Bayesian Approaches to Therapeutic Drug Monitoring

Wenqi YOU
LSI, I&C, EPFL

Abstract—Innovations in drug regimens and therapeutic protocols are improving the survival rate and life quality of patients affected by serious, life-threatening conditions. Such treatments require daily drug administration over long-term periods. Therapeutic Drug Monitoring (TDM) guided dose individualization according to patients' inter-differences has been recognized to be safe and beneficial in these treatments. Applying Bayesian methods to TDM is one of the monitoring methods and has shown to have good predictive performance using minimal data. It helps to control the drug concentration within the effective but nontoxic level, or therapeutic range. A clinical example of dose reduction guided by TDM shows one of the potential applications. Some other general mathematical models will also be introduced in this proposal.

Index Terms—Bayesian Network, Therapeutic Drug Monitoring (TDM), Dose Individualization or Personalized Medicine, Mathematical Models.

I. INTRODUCTION

THERAPEUTIC DRUG MONITORING (TDM) is a measurement made in a laboratory of a parameter which, with appropriate interpretation, will directly influence prescribing procedures[4]. It plays a far greater role than just therapeutic drug measuring in that its central goal is to understand the regulation of a certain drug influenced by the inter- and intra-personal differences, so that a personalized medicine guidance could be derived to help maintaining the drug concentration at an effective but nontoxic level (therapeutic range). More emphasis has been casted on TDM of some serious and life-threatening situations such as HIV, cancer, organ transplantation, etc, since those diseases are subject to variability and require careful monitoring. Hence it may be a useful tool for optimizing individuals’ drug exposure and improving treatment responses[8]. Studies have revealed a tight relationship between the drug concentrations and viral response, which indicates the need for a fine grained control over the drugs with a narrow therapeutic range.

In pharmacology, nowadays, many medicines are assigned to the patients according to the illness and monitored empirically without a scientific way to measure the drug concentration in the blood. However there is a small group of medicines, such as efavirenz (EFV) for HIV disease[2], which effective therapeutic range is narrow and hence is easy to be under- or overdosed. Being under-dosed will cause ineffective treatment, while being overdosed will lead to toxicity and tissue damage. Therefore, as one of the sub-projects in Nano-Tera, ISyPeM, Intelligent Integrated Systems for Personalized Medicine, aims at offering advanced technologies to healthcare providers and patients for:

- assessing drug exposure and response by measuring drug concentrations or biomarkers with miniaturized sensors;
- providing drug treatment indications to prescribers and/or patients, based on appropriate processing of statistical and personal data, to translate measurement results into therapeutic recommendations and decisions;
- enabling a seamless connection of novel drug delivery techniques with monitoring in some suitable situations.

The purpose of the project is also to advance the state-of-the-art in personalized medicine by creating new enabling technologies for drug monitoring and control mechanisms, while reducing the health care costs.

Monitoring the drug concentration and giving the dosage correspondingly is one of the solutions to control over the effects of this small group of medicines. But is is often inconvenient to the patients, and it is also expensive to take blood samples and do measurements during the treatment. Many examples from the literature use Bayesian approaches as an assisting tool[5][6][8]. Based on some prior knowledge of the patients, the method is able to learn, to predict and
to optimize the dosage to have a sufficient but nontoxic concentration level. Bayesian approaches are not the only method to solve this problem. In [3], Stevens et. al. give some other possible models to do the analysis.

In this paper, we are going to mainly give an introduction to Bayesian networks and its potential application to therapeutic drug monitoring. A clinical experiment of dose individualization (e.g. dose reduction) led by TDM will also be introduced. At last, a general mathematical model will also be discussed briefly.

II. SURVEY OF THE SELECTED PAPERS

Dose-individualization has been studied for a long time and there are 3 major methods used to target this problem: (1) population-based methods, e.g. nomograms, (2) pharmacokinetic methods based on linear regression analysis, (3) Bayesian estimation procedures [6] Methods (2) and (3) are superior to (1) [9], while the Bayesian approach is recommended over (2) due to its greater effectiveness [6].

In this section, Bayesian networks will be analyzed and their application in TDM-guided dose reduction for the medicine efavirenz will be also discussed. Finally, a general mathematical model will be introduced.

A. Bayesian Network in Analyzing Proteomic Data

A Bayesian network is a probabilistic graphical model that explores the relationships among a set of random variables. It is a directed acyclic graph (DAG) which is comparatively robust to reconstruct and represent the uncertainty of unobserved events [1]. The key property of Bayesian network is that it can use simplified structure in the joint distribution to represent a set of high-dimensional data. In the graph of a Bayesian network, the nodes represent the random variables of the data and the edges, also referred as paths, represent the influence of one datum on another. In a directed graph, each edge is oriented, e.g. \( X \rightarrow Y \) means an edge from \( X \) to \( Y \). According to the Markov assumption, each node, or variable \( X_i \), only depends on its direct parents \( Pa_i \) in the graph; hence, the relationship between the node and its parents' parents' is called conditional independence given its direct parents' information, as shown in equation (1).

\[
P(X_i|Pa_i, Pa_Pa_i) = P(X_i|Pa_i),
\]

where \( Pa_Pa_i \) denotes the parents of parents of the node.

Furthermore, as claimed in the chain rule of probabilities, any joint distribution can be expressed as a product of conditional probabilities as follows:

\[
P(X_1, \ldots , X_n) = \prod_{i=1}^{n} P(X_i|X_1, \ldots , X_{i-1}).
\]

(2)

Considering the equation (1), (2) can be rewritten as:

\[
P(X_1, \ldots , X_n) = \prod_{i=1}^{n} P(X_i|Pa_i),
\]

(3)

which obviously saves data entries compared with equation (2). Assuming the data are all binary variables and the maximal number of parents in the graph is bounded by \( k \), to represent the joint distribution using equation (2) requires \( 2^k - 1 \) independent parameters, but \( 2^{kn} \) parameters with the equation (3). If \( k \) is not too big, (3) surpasses the (2) in analyzing data with a large number of random variables.

In order to use equation (3) to draw a joint distribution of the given data, we need to obtain the conditional probability distribution \( P(X_i|Pa_i) \) for each variable \( X_i \), and \( P(X_i|Pa_i) \) can be viewed as a probabilistic function of \( X_i \) with the inputs as \( Pa_i \), \( X_i \)'s parents. Hence, in order to be efficient, it is critical to build up the Bayesian network to know the dependencies among the variables.

To draw the network, we could first learn from the given data whether every two random variables are independent or not. Then a Partially Directed Acyclic Graph (PDAG), a graph containing both directed and undirected edges, could be derived considering two structures:

- \( X \rightarrow Y \rightarrow Z \), where \( X \) and \( Z \) are conditionally independent given the information of \( Y \).
- \( X \rightarrow Y \leftarrow Z \), where \( X \) and \( Z \) are dependent given the information of \( Y \).

Afterwards, together with the knowledge that there is no cycles in the graph, the direction of some of the remaining undirected edges could also be determined.

The goal of the author [1] is to automatically find out the structure of the molecular signaling network from a large data set of phosphorylated protein activity levels; that is, to learn a Bayesian network which could give a high probability to the obtained samples data. Since nevertheless there will be some noise in the data, it is rarely possible to recover a true network \( G^* \). Instead, we try to recover a \( G \) that is equivalent to \( G^* \) or has the same independence statements as \( G^* \). In the first step, the author [1] assumes that the graph structure \( G \) is given, and then determines the conditional probability distribution parameters \( \theta \) as well as a score. Then, by scoring the graph structure, a high-scoring network will be discovered.

To do this, the method of Maximum Likelihood Estimation of Parameters is used based on the assumption that the graph \( G \) is given. Hence the task becomes to find the conditional probabilities \( \theta \) that maximize the likelihood that the obtained data \( D \) was generated from the Bayesian network \( B^* = (G, \theta) \). Since we assume that the samples are independent, the calculation of the likelihood is simply

\[
L(\theta : D) = \prod_{m=1}^{M} P(X[m]|\theta),
\]

(4)

where \( X[m] \) denotes sample \( m \). As discussed before, each variable only depends on its direct parents, thus equation (4) can be rewritten as:

\[
L(\theta : D) = \prod_{m=1}^{M} \prod_{i=1}^{N} P(X_i[m]|Pa_i[m], \theta),
\]

(5)

where \( X_i[m] \) denotes variable \( X_i \) of sample \( m \). Exchanging the order of the product, the calculation using the local
likelihood function can be rewritten as follows:

\[ L(\theta : D) = \prod_{i=1}^{N} \left( \prod_{m=1}^{M} P(X_i[m] | Pa_i[m], \theta) \right) \]

where \( L_i \) is called the local likelihood of variable \( X_i \). Furthermore, in order to overcome one of this method’s drawbacks, namely the overfitting problem, the Bayesian approach is used to add a prior belief on the parameters \( \theta \), i.e., an initial distribution \( P(\theta) \), in a principled manner. In this paper, the author[1] uses the Dirichlet priors for the multinomial distributions. This prior concept could ensure that every event has some nonzero probability, and makes use of all the available prior knowledge on the problem. The formula \( P(\theta | D) = \frac{P(D | \theta) P(\theta)}{P(D)} \) relates the prior belief and the observations.

However, the discussion above is based on a known structure of the graph \( G \). In reality, most of the time, we need to reconstruct \( G \) from the observed data. Problem is that it is hard to distinguish between equivalent graphs from the observations. Hence, a score-based approach is used in this article to solve this problem, and a probability distribution is casted whenever there is any uncertainty:

\[ \text{Score}_B(G : D) = \log P(D | G) + \log P(G) \]

This score could be decomposed into local contributions of each variable, called ‘FamScore’ to simplify the computation.

\[ \text{Score}_B(G : D) = \sum_i \text{FamScore}_B(X_i, Pa_i : D) \]

A desired property for this scoring strategy is that: given a sufficient set of samples, graph structures capturing all dependencies in the distribution will obtain a higher score than any other graph. However to find the structure \( G \) that maximizes the scores is known to cost exponentially to the number of variables, so a search space and a set of operators have been defined. At each step, one of the following 3 operations will be carried out to the local nodes’ structure:

- add an edge
- remove an edge
- reverse an edge

Hence, the decomposition can simplify the re-evaluations only on the local variables that are involved in an edge operation. Together with the acyclic graph and/or maximal node’s degree property, the algorithm would return a local maximum graph in scoring. But in practice, the local maximum drawback could be overcome by combining other heuristics e.g. random restarts.

Choosing the network graph with the highest score might be influenced by spurious artifacts, especially when there are several graphs with close scores. ‘Model Averaging’ method is proposed to handle this problem, where a set of high-scoring graphs are chosen via a bootstrap method[7] and averaged on their structures by finding the common features. This method works well because it deals with the effect of small perturbations to the data on the learning process. Simulation results show that features induced with high confidence are rarely false positives, even in cases where the data sets are small relative to the size of the system being learnt.

Usually the data observed in biological tests appear to be independent variables, so it is important to find out the relationship between/among them, so as to be more efficient.

A causal network is a model providing the causality in the system, for example protein \( X \) activates protein \( Y \). It is mathematically represented similarly to a Bayesian network, a DAG where each node represents a random variable along with a local probability model, but a causal network has a stricter interpretation on the meaning of the edges: The parents of a variable are its immediate causes.

However, in a Bayesian network, \( X \rightarrow Y \) is equivalent to \( X \leftarrow Y \) and cannot be distinguished on the basis of the observed data alone. In a causal network, if \( X \rightarrow Y \), the modification to \( X \) will influence the value of \( Y \). Inversely, the changes on \( Y \) will not affect \( X \). Modeling the interventions to the PDAG is a common way to find out the causal relationship of an undirected edge. It is assumed that the intervention only affects \( X \)'s causal mechanism and leaves intact all other causal mechanisms in the models. Differently from finding the Bayesian network with scoring, here scoring with interventions adapts the score without counting the intervened variables.

This paper gives in detail a way to build up Bayesian network with the observed data. Directed Acyclic Graph is assumed to be the network’s property and the input data is proteomic data. In our work, the Bayesian approach will be used in Therapeutic Drug Monitoring, where we are going to use the clinical data including individual patient’s information (age, gender, weight, etc) and the data for some monitored drug (dosage per day, intervals between two doses, etc). The causal relationship behind the data is acyclic but could be with any other unobserved parameter. Hence we choose to use Bayesian network to illustrate our case. To properly deal with the dose-individualization, the network should easily adapt itself to both the mass and specific observations.

### B. TDM-Guided Efavirenz Dose Reduction

Efavirenz (EFV) is used for the treatment of HIV (Human Immunodeficiency Virus) type 1. In [2], a simplified algorithm based on a Bayesian pharmacokinetic approach is applied to guide the dose reduction in patients with EFV concentration above 75% (P75) with documented virological efficacy. [2]

Since nowadays the dose recommendations of most of the medicines are often derived from clinical trials that were not designed to a specific patient, minimum effective dose can vary with inter-individual differences of patients. Many works from the literature have demonstrated that a high or low efavirenz concentration is responsible for for the clinical toxicity or decreased efficacy in long-term users, thus making therapeutic drug monitoring (TDM) a necessary support to help decide the appropriate dosage.

In the paper, the author[2] carried out a study among 13 patients whose EFV concentration is higher than the recommended interval. After the dosage adjustment over one to two dose-reduction cycles, all of them reaches targets for EFV plasma concentration. Based on a Bayesian population
pharmacokinetic model developed in the same group[10], patients treated by a stable EFV 600 mg QD based regimen and presenting with an increasing plasma concentration, were selected for the dose-reduction study.

The target therapeutic range of EFV plasma concentration is 1000 to 4000ng/ml and all the patients observed had a concentration over 4000ng/ml. The approach evaluates the most likely contribution of intra- and inter-individual variability in EFV kinetics, and adjusts the dosage according to inter-individual variability. In details, patients with EFV concentration between P75 and P95 received EFV 400 mg QD, taking 200 mg twice a day; and those with concentration above P95 received EFV 200 mg QD, one tablet. The plasma concentration checks were then done on the 6th week to decide according to the results whether a second cycle of dose reduction was needed or not. On the 10th and 24th week, the EFV plasma concentrations were checked again for all patients.

During the dose adaptation stage, screening and baseline samples were drawn between 9:25h and 22:3h after last dose intake. On baseline, 5 patients out of 13 were qualifying for a 400mg dose and the other 8 underwent a dose reduction to 200mg. After the dose reduction, EFV drug concentrations kept above the minimum threshold of 1000ng/ml in all patients, and below the peak threshold of 4000ng/ml on the 10th week.

Another drug dosage reduction method, based on genotype, is compared in the paper with nearly similar precision using genetic testing, although no full genetic characterization could be provided for the patients. Three out of 13 surveyed patients have a different prediction via two methods, but the results might be similar (both would help the patients to reach a safe EFV plasma concentration) after 24 weeks.

The results of the questionnaires on other symptoms such as depression, anxiety and stress did vary before and after the study. Besides, the sleep quality was also taken into measurement via Groningen Sleep Quality Score (GSQS)[11]. During the 24 weeks’ study, no change in sleep quality or length was identified. Although neuropsychological toxicity might persist longer because of the long term therapy with less dosage, it could reduce the toxicity symptoms in the patients.

Several limitations to the study were pointed out:

• General conclusions could not be drawn due to the limited sample size.
• No randomized test was available to make a comparison.
• Only two cycles of dose reduction were monitored.

As a clinical trial, there are often many limitations such as the number of technical subjects, the environments, the methods, etc. Besides, the author[2] also observed some non-linearity in EFV clearance, which could be considered to adjust their model of Bayesian approach.

In this paper, a clinical dose-individualization experiment for dose reduction was described in detail as an example of Therapeutic Drug Monitoring. The inter-personal difference is utilized assuming no intra-personal difference occurring. However, considering the intra-personal variances, the modeling results will be closer to real life where the abnormal reaction in taking a drug for some chronic disease usually results from a change in patient’s intra conditions. Also the author[2] here applied Maximum Likelihood Method in his Bayesian model, which is a simple way to do the prediction. But many strict assumptions have to be made such as linear model, independent variables, Gaussian distribution, etc. In our trial, we are going to develop a more flexible model based on Bayesian network to describe the joint probability distribution among the observed data.

C. Other Mathematical Models in Literatures

As the percentage of population having at least one chronic health problem rises, the management of long term illness becomes an important and expensive element of health care in the society. In the control phase of clinical monitoring, there are usually some common statistical models that could be applied. A general framework for modeling this phase is presented in [3] together with the clinical examples on the monitoring of cholesterol, blood pressure, diabetes and HIV infection. As noticed by the author[3] that over-frequent monitoring turns out to be a bigger problem than under-frequent monitoring, a good statistical model is key to obtaining a good prediction in the clinical tests’ frequency during the control phase. In practice, monitoring is usually done in the form of regular clinical measurements to a patient before making the decision whether this patient needs treatment change or not. Similarly, diagnosis may be same as regular clinical measurements to decide whether to initialize the treatment or not. There are five different phases of monitoring:

• Detection of abnormal measurement.
• Treatment initiation.
• Control.
• Out of control.
• Treatment cessation.

For all the above phases, the variables such as what to measure, measurement intervals, etc, should be chosen with full considerations of patients’ personal conditions.

Generally, a model for the clinical control phase could be written as follows:

\[ U_{it} = \alpha(i) + \beta(i,t) \]

\[ Y_{it} = U_{it} + \omega(i,t) \]

where \( U_{it} \) indicates the \( i \)-th true value that is the initial value at time 0, \( \alpha(i) \), \( \beta(i,t) \) is the true change in the \( i \)-th patient between time 0 and time \( t \), and \( \omega(i,t) \) is a random error term for patient \( i \) at time \( t \). Those values are assumed to be independent and identically distributed for any \( i \neq j \), and the vectors such as \( (\beta(i,t_0), \beta(i,t_1), \ldots)^T \) and \( (\beta(j,t_0), \beta(j,t_1), \ldots)^T \) are also independent and identically distributed, and similarly for the vector of \( \omega(i,t_i) \). It is also assumed that \( \omega(i,t) \) are normally distributed.

After declaring the general model, different methods could be applied to fit the model. The author[3] has listed four approaches:

1) Maximum Likelihood methods, which can be used to fit a fully specified model. Several measurements \( Y_{it} \) are required on each individual \( i \) in this method in order to minimize the influence of errors on the variance \( \sigma^2_{it} \).
2) Moment-based methods, where the mean values of $\alpha(i)$ and $\beta(i, t)$ could be computed as
\[
\hat{\alpha} = \frac{1}{n} \sum_{i=1}^{n} Y_{it}, \quad (10)
\]
\[
\hat{\beta}(t) = \frac{1}{n} \sum_{i=1}^{n} (Y_{it} - Y_{i0}). \quad (11)
\]
However, due to the influence of the error term $\omega(i, t)$, we need to compute the variation of $\omega(i, t)$ before knowing the variances of $\alpha(i)$ and $\beta(i, t)$. Two methods are introduced to get $\sigma^2_\omega$. Firstly, it could be computed by $\frac{1}{2} \text{Var}(Y(i, t) - Y(i, s))$ under the estimation of $\beta(i, t) - \hat{\beta}(i, t)$, with confidence intervals $t - s$ approximately 0 and the estimation of uncorrelated $\omega(i, t)$ and $\omega(i, s)$ with sufficiently larger difference in $t$ and $s$. Secondly, $\text{Var}(Y_{it} - Y_{i0})$ could be estimated with a stronger assumption that $\text{Var}(\beta(i, t)) = \lambda t$, which is known as ‘variogram’ method. With either of the above two methods and together with an additional independence assumptions of $\alpha(i)$, $\omega(i, 0)$ and $\beta(i, t)$:
\[
\text{Var}(\alpha(i)) = \text{Var}(Y_{i0}) - \sigma^2_\omega \quad (12)
\]
\[
\text{Var}(\beta(i, t)) = \text{Var}(Y_{it} - Y_{i0}) - 2\sigma^2_\omega. \quad (13)
\]
3) Literature-derived models, in which the estimation of $\sigma_\omega$ could be found in a review of the literature. Then the remaining parameters could be drawn with this known $\sigma_\omega$. However, the drawback of this method is obvious due to the different conditions of clinical tests carried out in different studies.

4) Comparison of estimation methods. Since methods 1) and 2) deal differently with their handling of missing data. Maximum Likelihood Estimation focus on observations $Y_{it}$, hence equal weight is assigned to each observation, while Variogram method gives an equal weight to each time point $t$, which in practice seems to be better than the first method. In [12], moment-based method was recommended due to its facilitated estimation.

After being fitted, the model should be applied to the calculation in the control phase. Three approaches are introduced here:

1) Analytic methods. Normal distribution is assumed for $\alpha(i)$ and $\beta(i, t)$ so that at time 0 and $t$:
\[
\begin{pmatrix}
U_{it} \\
Y_{i0} \\
Y_{it} \\
\end{pmatrix}
\sim N
\begin{bmatrix}
\alpha(i) \\
\alpha(i) \\
\alpha(i) + \beta(i, t) \\
\end{bmatrix}
\Sigma_t
\]
where $\Sigma_t$ is some covariance matrix. If there is a threshold $Y^*$ above which the clinical test will report abnormal, the proportion of the true positive tests could be drawn as $P(Y_{it} > Y^*, U_{it} > Y^*)$, together with the distribution of $Y_{it}$ calculated from equation (14).

2) Calculation by simulation, which is more general than the first method to deal with complex models. In this method, a large number of samples are simulated from the modeled distributions with estimated parameters to compute the value of $P(Y_{it} > Y^*, U_{it} > Y^*)$.

3) Uncertainty estimates. To be more general, if the model parameters are not known but only their estimates could be obtained such as $\hat{\alpha}$ and $\hat{\beta}$, a general solution is the use of resampling procedures such as bootstrap methods[13]. The hierarchical nature of the original sample in the resampling algorithm is the key to this situation.

To apply the methods discussed above, four clinical examples on the control phase are given in the paper. Cholesterol monitoring of heart diseases, long-term blood pressure monitoring, HbA1c monitoring of type 2 diabetes, and CD4 monitoring in HIV infection.

Cholesterol monitoring[12]: More than 9000 patients were included in the study over 5 years. A general model by equation (8) and (9) has been used with normal distribution assumption of $\alpha(i), \beta(i, t)$ and $\omega(i, t)$.
\[
Y_{it} = \alpha_i + \beta_t + \omega_{it} \quad (15)
\]
Variogram method was applied to specify the variance of $\beta(i, t)$, hence $\sigma^2_\omega$ and $\sigma^2_t$ could be estimated. The rates of Type 1 and Type 2 errors were monitored at an interval of 1, 3, and 5 years, and it was found that cholesterol levels monitoring should be reduced from annually to every 3 to 5 years.

Blood pressure monitoring[14]: The monitoring control test was done on more than 6000 patients taking a blood pressure lowering medicine. The model of equation (15) was also applied here and $\beta_t = 0$ was assumed. The ratio of false-positive to true-positive blood pressure tests at various time points shows that over-frequent monitoring would reflect only measurement error to most patients.

HbA1c monitoring[15]: The model of equation (15) was applied with a linear assumption of $\beta_{it} = t \beta_t$, where
\[
\begin{bmatrix}
\alpha_i \\
\beta_t \\
\omega_{it} \\
\end{bmatrix}
\sim N
\begin{bmatrix}
\alpha \\
\beta \\
\omega \\
\end{bmatrix}
\begin{pmatrix}
\sigma^2_\alpha & \rho\sigma_\alpha\sigma_\beta & \rho\sigma_\alpha\sigma_w \\
\rho\sigma_\beta\sigma_\alpha & \sigma^2_\beta & \rho\sigma_\beta\sigma_w \\
\rho\sigma_w\sigma_\alpha & \rho\sigma_w\sigma_\beta & \sigma^2_w \\
\end{pmatrix}
\]
\[
(16)
\]
\[
\begin{bmatrix}
\alpha_i \\
\beta_t \\
\omega_{it} \\
\end{bmatrix}
\sim N
\begin{bmatrix}
0 \\
0 \\
0 \\
1 \\
\phi(t_{z-t1}) \\
\phi(t_{z-t1}) \\
\end{bmatrix}
\]
\[
1 \\
1 \\
1 \\
1 \\
\]
\[
(17)
\]
for any $i$ and $i < t$. The model was fitted by Maximum Likelihood Estimation. After comparing the false to true positive ratio, a conclusion could be drawn that this ratio would be halved if the interval between two tests were close to one year.

CD4 monitoring[16]: Proper CD4 test frequency could minimize the cost and prevent the delay in detecting severe infections. Similar to the third case, a linear assumption of $\beta_{it}$ was applied to the equation (15), where equation (16) also holds and $\omega_{it} \sim N(0, \sigma^2_w)$. But the correlation between $\rho$ and $\alpha_i, \beta_t$ was different from the case 3. The parameters used in [16] were derived by a thorough review of the literature. Instead of using point estimates of $Y_{it}$, one-sided 95% prediction interval was calculated for the time $t^*$ when the prediction interval first crosses the threshold. The results showed that between 4% to 32% of CD4 tests can be safely omitted if other conditions remained conservative.

However, all the discussions and examples above assume to use the observed values $Y_{it}$ instead of the real value $U_{it}$. The
author[3] of the paper extended some possible improvements:
(1) It is better to condition on the unobserved $U_{i0}$ when using
theoretical patients to motivate population-wide monitoring,
while $Y_{i0}$ would be better for the patient-specific monitoring.
(2) Threshold $Y^*$ considered in the fourth clinical example
was based on observed value $Y_{it}$, hence the real retesting
time could be even later than the one based on [16]’s model.
(3) Averaging the observed values is an effective way to
improve the test accuracy.

In this paper, one general mathematical model has been
discussed and compared with different fitting methods in the
sense of clinical monitoring. It is appealing to clinicians to
use a linear assumption with normal distribution due to the
simplicity. However the real world is not necessarily linear
and even some clinical parameters are already known to
be nonlinear such as the drug concentration in one patient
over time. Hence the model we are going to apply will be
a Bayesian network with underlying causal relationship in
between the observed clinical data to try to reduce the false-
positive and false-negative predictions.

III. RESEARCH PLAN

The following topics and questions will be covered in my
future work.

A. Descriptions of Observed Clinical Data

In Therapeutic Drug Monitoring, the purpose is to provide
a specific patient with dosage recommendation according to
our prediction model. The model will be built up and trained
from a database containing the clinical test results of the drug.
The data we are going to process includes three types:
- Patient’s personal data such as gender, age, weight,
  height, etc.
- Patient’s drug intakes such as dosage per day, intervals
  between two doses, drug’s pharmacokinetics parameters,
  other drugs’ influences, etc.
- The measurement results such as drug concentration,
  analyzing time, other symptoms, etc.

Moreover, the therapeutic range of a drug’s concentration in
blood is known as well as our target range. Thanks to
the practical meaning of the data, some of the causal
relationship in between the variables are obvious to observe.
But in our research, what we want to get is that given a specific
quantity of drug, our model could tell the probability that the
drug concentration in a specific patient is going to be within
the therapeutic range. Considering the cost, we also want to
minimize the amount of concentration measurements while
still guaranteeing the effectiveness and non-toxicity.

In our approach, we might meet with two situations: (1) give
the dosage recommendation of a drug that is new to a patient;
(2) give the dosage adjustment recommendation of a drug that
is taken by a patient regularly. In the first case, we are going to
build up the model based on a mass study, specifically on the
previous records of other people taking the same drug. Hence
we assume that the difference in dosage is only dependent on
the inter-personal relationship. In the second case, it is usually
the intra-personal difference causing the problem like the drug
concentration getting suddenly higher than the toxic level or
lower than effective level. The model will be trained based
on the personal clinical data over time. However, since such a
problem often indicates some other disabilities in the body,
a mass analysis and test results will also be included when
training the model.

In both cases, we could apply the Bayesian network method.
Each node of our network represents an observed parameters
e.g. age, gender, weight, etc; with a specific prior distribution.
This prior distribution could be derived from the literature
such as how the drug concentration varying in the body over
time, or from an experience model such as Dirichlet prior.
The former one is usually tested and measured over a large
population[17]. The edges in the Bayesian network indicate
the dependency and/or causal relationship between two nodes
(variables). Currently, we only focus on the Directed Acyclic
Graph as described in the first paper.

The method of Buclin et al.[16] applied to his clinical ex-
periments is based on a strong assumption of linear model and
independent variables, while in [2], the recommendation on a
drug’s dosage to a specific patient does not consider the current
drug concentration. Hence, in our approach, Bayesian network
method will be applied to describe the relationship of all the
observations related to give a joint probability distribution.
Since we deal with continuous data, some modification will be
applied to adapt the method computing Bayesian network. One
way to do it is by using grids. This is reasonable as our target
drug concentration is a range of data on its pharmacokinetic
curve; thus it can be divided into a grid.

To be more general, since the observed data might be influ-
enced by noises e.g. measuring tools’ conditions, temperature
etc; sometimes some data might not reflect the true value in
a patient as $Y_{it} = U_{it} + \omega(i,t)$ which was discussed in
[3]. This could be analyzed first to get the probability of
$P(U_{it} > Y^*|Y_{i0} > Y^*)$, where $Y^*$ denotes a threshold. Since
the noises are independent of each others, there are many ways
to reduce their influence. One simple way is to average the data
samples. Besides, assigning a distribution to the noise is also a
common way.

B. Learning Algorithms

In Bayesian network methods, we expect the algorithm to
automatically build up the network from the given data. To
do this, some assumptions are required: (1) The data obtained
is a good representation of the population. (2) The observed
data is all that should be considered to solve the case; that is,
there is no other unobserved dependency. (3) We can control
the overfitting problem by putting a prior distribution on the
parameters. (4) The causal relationship in between the
variables is assumed to be acyclic, e.g. we only consider
the influence of a change in the dosage on our target drug
dosage, while the variance of the concentration will not
affect other variables such as patient’s weight, etc.

The Bayesian network approach will also be compared with
other learning algorithms. Since the objective is to find a
suitable dosage based on some specific patient’s data, it could
be analogous to the classification or regression problem. Many
advanced algorithms for such problems have been developed in
the machine learning literature such as neural networks,
decision trees, support vector machine (SVM), etc. The benefit
of using those algorithms is that no prior distribution is
required as they simply learn everything automatically from
the observed data alone. Among the three algorithms listed
here, all of them could work with real data, and the SVM and
neural network can also work for regression.

The main idea of SVM is to separate the data by an
hyperplane in a high-dimensional space. Those data could
be all the clinical information of patients as $D_i \in \mathbb{R}^n$
and $n = \{\text{age, gender, weight, etc.}\}$[18]. A neural network
performs the classification in the normal data space, but
differently from an SVM it is using a network of hyperplanes
to separate the data[19]. Thus the hyperplanes are learnt to
divide the data into different classes. Decision trees directly
segment the data with axis-aligned planes, and could also
help to compute the conditional probabilities[20]. All these are
called learning step that could be achieved by training the
algorithms with observations. After the algorithm has learnt
the classifying structure, given a new patient’s clinical data,
it could return the class (indicating the dosage in our case) where
the data belongs.

Hence these learning algorithms will be compared with the
Bayesian network in our future work.

C. Clinical Monitoring

The overall benefit of this research is bettering medical
practice by enabling personalized medicine while reducing
health care costs, and offering guidance to doctors, especially
in clinical dosage assignment. We are going to advance the
state-of-the-art by providing mathematical models that could
work in an electronic-control dimension to drug treatment,
based on real-time sensing and on safe and optimal dosing
policies. Plus, the model should be acceptable by doctors.
Hence we need to consider the simplicity and effectiveness
to be two main properties of our models.

We are working with CHUV Hospital (Lausanne, Switzerland)
on this project. To obtain the patient samples and validate
the measurements are in the charge of them. Currently, the
clinical case we are dealing with is the dose-individualization
the measurements are in the charge of them. Currently, the
land) on this project. To obtain the patient samples and validate
the classifying structure, given a new patient’s clinical data,
it could return the class (indicating the dosage in our case) where
the data belongs.

Hence these learning algorithms will be compared with the
Bayesian network in our future work.

IV. CONCLUSIONS

Dose-individualization or Personalized Medicine, including
dose-reduction, is a main topic in Therapeutic Drug Mon-
itoring. It is believed that inter- and intra-personal differ-
ences have an influence over the drug concentration to be
in or out of the drug’s therapeutic range. In this proposal,
a general mathematical model and a clinical experiment have
been introduced, where strong assumptions have been made
in practical work. In order to obtain a more robust result,
Bayesian network is also discussed here to learn and build
up a network giving the joint probability distribution of all
the variables. Based on these methods, the model could be
used to help doctors in making decisions on the dosage as
well as long-term monitoring.

REFERENCES

[1] Dana Pe’er, “Bayesian Network Analysis of Signaling Networks: A

Delhumeau, Julia di Lulio, Alessandra Fleurent, Marie-Paule Schneider,
Matthias Cavassini, Amalio Telenti, Bernard Hirschel and Alexandra
Calmy, “Successful efavirenz dose reduction led by therapeutic drug
monitoring (TDM)”, International Medical Press - Antiviral Therapy
2010.

the control phase of clinical monitoring”, Statistical Methods in Medical

Delhumeau, Julia di Lulio, Alessandra Fleurent, Marie-Paule Schneider,
Matthias Cavassini, Amalio Telenti, Bernard Hirschel and Alexandra
Calmy, “Successful efavirenz dose reduction led by therapeutic drug
monitoring (TDM)”, International Medical Press - Antiviral Therapy
2010.

the control phase of clinical monitoring”, Statistical Methods in Medical

[6] Miguel Goicoechea, Andrea Vidal, Edmund Capparelli, Andrew Rigby,
Carol Kemper, Catherine Diamond, Mallory D Witt, Richard Haubrich,
“A computer-based system to aid in the interpretation of plasma con-
centrations of antiretrovirals for therapeutic drug monitoring”, the 12th
Conference of Retroviruses and Opportunistic Infections, USA, February

[7] Helene Bourgoin, Gilles Painaud, Matthias Buchler, Yvon Lebranchu,
Elisabeth Aurrent-Leca, France Mentre & Chantal Le Guellec, “Bayesian
estimation of cyclosporin exposure for routine therapeutic drug moni-
toring in kidney transplant patients”, British Journal of Clinical Pharmacol-

Bayesian approaches to dose-individualization for once-daily aminogly-


[10] Miguel Goicoechea, Andrea Vidal, Edmund Capparelli, Andrew Rigby,
Carol Kemper, Catherine Diamond, Mallory D Witt, Richard Haubrich,
“A computer-based system to aid in the interpretation of plasma con-
centrations of antiretrovirals for therapeutic drug monitoring”, the 12th
Conference of Retroviruses and Opportunistic Infections, USA, February

Bayesian feedback method of aminoglycoside dosing”, Clin Pharmacol
Ther 1985; 37; pp 349-357.

and effects of efavirenz in patients with human immunodeficiency virus

Groningen Sleep Quality Scale”, Groningen: Heymans Bulletin 1988: 88-
13-EX.

cholesterol levels: measurement error or true change?” Annals of Internal
Medicine 2008; 148(9); pp 656-661.

[15] Efros B, Gong G, “A leisurely look at the bootstrap, the jackknife,
and cross-validation”, The American Statistician 1983; 37(1); pp 36-48.

patients receiving treatment to lower blood pressure: analysis of data from
placebo controlled randomised controlled trial”, British Medical Journal
2009; 338: b1492.

diabetes”, Diabetes Care 2004; 27(7); pp 1761-1773.

counts to decide when to start antiretroviral therapy in HIV-1 infected
individuals”, Personal Communication.

Rosselet, B. Rochat, C. B. Eap, H. Henry, J. Bodilaz & T. Buclin, “Popula-
tion pharmacokinetics of imatinib and the role of α1-acid glycoprotein”,

[20] Nello Cristianini and John Shawe-Taylor. An Introduction to Support
Vector Machines and other kernel-based learning methods. Cambridge