Supplementary Material: Discussion and SAS Code

Assumption of patients' exchangeability

The assumption of patients' exchangeability underlies this analysis. Indeed, inference is a process whereby one passes from data on a set of units to statements about a further unit. This linkage can be formulated in terms of judgments of exchangeability between the unit and the data; or, alternatively expressed, judgments of which subpopulation the unit belongs to.[1] Specifically, we assumed that the optimal coumarin dose calculated for patients in our multicenter study may be applied to a further, new patient. Evidence of clinical utility for the studied markers is a key issue in translating pharmacogenomics into clinical practice and the extent to which the optimal coumarin dose differs between subgroups defined by the markers is an important component of assessing clinical utility. In order to properly take into account genetic differences across subpopulations, we assumed exchangeability within subgroups of individuals from the same race and taking the same type of drug (warfarin or acenocoumarol), and therefore performed separate analyses according to each race/drug combination.

Nevertheless, beyond the identification of appropriate subpopulations, another important aspect to be taken into account is the assumption of "no unmeasured confounders" underlying exchangeability assumption.[2] We tried to assess the issue of possible unmeasured confounding by creating, as sensitive analyses, new models including further covariates beyond the ones included in the main analysis, and we obtained similar R Square values. A possible residual confounding, however, could not be completely ruled out.

In general, when the assumption of exchangeability does not hold, possible alternative analyses may be performed, including promising Bayesian approaches that enables personalized treatment selection for new patients.[3,4] Indeed, implicit to the concept of precision medicine is heterogeneity of treatment benefit among patients and patient subpopulations. Further researches and different

types of analyses may be helpful in the future to obtain optimal coumarin doses estimates in the era of precision medicine.

References

1. Lindley, D. V., Novick, M. R. The Role of Exchangeability in Inference. *Ann. Stat.* 9, 45-58 (1981)
2. Greenland, S., Robins, J. M., Pearl, J. Confounding and collapsibility in causal inference. *Stat. Sci.* 14, 29-46 (1999)
3. Ma, J., Stingo, F. C., Hobbs, B. P. Bayesian predictive modeling for genomic based personalized treatment selection. *Biometrics* 72, 575-583 (2016)
4. Hobbs B.P., and Landin R. Bayesian basket trial design with exchangeability monitoring. *Stat. Med.* 37, 3557-3572 (2018)

SAS code

```
/*1)Two-stage analysis for the association between CYP4F2*3 polymorphism and
stable coumarin dose */

/*Study-specific estimates*/
/*PARAMETERS:
      COHORT: name of dataset containing subjects for each specific cohort;
      CLASS: study-specific categorical variables (i.e. race);
      COVARIATES: study-specific covariates as described in Table 1 + CYP2C9*2
and *3 combined and VKORC polymorphisms (both genes coded as 0, 1, 2 for no, 1
or 2 variant alleles)*/

%MACRO GLM_ANALYSIS (COHORT= , CLASS= , COVARIATES= );
proc glm data=&COHORT.;
class &CLASS.;
model lndose=CYP4F2_V433M_D &COVARIATES./solution; /*lndose=logarithmic
transformation of dose, CYP4F2_V433M_D coded as 0 (CC) or 1 (CT+TT)*/
ods output ParameterEstimates=beta;
run;quit;
%MEND;

/*meta-analysis and meta-regression with STATA, 'metan' command:
metan Beta StdErr **possible covariate for meta-regression**, random classic
lcols( Study ) sortby( newID ) nowarning  xlabel(-0.50, 0.00, 0.50) boxsca(50)
textsize(100)  nowt diamopt(lcolor(black))*/
```

```
/*2)Stable coumarin dose predictive model*/

/*A)MAIN ANALYSIS: TABLE 3*/
/*Sampling*/
proc surveyselect data=WAR method=srs samprate=0.66 seed=0 out=sampling;
strata cohort_n;
run;
data warfarin.test;
set sampling;
test=1;
run;
proc sort data=WAR;by IDcumulative;run;
proc sort data=warfarin.test;by IDcumulative;run;
data tot;
merge WAR warfarin.test;
by IDcumulative;
if test=. then test=0;
run;
data warfarin.valid;
set tot;
if test=0;
run;
/*Create different datasets for each available combination of race and drug and
perform analysis for each of these combinations*/
/*Example for whites+warfarin*/
data white_warfarin;/*test set*/
set warfarin.test;
if race='White';if Warfarin_0_Aceno_1=0;
run;
data white_warfarin_valid;/*validation set*/
set warfarin.valid;
if race='White';if Warfarin_0_Aceno_1=0;
run;
/*Glm models on training set and validation*/
/*COVARIATES:
Age_years: continuous;
BMI: continuous;
sex: 0=Female 1=Male;
IND: 0=DVT OR PE OR heart valve OR stroke OR others  1=fibrillation/flutter OR
cardiomyopathy/LV dilation OR post orthopedic;
CYP2C9_a2_3: 0=2 variant alleles   1=1 variant allele     2=0 variant alleles
(reference);
CYP2C9_a3_3: 0=2 variant alleles   1=1 variant allele     2=0 variant alleles
(reference);
*only for Blacks: CYP2C9_a5_3: 0=2 variant alleles   1=1 variant allele     2=0
variant alleles (reference)*;
VKORC_3:     0=2 variant alleles   1=1 variant allele     2=0 variant alleles
(reference);
CYP4F2_3:    0=2 variant alleles   1=1 variant allele     2=0 variant alleles
(reference)*/
proc glm data=white_warfarin;
class CYP2C9_a2_3 CYP2C9_a3_3 VKORC_3 CYP4F2_3;
model lndose=AGE_years BMI sex IND CYP2C9_a2_3 CYP2C9_a3_3 VKORC_3
CYP4F2_3/solution clparm;
store work.Score_white_warfarin;
run;quit;
```

```
proc plm restore=work.Score_white_warfarin;
score data=white_warfarin_valid out=Pred1;
run;
proc glm data=Pred1;
model lndose=predicted/solution;
run;quit;

/*B)SENSITIVITY ANALYSIS: SUPPLEMENTARY TABLE S1*/
/*Create different datasets for each available combination of race and drug and
perform analysis for each of these combinations*/
/*Example for whites+warfarin*/
data white_warfarin_all;/*test+validation cohort*/
set WAR;
if race='White';if Warfarin_0_Aceno_1=0;
run;
/*Glm models on the whole cohort*/
proc glm data=white_warfarin_all;
class CYP2C9_a2_3 CYP2C9_a3_3 VKORC_3 CYP4F2_3;
model lndose=AGE_years BMI sex IND CYP2C9_a2_3 CYP2C9_a3_3 VKORC_3 CYP4F2_3
amiodarone/solution;/*amiodarone code 0=no     1=yes*/
run;quit;
proc glm data=white_warfarin_all;
class CYP2C9_a2_3 CYP2C9_a3_3 VKORC_3 CYP4F2_3;
model lndose=AGE_years BMI sex IND CYP2C9_a2_3 CYP2C9_a3_3 VKORC_3 CYP4F2_3
CYPind CYPinhib/solution;/*CYPind (All CYP-inducers) code        0=no  1=yes
CYPinhib (All CYP-inhibitors) code  0=no  1=yes*/
run;quit;
proc glm data=white_warfarin_all;
class CYP2C9_a2_3 CYP2C9_a3_3 VKORC_3 CYP4F2_3;
model lndose=AGE_years BMI sex INDICATION_01 CYP2C9_a2_3 CYP2C9_a3_3 VKORC_3
CYP4F2_3 amiodarone smokerstatus/solution;/*smokerstatus code 0=no     1=yes*/
run;quit;
```

```
/*C)PREVIOUS MODELS COMPARISON: SUPPLEMENTARY TABLE S3*/
/*Create different datasets for each available combination of race and drug and
perform analysis for each of these combinations*/
/*Example for whites+warfarin*/
/*Consortium IWP: create data set NE_validation from white_warfarin_valid by
recoding variables as follow:
age_dec: 1=10-19 years  2=2-29 years      3=30-39 years     ...;
heightcm:continuous;
weightkg: continuous;
VKORC1_NE:              0=GG/AA/missing              1=AG (dummy variable);
VKORC2_NE:              0=GG/AG/missing              1=AA (dummy variable);
VKORCmiss_NE:           0=GG/AG/AA                   1=missing (dummy variable);
CYP2C912_NE:            0=11/22/13/33/23/missing     1=12 (dummy variable);
CYP2C913_NE:            0=11/22/12/33/23/missing     1=13 (dummy variable);
CYP2C922_NE:            0=11/12/13/33/23/missing     1=22 (dummy variable);
CYP2C923_NE:            0=11/12/13/33/22/missing     1=23 (dummy variable);
CYP2C933_NE:            0=11/12/13/22/23/missing     1=33 (dummy variable);
CYP2C9miss_NE:          0=11/12/13/22/23/33          1=missing (dummy variable);
CYPind (All CYP-inducers):         0=no  1=yes;
Amiodarone:                        0=no  1=yes;
*/
data NE_model;
set NE_validation;
score_NE=5.6044-0.2614*age_dec+0.0087*heightcm+0.0128*weightkg-0.8677*VKORC1_NE-
1.6974*VKORC2_NE-0.4854*VKORCmiss_NE-0.5211*CYP2C912_NE-0.9357*CYP2C913_NE-
1.0616*CYP2C922_NE-1.9206*CYP2C923_NE-2.3312*CYP2C933_NE-
0.2188*CYP2C9miss_NE+1.1816*CYPind-0.5503*Amiodarone;
expscore_NE=score_NE*score_NE;
run;
proc glm data=NE_model;
model WEEKLYSTABLEDOSEmg=expscore_NE;
where score ne .;/*for comparison with our score: only in the subgroup of
subjects for which we can calculate our score*/
run;quit;
/*Gage: create data set GAGE_validation from white_warfarin_valid by recoding
variables as follow:
age_years: continuous;
BSA: sqrt(heightcm*weightkg/3600);
INRtarget: continuous;
VKORC11639GgtA_CF:      0=GG              1=AG        2=AA;
CYP2C9_a3_CF:           0=11/22/12        1=13/23     2=33;
CYP2C9_a2_CF:           0=11/33/13        1=12/23     2=22;
Amiodarone:             0=no  1=yes;
smokerstatus:           0=no  1=yes;
DVT_PE:                 0=no  1=yes;
*/
data GAGE_model;
set GAGE_validation;
score_GAGE=0.9751-0.00745*age_years+0.4317*BSA+0.2029*INRtarget-
0.3238*VKORC11639GgtA_CF-0.4008*CYP2C9_a3_CF-0.2066*CYP2C9_a2_CF-
0.2538*Amiodarone+0.0922*smokerstatus+0.0664*DVT_PE;
expscore_GAGE=7*(exp(score_GAGE));
run;
proc glm data=GAGE_model;
model WEEKLYSTABLEDOSEmg=expscore_GAGE;
```

```
where score ne .;/* for comparison with our score: only in the subgroup of
subjects for which we can calculate our score*/
run;quit;



/*D) GENE-GENE INTERACTION: TABLE 4*/
/*Create different datasets for each available combination of race and drug and
perform analysis for each of these combinations*/
/*Example for whites+warfarin*/
/*Gene codes:
CYP2C9:     0=11  1=12/22/13/33/23;
VKORC:      0=GG  1=AG/GG;
CYP4F2:     0=CC  1=CT/TT;
intCYP2C9: CYP2C9*CYP4F2;
intVKORC: VKORC*CYP4F2;
intCYP2C9VKORC: CYP2C9*VKORC;
*/
proc glm data=white_warfarin_all;
class cohort_n;
model lndose=cohort_n AGE_years BMI sex IND CYP2C9 VKORC CYP4F2 intCYP2C9
intVKORC intCYP2C9VKORC/solution;
run;quit;
```

```
/*E) GENE-DRUG INTERACTION AND STRATIFIED ANALYSIS: SUPPLEMENATARY TABLE S2*/
/*Create different datasets for each available combination of race and drug and
perform analysis for each of these combinations*/
/*Example for whites+warfarin*/
/*Gene codes:
CYP2C9_2:   0=11/13/33  1=12/22/23;
CYP2C9_3:   0=11/12/22  1=13/33/23;
VKORC:      0=GG        1=AG/GG;
CYP4F2:     0=CC        1=CT/TT;
*/
%MACRO DRUGS;
%let drugs = amiodarone azoli CYP2C9inducers CYP2C9inhibitors PPI Statin aspirin
carbamazepina otherCYPinducers otherCYPinhibitor rifampin CYPinhib CYPind;
%let i=1;
%do %while (%scan(&drugs, &i) ne );
   %let next_drug = %scan(&drug, &i);

proc glm data=white_warfarin_all; /*drug effect (column 1)*/
class cohort_n;
model lndose=cohort_n AGE_years BMI sex  IND &next_drug CYP2C9_2 CYP2C9_3 VKORC
CYP4F2/solution;
ods output Nobs=N&next_drug;
ods output ParameterEstimates=estimates&next_drug;
run;quit;

proc glm data=white_warfarin_all; /*gene-drug interaction (columns 2--4*/
class cohort_n;
model lndose=cohort_n AGE_years BMI sex IND &next_drug CYP2C9_2 CYP2C9_3 VKORC
CYP4F2 &next_drug*CYP2C9_2 &next_drug*CYP2C9_3 &next_drug*CYP4F2/solution;
ods output ParameterEstimates=int&next_drug;
run;quit;

proc sort data=white_warfarin_all;/*stratified analysis (columns 5--10 */
by &next_drug;run;
proc glm data=white_warfarin_all;
class cohort_n;
model lndose=cohort_n AGE_years BMI sex IND CYP2C9_2 CYP2C9_3 VKORC
CYP4F2/solution;
by &next_drug;
ods output ParameterEstimates=stratestimates&next_drug;
run;quit;

%let i = %eval(&i + 1);
%end;
%mend;
%DRUGS;
```