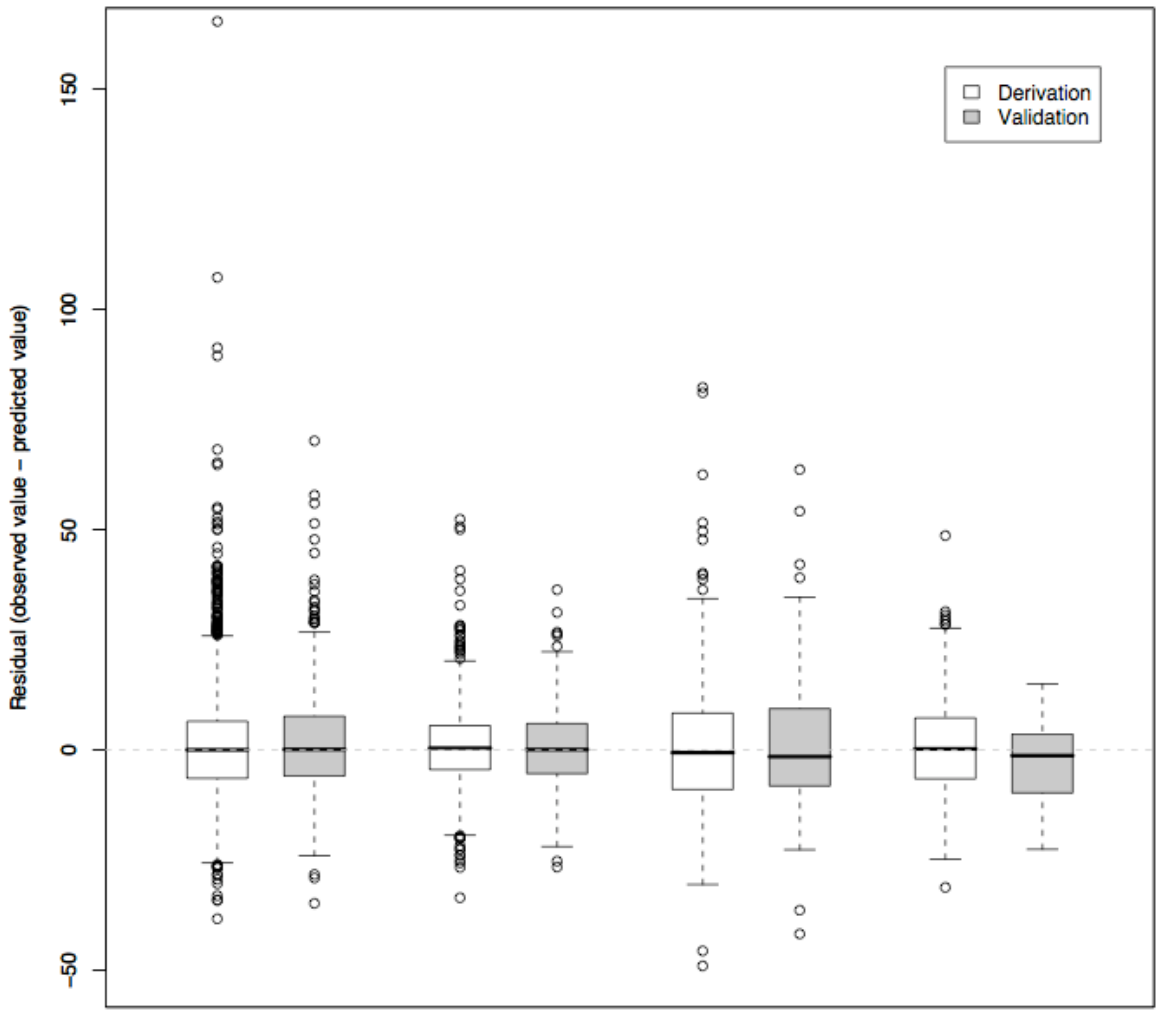
Section S9. Model performance across sites and by race

S9a. Model performance, reported as model R^2 and MAE for each of the 21 sites



Site	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Derivation n	290	97	631	167	251	224	349	207	171	113	89	101	50	240	252	66	67	216	162	70	230
Derivation MAE	9	7.4	6.8	5.2	8.4	8.6	6.6	8.9	6.4	9	11.5	6.6	9.7	11.1	13.9	12.6	5.6	7.9	8.1	10.2	10.4
Derivation r2	52.4	33.9	14	37.5	53.6	47.9	51.3	36.4	34.2	39.9	40.7	58.6	57.2	32	39.7	41.3	44.8	53.6	35.5	35.4	38.4
Validation n	75	28	158	39	60	56	87	52	42	26	19	28	13	60	62	16	20	50	41	19	57
Validation MAE	9	7.4	6.8	5.2	8.4	8.6	6.6	8.9	6.4	9	11.5	6.6	9.7	11.1	13.9	12.6	5.6	7.9	8.1	10.2	10.4
Validation r2	37.8	54.7	16.6	53.6	53	37.6	54.2	27.1	56.4	30.3	31.3	61.4	33	29.6	25.9	1.3	67.3	51.2	58.2	25.2	29.1

S9b. Model performance, reported as model R^2 and MAE for the major race groups



Race	White	Asian	Black or African American	Missing or Mixed
Derivation n	2233	1229	353	228
Derivation MAE	8.708	6.429	11.563	8.907
Derivation r2	44.9	32.9	25.9	35.5
Validation n	561	300	97	50
Validation MAE	8.711	7.177	11.984	7.272
Validation r2	39.5	34.2	19.2	47.3

Section S10. Number needed to genotype analysis

The number needed to genotype (NNG) in order to see a significant benefit of using genetics was computed using standard number needed to treat method, and was the inverse of the absolute risk reduction (ARR). The ARR was computed as the absolute difference between the event rate (ER) for the pharmacogenetic algorithm and the event rate for the clinical algorithm. An "event" was defined as the algorithm predicting a dose more than 20% above or below the actual therapeutic dose (i.e. an event is a poor dose estimate). The ER was the ratio of the number of patients for which an algorithm estimates a poor dose over the total number of patients. Table S10a provides the values for the clinical vs. pharmacogenetic algorithm with this 20% criteria. Table S10b provides the values for the fixed vs. pharmacogenetic algorithm with the 20% criteria.

S10b. Clinical vs. pharmacogenetic with 20% criteria

	All doses
clinical > 20% from actual	3110
clinical ≤ 20% from actual	1942
pharmacogenetic > 20% from actual	2730
pharmacogenetic ≤ 20% from actual	2322
absolute risk reduction	0.0757
number needed to genotype	13.2

S10b. Fixed vs. pharmacogenetic with 20% criteria

	All doses
fixed > 20% from actual	3568
fixed ≤ 20% from actual	1484
pharmacogenetic > 20% from actual	2730
pharmacogenetic ≤ 20% from actual	2322
absolute risk reduction	0.166
number needed to genotype	6.0

Section S11. Analysis with single very large outlier

The validation cohort included one case with a daily dose of 45 mg/day (315 mg/week). This dose was considered extraordinarily high (and rare), and may represent an artifact, e.g. non-adherence to therapy. Table 2 provides a report without this case included in the validation or total counts. This supplement provides the data if this case is included. The derivation MAE does not change.

Comparison of prediction performance on validation set relative to final stable warfarin dose of the pharmacogenetic algorithm, clinical-only algorithm, and fixed 5 mg/day starting dose.

	Derivation MAE[a] (mg/week) (Std Error, R^2)	Validation MAE (mg/week) (Std Error, R^2)
Pharmacogenetic algorithm*,**	8.3 (1.7, 0.47)	8.8 (2.1, 0.33)
Clinical-only algorithm**	10.0 (2.0, 0.28)	10.1 (2.3, 0.21)
Fixed 5 mg/day	13.3 (2.4, -0.07)	13.2 (2.3, -0.04)

[a]MAE – Mean Absolute Error, reported with standard error and R^2 for pharmacogenetic and clinical algorithms.

*Pharmacogenetic versus clinical: significantly different with $p = 0.000127$ by McNemar's test of paired proportions.

**Pharmacogenetic versus fixed dose; and clinical versus fixed dose differences: significantly different with $p < 0.000001$ by McNemar's test of paired proportions.