STAT 9530 – Data Mining and Machine Learning Methods

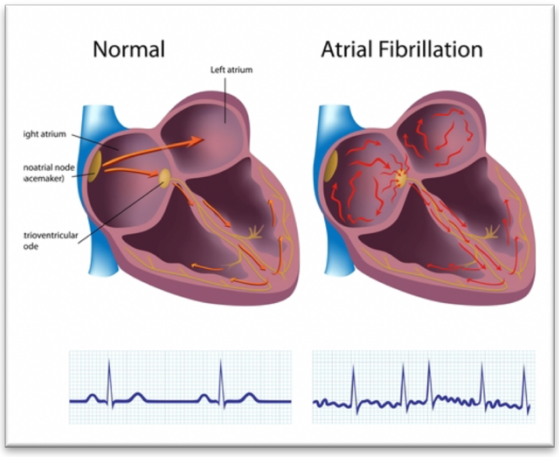
Project: PERSONALIZED ANTI-COAGULATION: OPTIMIZING WARFARIN MANAGEMENT USING GENETICS AND SIMULATED CLINICAL TRIALS

Ravvaz et al.

Submitted by Kat Usop

Submitted on May 11th 2018

INTRODUCTION AND PROBLEM DESCRIPTION



There are conflicting results with clinical trials testing pharmacometric-guided warfarin therapies for patients with atrial fibrillation. Patient safety is at stake when warfarin is not prescribed appropriately.

Warfarin, an oral drug caters to the condition of AF, is prescribed in such as a type of **anticoagulant** medication that helps prevent clots from forming in the blood. It relies heavily on the genetic markup and other biomarkers for an optimal drug dosing recommendation. Given that the genetic markup of each individual is fairly unique, this proposes a unique challenge for the medical community to develop a standard prescription recommendation across diverse populations. It is also time-consuming and perhaps difficult to implement, on a whole healthcare sphere that is, personalized one-to-one sequential diagnosing. This opens a unique opportunity for an exploration of whether precision medicine of "Subpopulation" can balance the need for personalized prescription on an individual level vs. standardization of gene-reliant drug prescription on a bigger spectrum.

BACKGROUND STUDIES

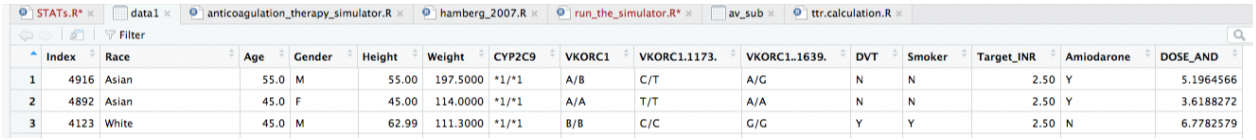
The paper "personalized Anticoagulation Optimizing Warfarin Management Using Genetics and Simulated Clinical Trials" by Ravvaz et al. (2014) is heavily based on the paper by Hamberg et. Al (2007) titled "PK- PD Model for Predicting the Impact of Age, CYP2C9, and VKORC1 Genotype on Individualization of Warfarin Therapy". The later paper focused on the characterization of the relationship between warfarin concentrations and INR (International Normalized Ratio) response to indicate plausible predictions for "dose individualization". Consequently, it aided on the formulation of the 5-arm protocols which is a neural network of 1 hidden layer which accepts partitioned synthetic dataset called Clinical Avatars from different

nodes. The predicted outcome is INR which is calculated from the ratio of a patients' prothrombin time to a normal (Control) sample, raised to the power of the international Sensitivity Index (ISI) value for the analytical system used.

The study design constitutes of 5 phases:

Medical Records extracted from Aurora Hospital Network, Bayesian Network model patient records, Synthetic dataset aka. Clinical AVATARS, cohorts without replacement into 5 arm protocol, aggregation and compare outcomes for targeted subpopulations.

DATA DESCRIPTION



STATs.R* | data1 | anticoagulation_therapy_simulator.R | hamberg_2007.R | run_the_simulator.R* | av_sub | ttr.calculation.R

Filter

	Index	Race	Age	Gender	Height	Weight	CYP2C9	VKORC1	VKORC1.1173.	VKORC1..1639.	DVT	Smoker	Target_INR	Amiodarone	DOSE_AND
1	4916	Asian	55.0	M	55.00	197.5000	*1/*1	A/B	C/T	A/G	N	N	2.50	Y	5.1964566
2	4892	Asian	45.0	F	45.00	114.0000	*1/*1	A/A	T/T	A/A	N	N	2.50	Y	3.6188272
3	4123	White	45.0	M	62.99	111.3000	*1/*1	B/B	C/C	G/G	Y	Y	2.50	N	6.7782579

The dataset is a data.frame with 5,700 observations of 15 variables, with a mix of categorical and numerical variables. They are derived from a hospital facility called Aurora HealthCare in Milaukee, Wisconsin USA. The study population in this dataset was extracted from longitudinal EMRs of patients with Atrial Fibrillation who were taking Warfarin over a period of 2002 to 2012. Without doubt, this patient dataset is de-identified. The attributes are classified into Biodata, Drug prescription, and Genotypes**. One major hindrance of this dataset is the lack of "time" stamps and "concentration" level. It became a root issue in the future analysis conducted in this project. A less than optimal "work-around" is established to extract estimations and genetic variation to individual-based analysis.**

The most significant attributes are:

Biodata:

Race: Asian with 1634 obs., African American with 462 obs., , Unknown, 482 obs, and White with 3122 obs.

Age: 1st Qu:. 55.00, Mean: 64.14, Median: 65.00, 3rd Qu.: 75.00, and the Max. Age observed is 95.00 **Gender:** Female with 2373 obs., and Male with 3327 obs. Genotypes:

CYP2C9 is a gene which provides instructions for making an enzyme that breaks down warfarin, with 4165 obs. On *1/*1 (normal warfarin response), 737 on *1/*2, 554 obs. On *1/*3, 56 obs. On *2/*2, 99 obs. On *2/*3 and. Finally 89 obs. On *3/*3 (abnormal warfarin response)

VKORC1..1639 encodes for the enzyme: Vitamin K epOxide Reductase Complex (VKORC) subunit 1. It is responsible for reducing vitamin K 2,3-epoxide to its active form. This is crucial for blood clotting. A/A has 1821 observations, A/G has 1904 observations and G/G has 1975 observations.

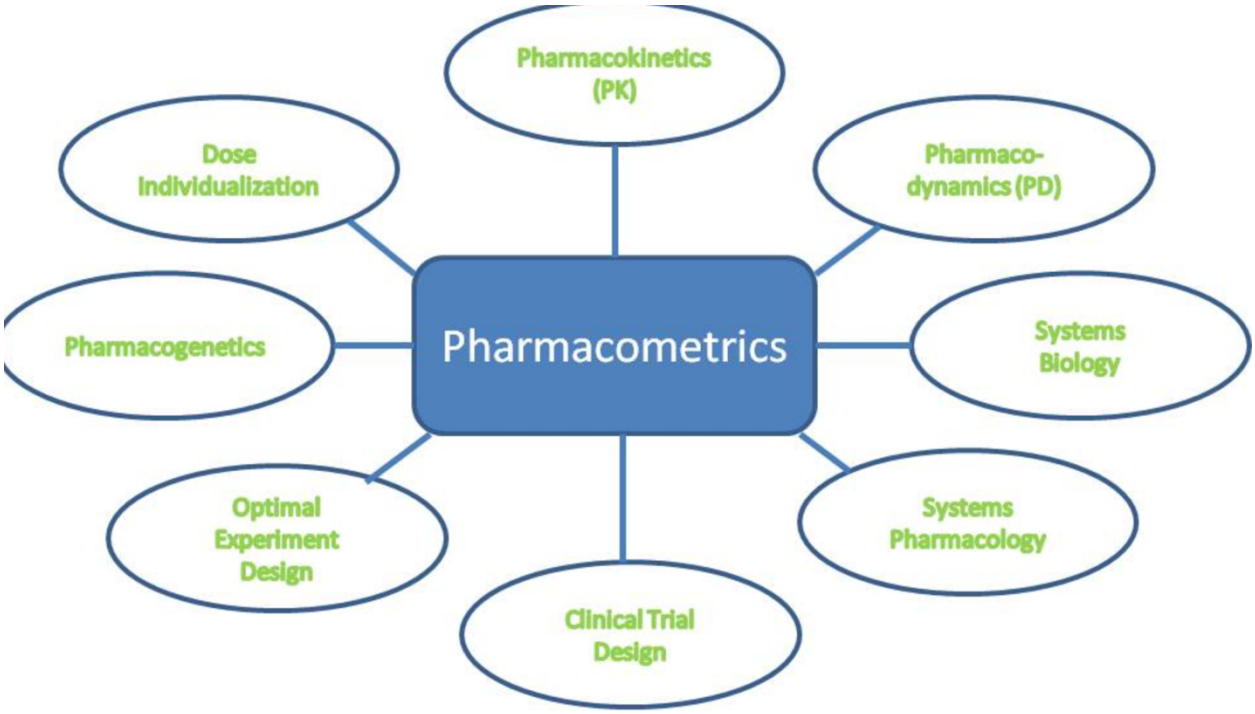
Drug Prescription:

DVT: Fluvastatin: keep (atherosclerosis) from getting worse has 572 records saying Yes, and 5128 observations that didn't take it.
Amiodarone: Anti-arrhythmic Agent has 1715 observations with a Yes and 3985 records who didn't take it.

The rest: Weight, Height, Smoker, Target INR: International Normalized Ratio, Index

MODELS TO BE USED

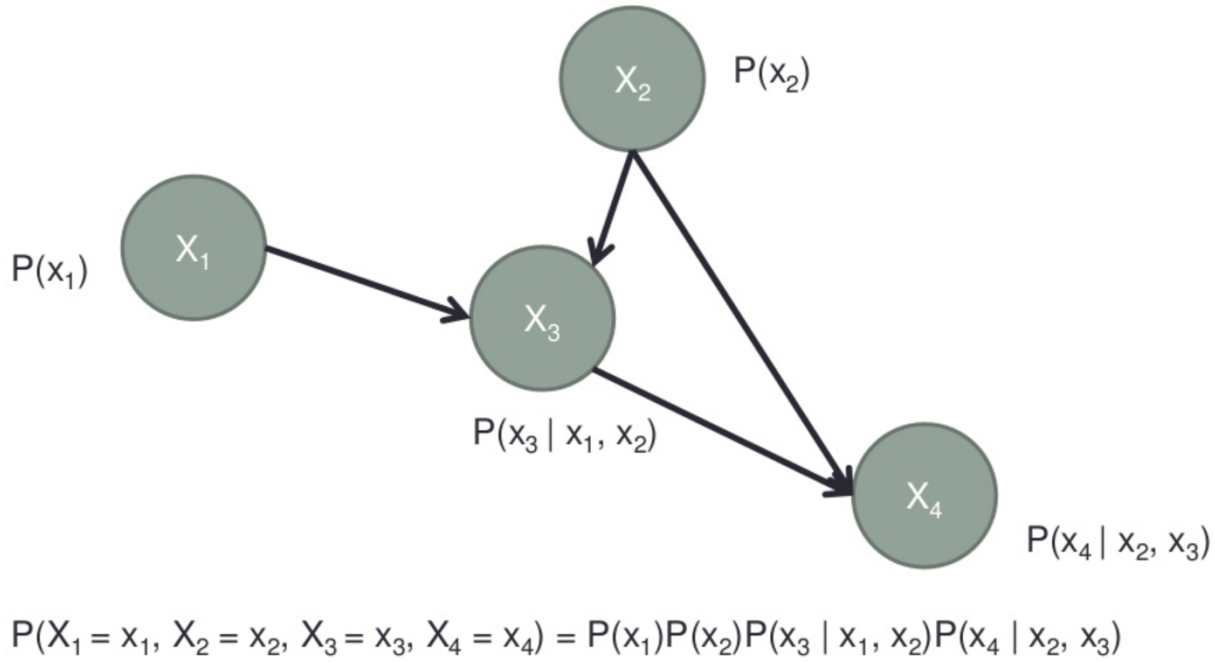
Introduction to Pharmacometrics



The multidisciplinary field of pharmacometrics is a applied quantitative field that pander to the interpretation of pharmacological observations. It is a plausible bridge between observation collection and the understanding of medical science. In application, a pharmacometric implementation involves mathematical and statistical models to aid in answering vital questions within this field. Mixed effects models became a primary analysis for population-based pharmacometric modeling (PK/PD) given the competence of this model to handle sparse data. In

more depth, pharmacometrics contains pharmacostatistical modeling and simulation techniques that are considerable for a wide variety of drug-response and relationship in pharmacotherapy.

Bayesian Network Modelling



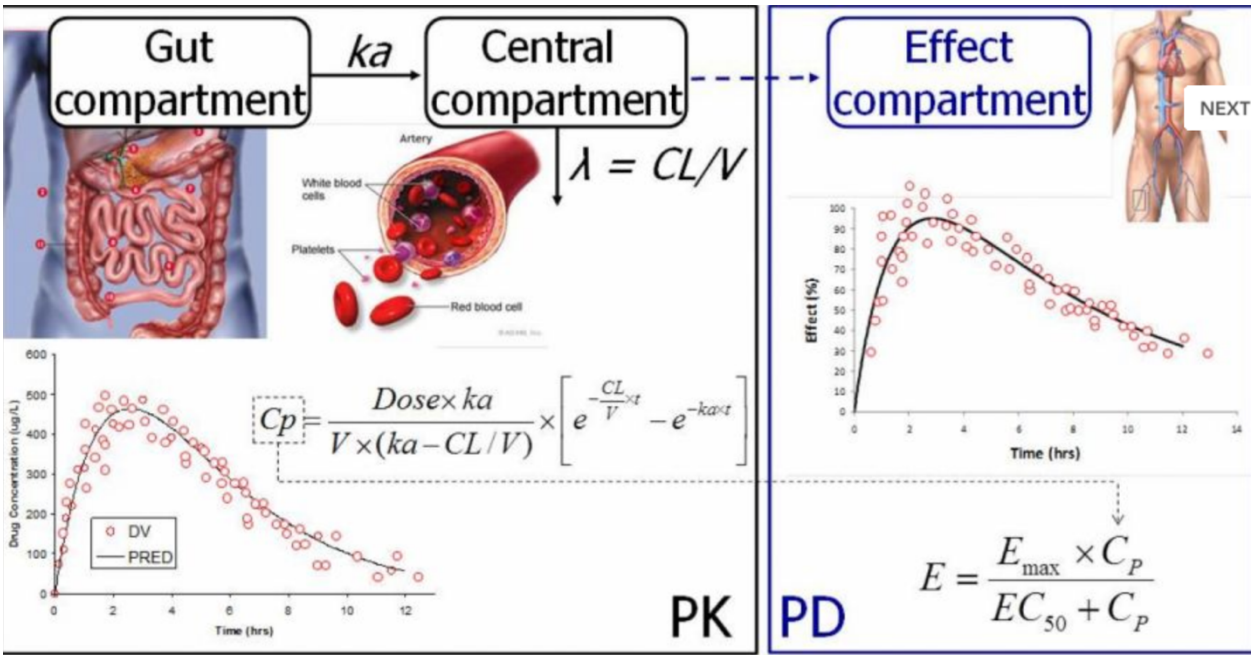
Bayesian network is a probabilistic statistical model that portrays the relationships and conditional dependencies between attributes via a mechanism called DAG, short for "Directed Acyclic Graph" from the observed data. Structural learning within a belief network is performed by the definition of the score metric and the search algorithm with a good amount of learning and training for a plausibly accurate DAG result. Outcomes from the BNM conditional probability of each vertex given its codependent variables.

Bayesian Network Modelling matters in the pharmacometric field due to its nature of being a probabilistic graphical model which is compatible in modelling different types of biological systems. It can help in representing causal associations between genetic markups of the population observed in this project.

Package used: BNlearn

BnLearn is an R package that caters to Bayesian Network Structure Learning, Parameter Learning and Inference. Given that the dataset is a mixed type with categorical and numerical variables, a conditional gaussian distribution is implemented. CPTs and DAGs were the most prominent outcomes that were vital in understanding the population observed.

Population PK/PD Analysis



Population PK/PD short for Pharmokinetic and Pharmacodyanmic modeling framework is a significant methodology for quantitating and explaining variability in drug exposure and response, especially for drug therapies that rely heavily on individualized dosage.

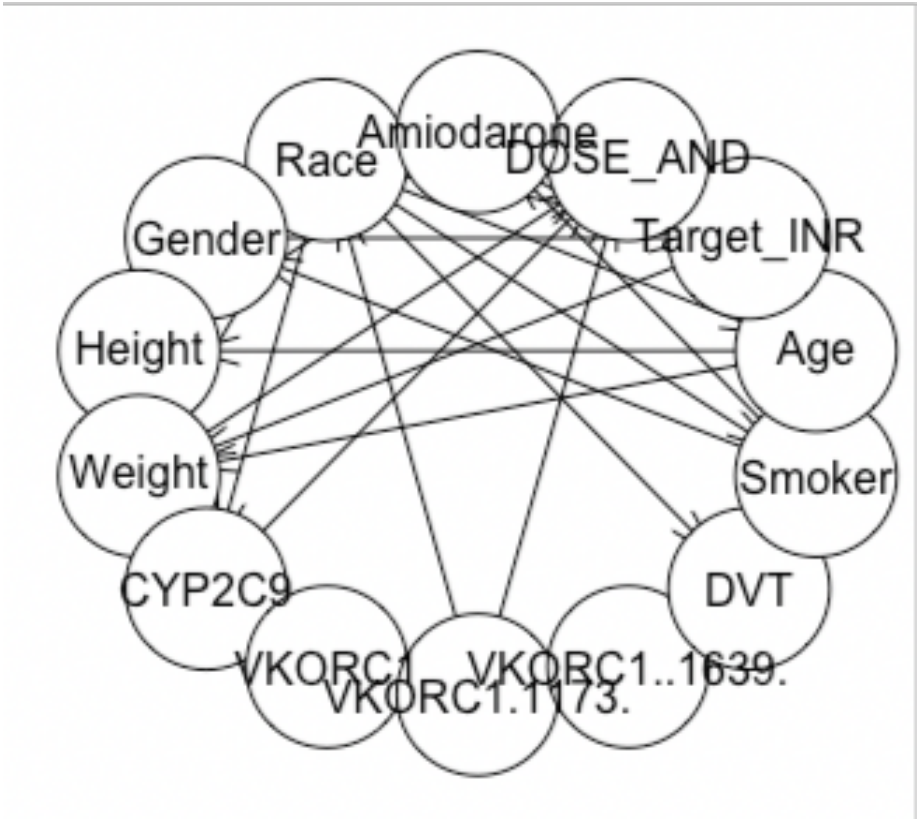
PK/PD models approximate both temporal and quantitative aspects of drug response. This helps us to distinguish the time delay of drug response. This compound model can aid in recommending dosage that are tailor-fit for a subpopulation for any targeted INR. It is, however, highly dependent on supplemental tools to solve a set of differential equations prior any predicted outcome to manifest. It is not a stand-alone tool which can hinder the initial exploration of such model.

MIXED EFFECTS MODELLING

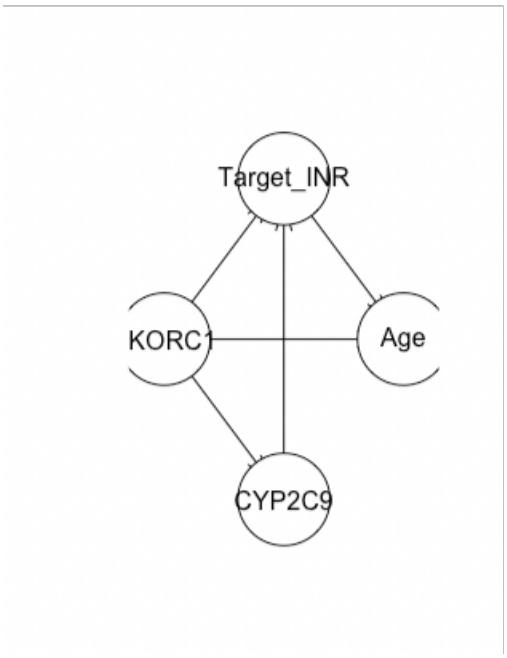
A mixed model is a statistical model which represent the relationship between a response variable and independent variables with coefficients that can change to one or more grouping variables. In this project, the genetic markup that are direct effectors in the drug response are grouped together "**VKORC1.1173. and CYP2C9**" while the **"TARGET_INR"** with ~ **"DOSE"** are regarded as the response variables. In the mixed effects modelling, the data points portrayed their co-interdependence. Within the mixed effects model, there are the fixed effects and the random effects, also known as variance components model. In biostatistics, fixed effects refer to "population-average" and random effects to "individual-specific effects". Fixed effects assumption can be summarized in this question "What is the best estimate of the dosage regimen?" and the random effects assumption will try to answer the question "What is the average treatment effect given these unique genetic markups?". Maximum likelihood approach is used in this analysis given that the random effects absorbs all possibilities what has been and could be.

PRELIMINARY DATA ANALYSIS (MIDTERM)

Bayesian Network Modelling: Directed Acyclic Graph Learning Structure



The DAG is derived from the Aurora AF dataset, which is a probabilistic graphic model in which the nodes represent the variables within the dataset, the lack of arcs represents "conditionally independence". The DAG is used to expound the parametric model of the data at hand. It uses the greedy algorithm within the ML Bayes simulator.



Given that the most significant variables in predicting INR are Age, and the 2 genotypes of VKORC1 and CYP2C9, I have implemented a BNM and DAG on the subset to explore the causal dependencines among them as well.

BNM: Conditional Probability Tables on observed dataset

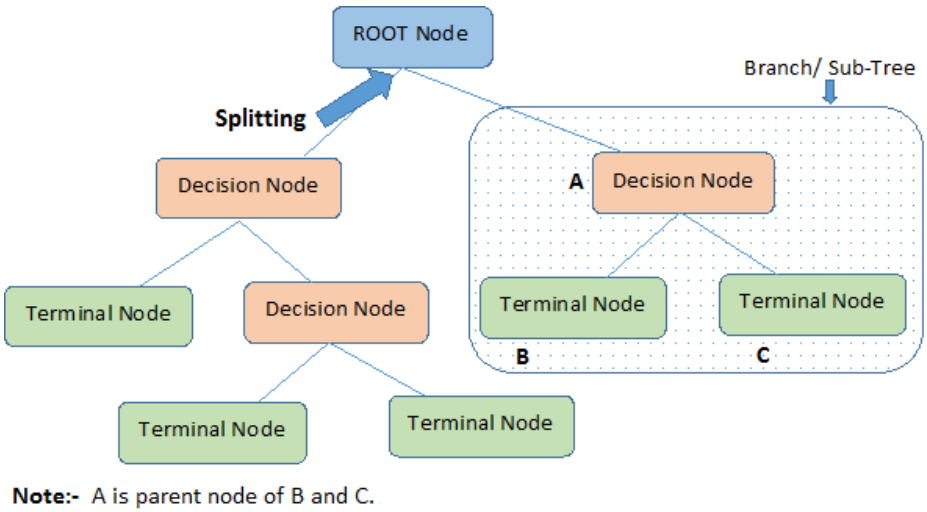
The outcome from Bayes Network Modelling upon output of a training structure, I was able to formulate a CPT for each node of every dependent attribute within the dataset. CPT which provides marginal probability of random values, this helps in the generation of a probabilistic reasoning and infernece of the simulated populations aka. Clinical avatars.

Categorical Data: multinomial distribution
Continuous Data: multivariate normal distribution
Mixed Data: conditional Gaussian distribution

(The full BNM: CPT Result is included in the submission folder)

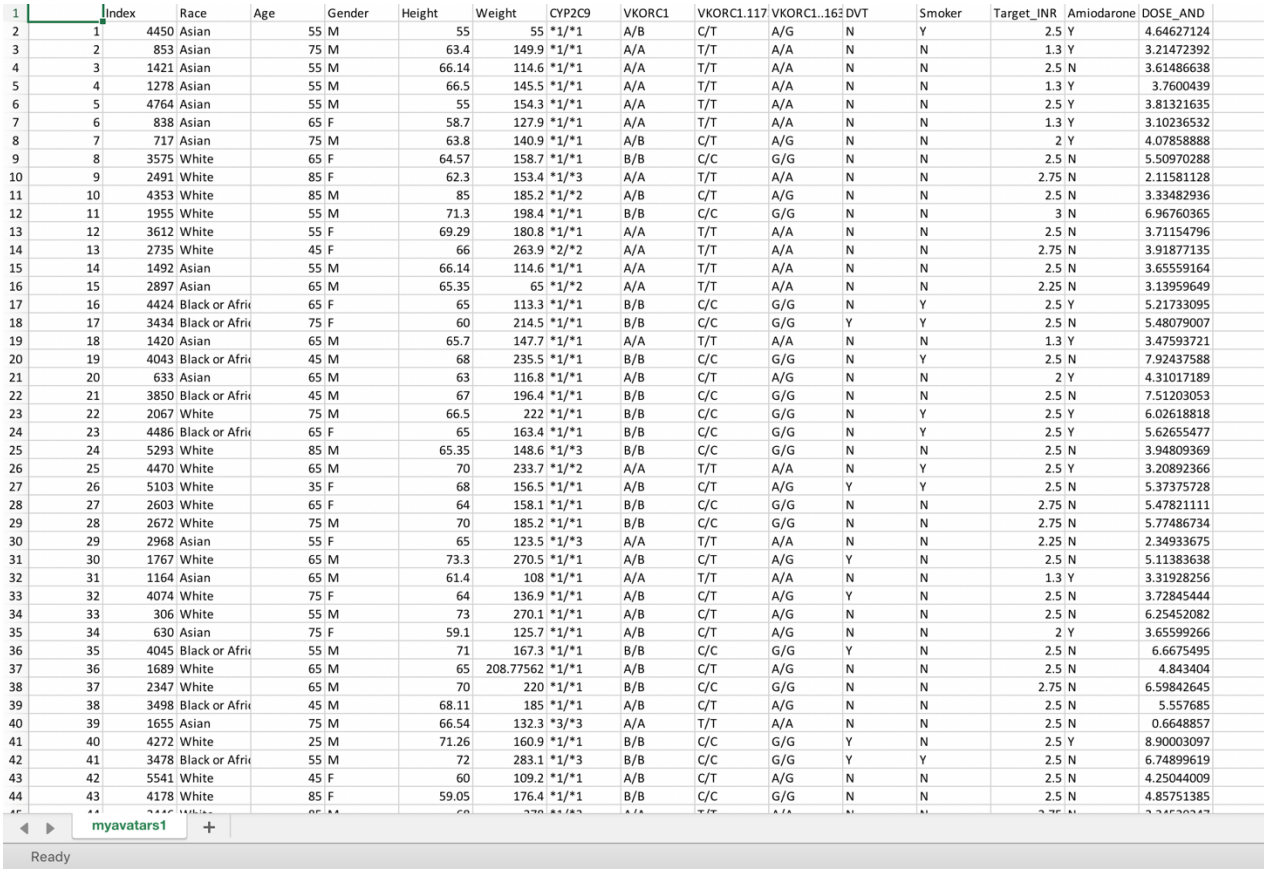
Clinical Avatars: CART Synthetic Data

Classification and Regression Trees (CART) Method



Note:- A is parent node of B and C.

I have used a combination of regression trees and classification trees given that I have a mixture between categorical and numerical data. In the regression tree, the value attained by the terminal nodes is the "mean" of the observations. In the classification tree, however, the value attained by the terminal node is the mode of the observations falling into that segment. As for the independent variables in the dataset, they are divided into distinct and non-overlapping regions. It keeps on splitting the nodes into sub-nosed until homogeneity of resultant sub-nodes is reached. Below the myavatars1.csv output of such process. I generated exactly 5700 synthetic observations more to verify the accuracy of the simulation.

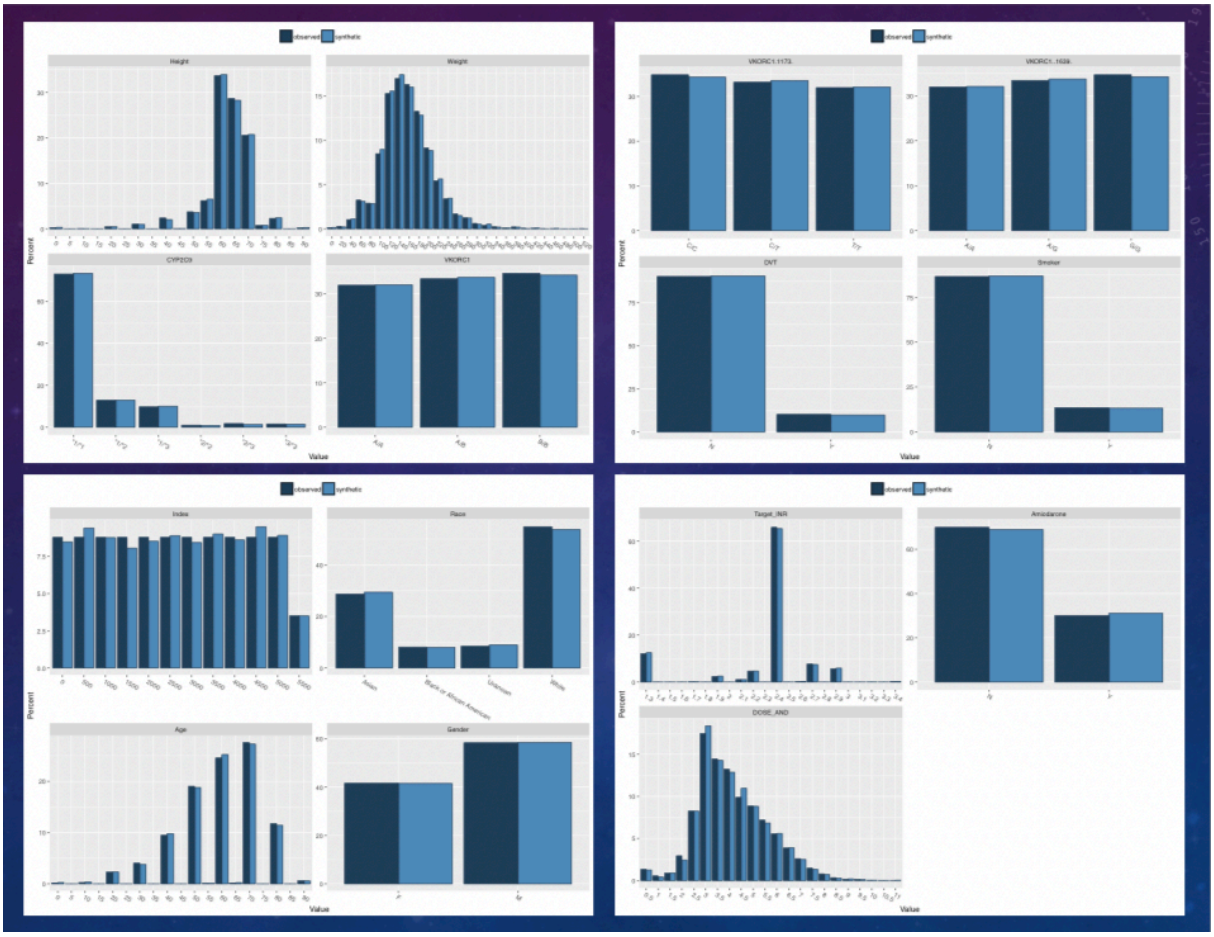


		Index	Race	Age	Gender	Height	Weight	CYP2C9	VKORC1	VKORC1.117	VKORC1..163	DVT	Smoker	Target_INR	Amiodarone	DOSE_AND
2	1	4450	Asian	55	M	55	55	*1/*1	A/B	C/T	A/G	N	Y	2.5	Y	4.64627124
3	2	853	Asian	75	M	63.4	149.9	*1/*1	A/A	T/T	A/A	N	N	1.3	Y	3.21472392
4	3	1421	Asian	55	M	66.14	114.6	*1/*1	A/A	T/T	A/A	N	N	2.5	N	3.61486638
5	4	1278	Asian	55	M	66.5	145.5	*1/*1	A/A	T/T	A/A	N	N	1.3	Y	3.7600439
6	5	4764	Asian	55	M	55	154.3	*1/*1	A/A	T/T	A/A	N	N	2.5	Y	3.81321635
7	6	838	Asian	65	F	58.7	127.9	*1/*1	A/A	T/T	A/A	N	N	1.3	Y	3.10236532
8	7	717	Asian	75	M	63.8	140.9	*1/*1	A/B	C/T	A/G	N	N	2	Y	4.07858888
9	8	3575	White	65	F	64.57	158.7	*1/*1	B/B	C/C	G/G	N	N	2.5	N	5.50970288
10	9	2491	White	85	F	62.3	153.4	*1/*3	A/A	T/T	A/A	N	N	2.75	N	2.11581128
11	10	4353	White	85	M	85	185.2	*1/*2	A/B	C/T	A/G	N	N	2.5	N	3.33482936
12	11	1955	White	55	M	71.3	198.4	*1/*1	B/B	C/C	G/G	N	N	3	N	6.96760365
13	12	3612	White	55	F	69.29	180.8	*1/*1	A/A	T/T	A/A	N	N	2.5	N	3.71154796
14	13	2735	White	45	F	66	263.9	*2/*2	A/A	T/T	A/A	N	N	2.75	N	3.91877135
15	14	1492	Asian	55	M	66.14	114.6	*1/*1	A/A	T/T	A/A	N	N	2.5	N	3.65559164
16	15	2897	Asian	65	M	65.35	65	*1/*2	A/A	T/T	A/A	N	N	2.25	N	3.13959649
17	16	4424	Black or Afric	65	F	65	113.3	*1/*1	B/B	C/C	G/G	N	Y	2.5	Y	5.21733095
18	17	3434	Black or Afric	75	F	60	214.5	*1/*1	B/B	C/C	G/G	Y	Y	2.5	N	5.48079007
19	18	1420	Asian	65	M	65.7	147.7	*1/*1	A/A	T/T	A/A	N	N	1.3	Y	3.47593721
20	19	4043	Black or Afric	45	M	68	235.5	*1/*1	B/B	C/C	G/G	N	Y	2.5	N	7.92437588
21	20	633	Asian	65	M	63	116.8	*1/*1	A/B	C/T	A/G	N	N	2	Y	4.31017189
22	21	3850	Black or Afric	45	M	67	196.4	*1/*1	B/B	C/C	G/G	N	N	2.5	N	7.51203053
23	22	2067	White	75	M	66.5	222	*1/*1	B/B	C/C	G/G	N	Y	2.5	Y	6.02618818
24	23	4486	Black or Afric	65	F	65	163.4	*1/*1	B/B	C/C	G/G	N	Y	2.5	Y	5.62655477
25	24	5293	White	85	M	65.35	148.6	*1/*3	B/B	C/C	G/G	N	N	2.5	N	3.94809369
26	25	4470	White	65	M	70	233.7	*1/*2	A/A	T/T	A/A	N	Y	2.5	Y	3.20892366
27	26	5103	White	35	F	68	156.5	*1/*1	A/B	C/T	A/G	Y	Y	2.5	N	5.37375728
28	27	2603	White	65	F	64	158.1	*1/*1	B/B	C/C	G/G	N	N	2.75	N	5.47821111
29	28	2672	White	75	M	70	185.2	*1/*1	B/B	C/C	G/G	N	N	2.75	N	5.77486734
30	29	2968	Asian	55	F	65	123.5	*1/*3	A/A	T/T	A/A	N	N	2.25	N	2.34933675
31	30	1767	White	65	M	73.3	270.5	*1/*1	A/B	C/T	A/G	Y	N	2.5	N	5.11383638
32	31	1164	Asian	65	M	61.4	108	*1/*1	A/A	T/T	A/A	N	N	1.3	Y	3.31928256
33	32	4074	White	75	F	64	136.9	*1/*1	A/B	C/T	A/G	Y	N	2.5	N	3.72845444
34	33	306	White	55	M	73	270.1	*1/*1	A/B	C/T	A/G	N	N	2.5	N	6.25452082
35	34	630	Asian	75	F	59.1	125.7	*1/*1	A/B	C/T	A/G	N	N	2	Y	3.65599266
36	35	4045	Black or Afric	55	M	71	167.3	*1/*1	B/B	C/C	G/G	Y	N	2.5	N	6.6675495
37	36	1689	White	65	M	65	208.77562	*1/*1	A/B	C/T	A/G	N	N	2.5	N	4.843404
38	37	2347	White	65	M	70	220	*1/*1	B/B	C/C	G/G	N	N	2.75	N	6.59842645
39	38	3498	Black or Afric	45	M	68.11	185	*1/*1	A/B	C/T	A/G	N	N	2.5	N	5.557685
40	39	1655	Asian	75	M	66.54	132.3	*3/*3	A/A	T/T	A/A	N	N	2.5	N	0.6648857
41	40	4272	White	25	M	71.26	160.9	*1/*1	B/B	C/C	G/G	Y	N	2.5	Y	8.90003097
42	41	3478	Black or Afric	55	M	72	283.1	*1/*3	B/B	C/C	G/G	Y	Y	2.5	N	6.74899619
43	42	5541	White	45	F	60	109.2	*1/*1	A/B	C/T	A/G	N	N	2.5	N	4.25044009
44	43	4178	White	85	F	59.05	176.4	*1/*1	B/B	C/C	G/G	N	N	2.5	N	4.85751385

myavatars1

Ready

Comparison between Synthetic Dataset and Observed Dataset -- Results



As we can notice, the CART was able to reach a high level of homogeneity in its partition to be able to mimic the original dataset. The probability outcome are listed below:

$Race

	Asian	Black or African American	Unknown	White
observed	28.66667	8.105263	8.456140	54.77193
synthetic	29.36842	7.982456	8.947368	53.70175

$Age

0 5	10 15	20 25	30 35	40 45	50	55	60

observed 0.1578947 0 0.2456140 0 2.280702 0 4.035088 0 9.473684 0 19.03509 0.0877193 24.57895

synthetic 0.2456140 0 0.3333333 0 2.298246 0 3.754386 0 9.807018 0 18.78947 0.0877193 25.22807

	65	70	75	80	85	90
observed	0.1578947	27.57895	0	11.75439	0	0.6140351
synthetic	0.2280702	27.26316	0	11.40351	0	0.5614035

$CYP2C9

	*1/*1	*1/*2	*1/*3	*2/*2	*2/*3	*3/*3
observed	73.07018	12.92982	9.719298	0.9824561	1.736842	1.561404
synthetic	73.45614	12.94737	10.035088	0.8245614	1.263158	1.473684

$VKORC1

	A/A	A/B	B/B
observed	31.94737	33.40351	34.64912
synthetic	32.05263	33.71930	34.22807

$VKORC1.1173.

	C/C	C/T	T/T
observed	34.84211	33.19298	31.96491
synthetic	34.31579	33.57895	32.10526

Randomization of subpopulations from Clinical Avatars to Protocols

They have used 5 models which are heavily relying on Hamberg2007.R that was originally developed in NONMEN. The clinical avatars are randomly distributed into these sub-models during the simulation, the weight outcome is then used to compute the TTR (Time in Therapeutic Range). The simulation requires heavy differential equations computing; thus, the partition and the simulation is of sequential manner.

DETAIL WORK POST-MIDTERM

PK/PD Mixed Effects Modelling of Observed Data & Clinical Avatars

Package used: nlmeODE

nlmeODE R package combines two distinct and useful packages of deSolve and nlme. deSolve package caters to "Ordinary Differential Equations" and nlme is simply for "Non-linear mixed-effects modelling". This package is commonly used for PKPD model analysis.

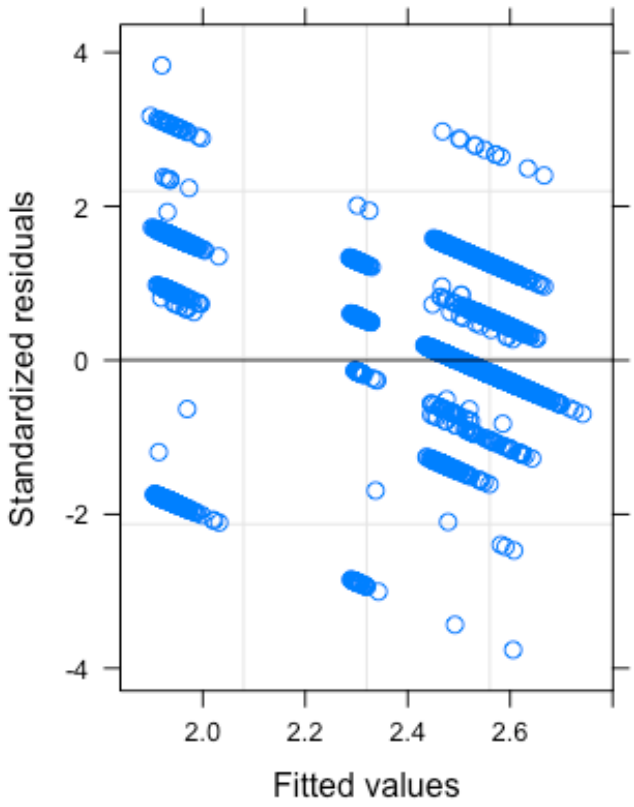
mlmeODE::Library PK/PD Models

In order to conduct a generalized analysis of PK/PD data is mainly concerned with identifying the relationship between the dosing regimen proposed and the body's exposure to the drug taken into account of the concentration time to determine a dose. Results portray the rate of distribution of the drug throughout the targeted compartment, in this project however, I focused on two-compartment model. Given the fact that there is a lack of time stamp and concentration level, I used the TARGET_INR, short for **"International Normalized Ratio"**, instead. INR is a calculation to standardize specific patient's "prothrombin" time against the normal mean prothrombin time of an observed population. **Prothrombin time** is a common test to unravel how quick the blood clots in patients receiving oral anti-coagulant drug. I treated INR as an indicator of time and concentration, though, this is not an optimal solution due to lack of the needed variables. If INR is too low, blood clots cannot be prevented which means "absorption level" by nearby tissues will be very slow due to clogging in veins or direct tissues. When the INR is too high, there is a risk of bleeding, which also mean that the circulation of blood is too "fast" and can reach "over-absorption". The analysis yielded a two-compartmental model of how the warfarin drug "moves" through the body. Basic assumptions that were considered during this analysis were the fact that there must be a "effect site" that is a two-compartment tissue and its direct neighbors. The "dose" variable portrays information regarding the magnitude of a possible "response" and "toxicity". Another assumption is that the drug cannot be "injected" directly to the targeted effect site but instead "move" towards the compartment. In this project, it is clearly an "oral" and dissolves towards the patient's capillaries (i.e tiny vessels). A definitive fact that we need to consider in this analysis is that the process must vary across all subjects due to its **inter-subject variation**. This is one of the fundamental reasons why within the Pharmacodynamic analysis, mixed-effect modelling with random effects can help shed light upon on inter-subject variation and estimation and characterize of dosage regimen within a "therapeutic window". The data used in this analysis is more appropriate for a "population" model, a linear mixed-effects model was implemented. This way we can achieve a balance between the fixed-effects anticipated outcome of population-average variability and random effects anticipated outcome of subject-specific variability.

2-compartment model generated as follows: *(WAR short for WARFARIN)*

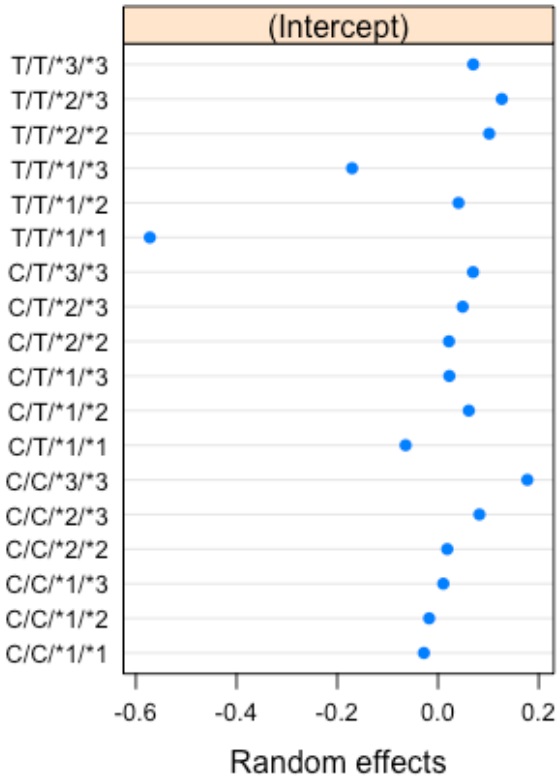
```
TwoComp <- list(DiffEq = list(dy1dt = ~ -(k12 +k10)*y1+k21*y2 ,
                              dy2dt = ~ -k21*y2 + k12*y1),
                ObsEq = list(
                  c1 = ~ y1,
                  c2 = ~ 0),
                  States = c("y1", "y2"),
                  Parms = c("k12", "k21", "k10", "start"),
                  Init = list("start", 0))
War2Model <- nlmeODE(TwoComp, WarData)
```

Implementation of Mixed-effects Modelling for PD Analysis



Linear mixed model was used instead of a non-linear mixed model to the characteristics of the dataset that lacks several vital variables for a non-linear mechanism to execute properly. Linear model, on the other hand, gives more parsimonious description of the data with linear parameters. It does, nevertheless, describe how a response variable varies with covariates. The empirical models can still observe relationship while excluding the mechanism producing the data underneath. This is arguable given that this is the raw data that has been extracted directly from the EMR dataset with the ETL Process.

Random Intercepts interpretation



To check how much better out-fit is compared to a fit that ignores individual effects

```
Linear mixed-effects model fit by REML
 Data: WarData
       AIC      BIC    logLik
  4194.788 4228.027 -2092.394

Random effects:
 Formula: ~1 | VKORC1.1173.
        (Intercept)
StdDev: 0.0002502368

 Formula: ~1 | CYP2C9 %in% VKORC1.1173.
        (Intercept)  Residual
StdDev:   0.1713266 0.3472078

Fixed effects: Target_INR ~ Dose
               Value  Std.Error   DF  t-value p-value
(Intercept) 2.3945539 0.04667223 5681 51.30576       0
Dose        0.0336352 0.00592237 5681  5.67934       0
 Correlation:
     (Intr)
Dose -0.406

Standardized Within-Group Residuals:
        Min          Q1         Med          Q3         Max
-3.76096009 -0.25125541 -0.06546283  0.52573712  3.82994045

Number of Observations: 5700
Number of Groups:
            VKORC1.1173. CYP2C9 %in% VKORC1.1173.
                       3                       18
```

With	(Intercept)	Dose
C/C/*1/*1	2.366588	0.03363518
C/C/*1/*2	2.377025	0.03363518
C/C/*1/*3	2.405128	0.03363518
C/C/*2/*2	2.412952	0.03363518
C/C/*2/*3	2.476842	0.03363518
C/C/*3/*3	2.572076	0.03363518
C/T/*1/*1	2.330310	0.03363518
C/T/*1/*2	2.455848	0.03363518
C/T/*1/*3	2.417152	0.03363518
C/T/*2/*2	2.416393	0.03363518

C/T/*2/*3 2.443610 0.03363518

C/T/*3/*3 2.464221 0.03363518

T/T/*1/*1 1.822302 0.03363518

T/T/*1/*2 2.435366 0.03363518

T/T/*1/*3 2.223988 0.03363518

T/T/*2/*2 2.496293 0.03363518

T/T/*2/*3 2.521330 0.03363518

T/T/*3/*3 2.464546 0.03363518

The result from AOV function which pander to "fit an analysis of variance model" is the following:

Df Sum Sq Mean Sq F value Pr(>F)

Dose 1 102.1 102.11 589.9 <2e-16 ***

Residuals 5698 986.3 0.17

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

CONCLUSION

The dataset that has been provided is more inclined towards the clinical avatar formation using Bayesian Network Modelling than the PK/PD modelling track. Mixed effects modelling for PK-PD Analysis was implemented, thus, providing with estimates of dosage regimen in correlation to the genetic relationship between **CYP2C9 and VKORC1.1173.** The random effect model was able to give all possibilities of a plausible inter-subject variation of dosage vs. INR. It will be much more informative and stronger if the dataset included time stamps, concentration levels, and specific dosage from the supplementary drugs of Amiodarone and Fluvastatin drug intake.

REFERENCE

https://cran.r-project.org/web/packages/bnlearn/index.html

https://ermongroup.github.io/cs228-notes/learning/structLearn/

https://cran.r-project.org/web/packages/nlmeODE/nlmeODE.pdf

https://cran.r-project.org/web/views/Pharmacokinetics.html

https://www4.stat.ncsu.edu/~davidian/webinar.pdf

http://www.heart.org/HEARTORG/Conditions/Arrhythmia/PreventionTreatmentofArrhythmia/A-Patients-Guide-to-Taking-Warfarin_UCM_444996_Article.jsp?appName=WebApp#.WvS5idMvxiI

http://holford.fmhs.auckland.ac.nz/teaching/medsci719/

https://onlinelibrary.wiley.com/doi/10.1002/9781118784860.ch1

https://stats.stackexchange.com/questions/9759/can-someone-shed-light-on-linear-vs-nonlinear-mixed-effects

https://www.ndm.ox.ac.uk/principal-investigators/researcher/joel-tarning

APPENDIX

```
#Midterm due April 2nd

#Finals due May 11
################################
data1 <-read.csv("/Users/ladykat/Downloads/PharmGKB_KNN2.csv") #data pre-processing (technically correct data, consistent data) install.packages("dplyr")
library(dplyr)
str(data1)
summary(data1)
glimpse(data1)
dim(data1)
#ANS: 5700 15

names(data1)
```

```
#Baysian network
install.packages("bnlearn")
library(bnlearn)
#remove index -> integer type
bnvars <- c ("Race", "Gender", "Height", "W eight", "CYP2C9", "VKORC1", "VKORC1.1173.", "VKORC1..1639.", "DVT",
"Smoker", "Age", "Target_INR", "DOSE_AND", "Amiodarone")

bndata <- data1[bnvars] bn_df <- data.frame(bndata) res <- hc(bn_df)
plot(res)

#Training
fittedbn <- bn.fit (res, data = bn_df)
print(fittedbn$Age)
print(fittedbn)
#some other variable such as other genotypes
#Infering
head(fittedbn)
#cpquery(fittedbn, event = (Age == "<55.0"), evidence = (Target_INR == "<2.5")) #Evaluating

######
#Subset Data to NUM vars: AGE, TARGET_INR, DOSE_AND, INDEX

myvars <- c("Age", "Target_INR", "VKORC1..1639.", "CYP2C9" ) numdata <- data1[myvars]
head(numdata)

#Age, CYP2C9, VKORC1 most important variables names(data1)

vars <- c("VKORC1..1639.", "CYP2C9", "Age", "Target_INR")

subdata1 <- data1[vars] head(subdata1)
#Target INR: 2.5

#Let's do bayesian network for these important variables

bndata1 <- data1[vars]
bn_df1 <- data.frame(bndata1) res1 <- hc(bn_df1)
plot(res1)

#Training
fittedbn1 <- bn.fit(res1, data = bn_df1) print(fittedbn1)

################################### #Synthetic Data out of the observed

install.packages("synthpop") library(synthpop)
#reproduce the data1
my.seed <- 123456780
myavatars <- syn(data1, seed = my.seed) names(myavatars)

str(myavatars)
print(myavatars)
myavatars$syn
```

```
#lets compare original vs synthesised data
compare(myavatars, data1, vars = c("Age", "Smoker", "Race")) compare(myavatars,data1)

#export synth myavatars dataset write.csv(myavatars$syn, "myavatars1.csv")

source("anticoagulation_therapy_simulator.R")
```

```
#Read in the avatars table
avatars = read.table("avatars", sep="\t", header=T)

avatars = data1

#Distribute the job
av_per = 5 # Number of avatars per file (per run) block = av_per - 1

av_index = 1

to = av_index*av_per from = to - block

av_sub = avatars[from:to,]

#Create random seeds
# need to create a global set of random seeds so when the jobs are distributed the values are random and
# don't repeat with every block
set.seed(43210)
numReplicates = 1
randomValues = array(round(abs(rnorm(nrow(avatars)*numReplicates)*nrow(avatars)*numReplicates)),
dim=c(nrow(avatars), numReplicates))
rand_sub = randomValues[from:to,]

#Run function anticoagulation_therapy_simulator
av_out = processAvatar(avatars=av_sub, protocol="coumagen_pharm", initialDose=16,

numDaysToSimulate=22, maxDose=15, numReplicates=numReplicates, maxTime=24, rseed=rand_sub)
```

```
install.packages("nlmeODE")

install.packages("deSolve")

install.packages("nlme")

install.packages("lattice")

library(PKPDmodels)

library(nlmeODE)
```

```
library(lattice)

library(nlme)

library(deSolve)

??nlmeODE

##MIXED-EFFECTS MODELING USING DE.

data(Theoph)

head(Theoph)

WarData<- read.csv("/Users/ladykat/Downloads/CBMI\ Training\ Materials/PharmGKB_KNN.csv")

head(WarData)



WarODE <- WarData

WarODE$Dose[WarODE$INR!=0] <- 0

WarODE$Cmt <- rep(1, dim(WarODE)[1])

########################################

#Pharmacokinetics

#One-compartment Model

########################################

OneComp <- list(

  DiffEq =list(


    dy1dt = ~ -ka*y1,

    dy2dt = ~ ka*y1 - ke*y2),

    ObsEq = list(
```

```
    c1 = ~ 0,
    c2 = ~ y2/CL*ke),

    Parms = c("ka", "ke", "CL"),
    States = c("y1", "y2"),
    Init = list(0,0)
  )
#Using Library NlmODE
War1Model <- nlmeODE(OneComp, WarODE)
###################################
#PK : 2-compartment model
###################################
TwoComp <- list(DiffEq = list(dy1dt = ~ -(k12 +k10)*y1+k21*y2 ,
                        dy2dt = ~ -k21*y2 + k12*y1),
             ObsEq = list(
               c1 = ~ y1,
               c2 = ~ 0),
               States = c("y1", "y2"),
               Parms = c("k12", "k21", "k10", "start"),
               Init = list("start", 0))

War2Model <- nlmeODE(TwoComp, WarData)

##############################
```

```
## PK MODEL VISUALIZATION ##

############################

#drug dosing is usually adapted to patient's weight.

#The blood concentration depends on the liberation, absorption,

#distribution, metabolism and the excretion of the drug.

install.packages("PK")


##no time/concentration though!

#################################################################

##ABSORPTION MODEL WITH ESTIMATION of TIME/RATE of INFUSION

#################################################################


OneCompAbs <- list(DiffEq = list(dA1dt = ~ -ka*A1,

                                 dA2dt = ~ ka*A1- CL/V1*A2),

                   ObsEq = list(

                     SC = ~ 0,

                     C = ~ A2/V1),

                   States =c("A1", "A2"),

                   Parms =c("ka", "CL", "V1", "F1"),

                   Init = list(0,0))


ID <- rep(seq(1:18), each = 11)

Time <- rep(seq(0,100, by = 10), 18)

Dose  <- c(rep(c(100,0,0,100,0,0,0,0,0,0,0),6),rep(c(100,0,0,0,0,0,0,100,0,0,0),6),

           rep(c(100,0,0,0,0,0,0,0,0,0,0),6))
```

```
Rate  <- c(rep(rep(0,11),6),rep(c(5,rep(0,10)),6),rep(rep(0,11),6))

#Cmt = ?

Cmt   <- c(rep(1,6*11),rep(c(2,0,0,0,0,0,0,1,0,0,0),6),rep(2,6*11))

#Concentration

Conc  <- rep(0,18*11)

Data <- as.data.frame(list(ID=ID,Time=Time,Dose=Dose,Rate=Rate,Cmt=Cmt,Conc=Conc))


#SimData <- groupedData(Conc ~ Time | ID, data = WarData, labels = list(x = "Time", y = "Concentration"))


#head(SimData)


OneCompAbsModel <- nlmeODE(OneCompAbs, WarData)

#kaSim <- rep(log(rep(0.05, 18)) + 0.3*rnorm(18), each = 11)

#CLSim <- rep(log(rep(0.5, 18)) + 0.2*rnorm(18), each = 11)

#V1Sim <- rep(log(rep(10,18)) + 0.1*rnorm(18), each = 11)

#F1Sim <- rep(log(0.8), 18*11)

#SimData$Sim <- OneCompAbsModel(kaSim, CLSim, V1Sim, F1Sim, SimData$Time, SimData$ID)

#SimData$Conc <- SimData$Sim + 0.3*rnorm(dim(SimData)[1])


#Data <- groupedData( Conc ~ Time | ID, data = SimData, labels = list(x = "Time" , y = "Concentration"))


plot (WarData, aspect = 1/1)

##################################

##Estimation of Model Parameters

##################################
```

```
#OneCompAbsModel <- nlmeODE(OneCompAbs, WarData)

############################################################

## Simulation and Simultaneous Estimation of PK/PD Data

############################################################


PoolModel <- list(

 DiffEq=list(

   dy1dt = ~ -ke*y1,

   dy2dt = ~ krel * (1-Emax*(y1/Vd)**gamma/(EC50**gamma+(y1/Vd)**gamma)) * y3 - kout * y2,

   dy3dt = ~ Kin - krel * (1-Emax*(y1/Vd)**gamma/(EC50**gamma+(y1/Vd)**gamma))*y3),

 ObsEq=list(

   PK   = ~ y1/Vd,

   PD   = ~ y2,

   Pool = ~ 0),

 States=c("y1","y2","y3"),

 Parms=c("ke","Vd","Kin","kout","krel","Emax","EC50","gamma"),

 Init=list(0,"Kin/kout","Kin/krel"))


#ID    <- rep(seq(1:12),each=2*12)

#Time  <- rep(rep(c(0,0.25,0.5,0.75,1,2,4,6,8,10,12,24),each=2),12)

#Dose  <- rep(c(100,rep(0,23)),12)

#Cmt   <- rep(rep(c(1,2),12),12)

#Type  <- rep(rep(c(1,2),12),12)

#Conc  <- rep(0,2*12*12)

#Data <- as.data.frame(list(ID=ID,Time=Time,Dose=Dose,Cmt=Cmt,Type=Type,Conc=Conc))
```

```
#SimData <- groupedData( Conc ~ Time | ID/Type,
#                   data = WarData,
#                   labels = list( x = "INR", y = "Dose"))
PKPDpoolModel <- nlmeODE(PoolModel,WarData)

#keSim    <- rep(log(rep(0.05,12))+0.1*rnorm(12),each=2*12)
#VdSim    <- rep(log(rep(10,12))+0.01*rnorm(12),each=2*12)
#EC50Sim  <- rep(log(rep(5,12))+0.1*rnorm(12),each=2*12)
#KinSim   <- rep(log(5),2*12*12)
#koutSim  <- rep(log(0.5),2*12*12)
#krelSim  <- rep(log(2),2*12*12)
#EmaxSim  <- rep(log(1),2*12*12)
#gammaSim <- rep(log(3),2*12*12)

#SimData$Sim <- PKPDpoolModel(keSim,VdSim,KinSim,koutSim,krelSim,EmaxSim,EC50Sim,
#                     gammaSim,SimData$Time,SimData$ID,SimData$Type)
#SimData$Conc[SimData$Type==1] <- SimData$Sim[SimData$Type==1]*(1 +
#                                          0.1*rnorm(length(SimData[SimData$Type==1,1])))
#SimData$Conc[SimData$Type==2] <- SimData$Sim[SimData$Type==2]*(1 +
#                                          0.05*rnorm(length(SimData[SimData$Type==2,1])))

#Data <- groupedData( Conc ~ Time | ID/Type,
#                data = SimData,
#                labels = list( x = "Time", y = "Concentration"))
plot(WarData,display=1,aspect=1/1)
```

```
#Fixed parameters

Data$Emax <-  rep(log(1),dim(Data)[1])


#Estimation of model parameters

PKPDpoolModel <- nlmeODE(PoolModel,Data)

########################################

########################################

#MIXED-EFFECTS MODELLING

library(nlme)

summary(WarData)

str(WarData)

#Wt,Age, Target_INR, Dose

plot(WarData)

#Model1 = LM

model1 <- lm(Target_INR ~ Dose , data = WarData)

summary(model1)

plot(model1)

coef(model1)

#Model2 = Mixed-Effects

model2= lme(Target_INR~ Dose, data = WarData, random = ~1 | VKORC1.1173./CYP2C9)

summary(model2)

coef(model2)

plot(ranef(model2))

plot(model2)

m2 <- aov(Target_INR~ Dose, data = WarData)
```

```
summary(m2)
```