Research Paper

The Non-Linear Child: Ontogeny, Isoniazid Concentration, and NAT2 Genotype Modulate Enzyme Reaction Kinetics and Metabolism

Zoe Rogers a, Hiwot Hiruy b, Jotam G. Pasipanodya c, Chris Mbowane d, John Adamson a, Lihle Ngotho a, Farina Karim a, Prakash Jeena d, William Bishai a,e, Tawanda Gumbo f,g

a KwaZulu-Natal Research Institute for TB and HIV, Durban 4001, South Africa
b Center for Tuberculosis Research, Department of Medicine, Johns Hopkins University, Baltimore, MD 21287, USA
c Center for Infectious Diseases Research and Experimental Therapeutics, Baylor Research Institute, Baylor University Medical Center, Dallas, TX 75204, USA
d Dept of Pediatrics, Nelson Mandela School of Medicine, UKZN, Durban 4001, South Africa
e Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA
f Department of Medicine, University of Cape Town, Observatory, South Africa
g Center for Tuberculosis Research, Department of Medicine, Johns Hopkins University, Baltimore, MD 21287, USA

A R T I C L E   I N F O

Article history:
Received 22 September 2015
Received in revised form 20 July 2016
Accepted 25 July 2016
Available online 27 July 2016

Keywords:
NAT2 reaction kinetics
Isoniazid concentration
Maturation
NAT2 genotype
Artificial intelligence

A B S T R A C T

N-acetyltransferase 2 (NAT2) catalyzes the acetylation of isoniazid to N-acetylisoniazid. NAT2 polymorphism explains 88% of isoniazid clearance variability in adults. We examined the effects of clinical and genetic factors on Michaelis-Menten reaction kinetic constants of maximum velocity (Vmax) and affinity (Km) in children 0–10 years old. We measured the rates of isoniazid elimination and N-acetylisoniazid production in the blood of 30 children. Since maturation effects could be non-linear, we utilized a pharmacometric approach and the artificial intelligence method, multivariate adaptive regression splines (MARS), to identify factors predicting NAT2 Vmax and Km by examining clinical, genetic, and laboratory factors in toto. Isoniazid concentration predicted both Vmax and Km and superseded the contribution of NAT2 genotype. Age non-linearly modified the NAT2 genotype contribution until maturation at 5.3 years. Thus, enzyme efficiency was constrained by substrate concentration, genes, and age. Since MARS output is in the form of basis functions and equations, it allows multiscale systems modeling from the level of cellular chemical reactions to whole body physiological parameters, by automatic selection of significant predictors by the algorithm.

© 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Isoniazid is a first line antituberculosis agent recommended for treatment of children with tuberculosis (TB). Its conversion is catalyzed by arylamine N-acetyltransferase (NAT) isoenzyme 2 (NAT2) in phase II xenobiotic metabolism. NAT2 is central to detoxifying many environmental-, industrial-, and food-based arylamines and hydrazines. NAT2 catalyzes the reaction of acetyl-CoA and isoniazid (substrates) to produce CoA and N-acetylisoniazid. NAT2 single nucleotide polymorphisms (SNPs) identified in adults led to classification of individuals as slow and rapid acetylators. Specific SNPs could be associated with changes in enzyme catalytic activity, as defined by enzyme reaction kinetic constants in liver tissue (Blum et al., 1991; Freland et al., 2001). The elimination of isoniazid from the body is a combination of NAT2 metabolism and direct elimination of parent compound, with parent isoniazid elimination in urine of up to 37%. However, in adult patients 88% of systemic clearance variability has been linked to NAT2 SNPs (Kinzig-Schippers et al., 2005). While the ontogeny of phase I enzymes is well established, the age-dependent maturation changes in phase II enzymes such as NAT2 are still poorly understood. This is important since isoniazid peak concentrations and 0–24 h area under the concentration-time curves (AUC0–24) are major determinants of cure rates, speed of sterilizing effect, and acquired drug-resistance, in TB patients (Chigutsa et al., 2015; Pasipanodya et al., 2013, 2012; Pasipanodya and Gumbo, 2011; Gumbo et al., 2014; Dheda et al., 2014). There could be an age-dependent difference in systemic clearance of isoniazid in children, however, the effect of maturation, or other pediatric factors, on NAT2 enzyme kinetics has hitherto not been investigated (Jeena et al., 2011; Rey et al., 2001). Here, we identified the rate of production of N-acetylisoniazid and the rate of elimination of isoniazid in children enrolled in the prospective PHATISA study (Hiruy et al., 2015) and investigated the possible effects of NAT2 SNPs, measures of body-size, age, and nutrition status, in altering NAT2 reaction kinetics and isoniazid elimination rate constants.

Human growth and physiological maturation are non-linear processes, with discordant changes that have direct effect on drug pharmacokinetics (Kearns et al., 2003). Indeed, in recent years it has become evident that even in adults the relationships between different
pharmacokinetic parameters and covariates such as age and weight are not only non-linear, but have high order interactions, with data zones of discontinuity in the relationships (Hall et al., 2011, 2013; Jain et al., 2013). Thus, standard statistical analyses could be limited on several fronts, including reliance on linear analyses and the restrictive normality, collinearity and sampling assumptions needed for valid inference. In 2012, Kiser et al. used standard statistical and pharmacokinetic approaches and identified that NAT2 genotype and child’s age were associated with isoniazid clearance in South African infants. With the slow acetylator genotype, weight adjusted clearance was unchanged from week 0 to week 12 but increased 28% in week 84; while for rapid acetylators it increased 13% from week 0 to week 12 but then decreased by 23% in week 84 (Kiser et al., 2012). In other words, the changes were non-linear. Isoniazid elimination is from a combination of NAT2 metabolism and loss of parent compound in urine, which could account for the complex picture observed. Here, we used non-linear science to identify the role of maturation, weight, and NAT2 SNPs on xenobiotic metabolism in children (Campbell, 1987; Campbell et al., 1985; Dokoumetzidis et al., 2002).

We applied artificial intelligence (AI)-based non-linear analyses to investigate NAT2 maximum velocity ($V_{\text{max}}$) and affinity ($K_{\text{m}}$) during the first decade of life in children with TB. We were interested in determining relationships that could be used to scale from the level of enzyme-based chemical reactions to the whole organism (patient) in the clinic, to allow for direct translation.

Machine learning is a branch of AI-based methods that we have applied to examine pathophysiological and clinical parameters as well as therapeutic outcomes in the past, whereby it outperformed standard statistical inference with datasets of as little as 28 patients (Chigutsa et al., 2015; Cumbo et al., 2014; Modongo et al., 2015, 2016; Jain et al., 2013; Pasianidoya et al., 2013, 2015). One method, multivariate adaptive regression splines (MARS), was introduced by Friedman (1991). MARS has been utilized by several groups for a variety of purposes, with both large and small clinical sample sizes for a larger number of predictors (Brasier et al., 2012; George and Chang, 2014; Lee et al., 2014; Lin et al., 2008; Liu et al., 2015; Ju et al., 2014). Studies included examination of concentrations of 8 proteins as potential predictors in 30 patients for helicobacter infection, RNA-sequencing data to differentiate genes with low and high expression in the entire human genome from sample sizes as low as four per group, and an examination of 419 plasma protein concentrations as potential predictors of dengue fever in 55 patients. MARS uses an optimization procedure that employs basis functions (BF) to examine candidates for both main and interaction effects in a piecewise manner. BF’s are flexible mathematical linear spline functions that describe relationship between variables in short segments of dataspace, which allows fit of both nonlinear and linear relationships as well as their interactions simultaneously. A backward elimination step is used to select a set of BF’s that minimizes mean-square errors, after which multifold cross validation procedure is employed for model validation. The general cross-validation (GCV) step gives a measure of how well the model will perform on an independent dataset. In AI, an interaction is encountered when one predictor modifies another predictor’s effect on the outcome (say $V_{\text{max}}$). In the same model, MARS identifies inflection thresholds (termed “hinges”), where relationship between predictor and target changes or even ceases to exist within specific and well-defined data-space. This essentially allows the slope of the relationship between target and predictor to be different for different data ranges of the same predictor variable, and allows for discontinuity of relationships in some data spaces. Finally, the variable importance score is used for ranking and selecting predictors, especially important when the number of predictors is relatively large compared to the number of observations (as was the case in our study) and when many of the predictors could be co-linear (Breiman, 2001). Here, we applied MARS to data obtained from an experimental clinical design to identify predictors of NAT2 $V_{\text{max}}$ and $K_{\text{m}}$ in South African children with TB.

The philosophy and purpose of AI methods such as MARS differ from standard statistical inferences (Breiman, 2001; Hastie et al., 2009). The underlying scientific philosophy in standard statistical inference is rejection of a pre-specified null hypothesis. Comparison of distributions and the measures of central tendency (or hazard rates) between two or more groups are often the main way to test hypotheses, with tests of significance designed to make sure that differences observed are not due to chance. The purpose of MARS is not for hypothesis testing, but rather for hypothesis generation. Since these algorithms are distribution free, measures of central tendency are not the issue. Moreover, MARS algorithm uses non-parametric regression: in other words, predictors do not take a pre-specified form but are derived from the data. The main objective of MARS is prediction and pattern recognition, based on present data. Thus, there is no test of significance. Rather, the algorithms determine how likely the prediction is likely to be true with different datasets. During the GCV-step, the algorithms randomly split the dataset into different sized smaller sets and calculate a GCV score, which tells us how well predictors will perform on an entirely new data-set. When predictors and their interactions are identified with these approaches, they become hypotheses to be tested using standard statistical inference.

2. Methods

2.1. PHATISA Study Design

The study design, regulatory approvals, and the summary pharmacokinetic parameter results of the PHATISA study have been published in the past (Hiruy et al., 2015). Clinical and demographic factors were recorded. Malnutrition was measured using Z scores for height-for-age, weight-for-age, and weight-for-height, and mid-upper arm circumference (MUAC), since it has been proposed that isoniazid elimination could be affected by malnutrition. The nutrition measures such as MUAC and Z-scores were interpreted based on WHO criteria (http://whqlibdoc.who.int/hq/1997/WHO_NUT_97.4.pdf). The children had blood draws after directly observed therapy between the fourth and twelfth day after initiation of anti-TB therapy. In order maximize experimental design and limit the number of samples drawn, and to reduce the number of children under study, times for blood draw were specifically identified using optimal sampling theory (Reed, 1999; Wang and Endrenyi, 1992). We drew blood for concentration of both isoniazid and its metabolite, N-acetylisoniazid at 0, 0.42, 1.76, 3.37, 10.31 and 24 h after an isoniazid dose. Blood was processed for extraction of isoniazid and N-acetylsalizosanid, as well as for DNA extraction.

2.2. Isoniazid and N-acetylsalizosanid Concentration Measurement Assays

N-acetylsalizosanid was synthesized from isoniazid standard using acetic anhydride and a previously published method. Following sample preparation using an acetonitrile-based protein precipitation process, measurement of drug and metabolite concentrations was carried out by liquid chromatography-tandem mass spectrometry on an AB Sciex 5500 Q-trap mass spectrometer coupled to an Agilent 1200 UPLC system. An Exterra 2.1 mm × 50 mm C18 column and a mobile phase of water with 0.1% formic acid at a flow rate of 200 µL/min was used for separation. The mass spectrometer was operated in the positive ion mode and the following transitions monitored: Isoniazid: 138.1 → 51.9 and 138.1 → 66, N-acetyl isoniazid: 180 → 66 and 180 → 78.6 and 6-aminonicotinic acid (internal control): 138.7 → 51 Analysys® 1.5 software version 1.5.1 was used for quantitation.

2.3. NAT2 Genotyping

DNA was extracted from buffy coats using the Qiagen Flexigene DNA Kit. The entire NAT2 gene was amplified by PCR using the sense primer (5’ GAC ATT GCA TAT TTT GAA AG 3’) and the antisense primer (5’ GAC ATT GCA TAT TTT GAA AG 3’).
GAT GAA ACT ATT TGA TGT TTA GG 3’ primer pair (Hickman and Sim, 1991). Amplification conditions were as follows: an initial denaturation of 5 min at 95 °C, followed by 35 cycles of denaturation at 95 °C for 30 s; annealing at 55 °C for 30 s; extension at 72 °C for 60 s, with a final extension of 5 min at 72 °C. Sanger sequencing (performed by Inqaba Biotechnical Industries) was used to determine NAT2 genotypes. NAT2 alleles were characterized by polymorphisms located at seven positions 191, 282, 341, 481, 590, 803, and 857, as described by the Database of arylamine N-acetyltransferases (NATs) (http://nat.mbg.duth.gr; McDonagh et al., 2014; Sabbagh et al., 2011). SNPs at these 7 positions have a genotype/acetylation status prediction of accuracy of 98.4% (Cascorbi et al., 1995; Hein and Doll, 2012).

2.4. Identification of Enzyme Kinetics and Pharmacokinetic Parameter Estimates

We implemented two types of models from the ADAPT software of D’Argenio et al. (2009). The parameter estimates for each child were estimated using the maximum-likelihood solution via the expectation-maximization algorithm (MLEM). The changes in concentration of isoniazid (INH) and N-acetylisoniazid (ANIH) with time were utilized in two types of pharmacometric models. The first was a pharmacokinetic model for the plasma concentrations of parent drug and its metabolite, with no assumptions of NAT2 saturability. The isoniazid elimination rate (Kmin) as well as the fraction of isoniazid metabolized (fmet) and the volume of distribution of metabolite VPATH were used to calculate the elimination ratio VPATH/Kmin. In order to identify NAT2 Vmax (mg/h) and Km (mg/L) for isoniazid metabolism in each child, we co-modeled the concentration versus time of isoniazid and N-acetylisoniazid using a 2-component model, chosen based on our prior work with this drug (Hiruy et al., 2015; Pasipanodya et al., 2013). A standard two-stage estimation method was used to generate initial Km and Vmax values and other pharmacokinetic parameter estimates. The parameters identified were then used in subroutine POPINIT of ADAPT, and enzyme reaction kinetic constants and pharmacokinetic parameter estimates for each child were then identified using MLEM.

2.5. Frequentist Statistical Analysis

We employed the D’Agostino and Pearson omnibus normality test and a p-value > 0.05 for normality tests. Comparison between proportions was made using the Fischer’s exact or χ-square tests. Comparison between medians or means made via the Kruskal-Wallis test. All statistical testing other than machine learning methods were performed using STATATA version 13 software (STATATA Software, College Station, Texas).

2.6. Machine Learning

MARS uses BFs to examine candidates for both main and interaction effects, which gives MARS flexible and adaptive capabilities to fit nonlinear and linear relationships and interaction components simultaneously. Over fitting procedures were used to grow large models with up to 15 BFs, which were then pruned back using GCV function during the backward pass. Each variable was assigned a measure of predictive importance by MARS, entailing both marginal and interaction effects involving this variable. We examined for predictors of four target phenotypic responses: Vmax, Km, Kinh, and VPATH/fBf. The potential predictors entered for each model for each of the four responses were each child’s NAT2 genotype, age in days, INH dose, observed peak concentration, HIV status, gender, height, weight, body mass index (BMI) in kg/m2, Z-scores for both height and weight, as well as MUAC. The coefficient of determination (R²) values, variable importance measures and the MARS equation for selected models are reported. All machine learning modeling was performed on the SPM Salford Predictive Modeler® software suite version 7.0 (SALFORD Software, San Diego, California).

Except for setting maximum allowable basis functions to 15, all other settings were set to default for both the exploratory legacy mode and in cross-validation.

3. Results

We enrolled children ≤10 years old who presented with TB to King Edward VIII Hospital in Durban, South Africa, between May 2012 and March 2013. We enrolled 19 boys and 11 girls, whose clinical characteristics are shown in Fig. 1. The figure shows that most children were one year-old or less. All children had chest X rays performed, 19 (63%) had consolidation and 4 (13%) had cavitation, thus most children had pulmonary TB. Fig. 2 shows Z-scores and demonstrates that up to 30% of children were malnourished, though only one was severely malnourished (Z-score standard deviation < −3).

Isoniazid acetylation follows a ping-pong Bi-Bi mechanism described by the standard Michaelis-Menten relationship for enzyme rate reaction rate (v) and substrate concentration [S] (Michaelis and Menten, 1913; Weber and Cohen, 1968):

\[ v = \frac{V_{\text{max}} \cdot [S]}{K_{\text{m}} + [S]} \]  

We generated Km and Vmax values for each child in ADAPT based on this relationship, with estimates shown in Fig. 3A-B. The mean population estimate for Km was 1.69 mg/L, which translates to 0.012 μM, while that of Vmax was 30.11 mg/h, which translates to 0.06 μM/s. Fig. 3A-B show that Km was tightly distributed around the mean, but there was greater variability in the Vmax. The isoniazid elimination rate (Kmin) is shown in Fig. 3C. Fig. 3D shows the elimination ratio VPATH/fBf, calculated using the fraction of isoniazid metabolized (fmet), and the volume of distribution of metabolite VPATH.

NAT2 sequencing results, including the base pairs and SNP calls for each allele, are shown in Extended Dataset 1. We identified 17 different alleles, including four reference high-activity NAT2*4 alleles. The most frequent genotypes were NAT2*5B/NAT2*5B encountered in 6/30 (20%) children, and NAT2*3C/NAT2*12C encountered in 6/30 (20%) children; the former genotype is commonly reported in literature to confer slow acetylator status, while the latter is predicted to confer the intermediate acetylator status (McDonagh et al., 2014). On the other hand, homozygosity for the rapid acetylator NAT2*4 reference allele (NAT2*4/NAT2*4 genotype) was encountered in 2/30 (7%) of enrolled children (Table 1).

As a first step, we employed standard frequentist statistics to compare the distributions of Km, Vmax, and elimination rate between different inferred genotypes and predicted acetylation status. Table 2 shows no significant differences in mean values. Compared the distributions of the same enzyme function parameters, as well as measures of body size, age, and dose by inferred acetylator status, shown in Table 2. The only statistically significant finding was that the isoniazid dose administered was lower among rapid acetylators.

Next, we used MARS algorithm in order to identify predictors of Vmax. Table 3 shows that dose (BF1), genotype (BF5), isoniazid peak concentration (BF2), and age (BF12) were identified as best predictors of Vmax. The table illustrates the nature of MARS output, which is the form of basis functions (BF). BF1 is a simple hinge function of Dose (0, Dose − 190) which means that the value of the expression is zero for all doses (mg) up to 190 (i.e., for all doses satisfying “Dose minus 190” less than zero), as illustrated in Fig. 4A. The dose of 190 mg is at the hinge. The variable importance score of 100% means that “Dose” is the most important variable for Vmax. Isoniazid peak concentration was also an important predictor (BF2), with a variable importance score of 90%, very close to that of “Dose”; thus both dose and peak concentration likely reflect the same thing, which is effect of isoniazid concentration on Vmax. BF5 means that isoniazid concentration had no effect on Vmax until the peak concentration achieved the threshold value of 8.83 μM/L at the hinge. Table 3 shows that the peak concentration (BF7) interacted with BF5 (genotype). However, the BF5 interaction with BF7
only applied to NAT2 genotypes selected as subset 1, which means that there is staggered or stratified effect so that relationship applied to some genotypes and not others. Specifically, BF5 comprised NAT2 genotypes that were grouped by MARS, which had a variable importance score of 90%, similar to that of isoniazid concentration. Of note, for MARS analyses we did not pre-categorize different genotypes by acetylator status (slow or rapid acetylator), since predictors do not take a pre-specified form but are derived from the data. Table 3 further shows that, in addition, the child’s age also had an effect on \( V_{\text{max}} \) as well, as reflected by BF10; the impact however was less than half that of either genotype or concentration. Fig. 4B shows the singular contribution of age; the effect of age starts at zero but reaches a maximum at 1935 days (about 5.3 years).
5.3 years) when it reaches a plateau, and remains flat into the next decade. Overall, the relation between \( V_{\text{max}} \) and these four predictors was given by the equation:

\[
V_{\text{max}} (\text{mg/hr}) = 31.297 + 0.030^*\text{BF}_1 - 1.025^*\text{BF}_2 - 0.001^*\text{BF}_{10}; \quad R^2 = 0.89
\]

Thus, age (maturation) affects speed of isoniazid acetylation by modifying effect of genotype, a non-linear interaction, however after 5.3 years the maturation is complete.

Table 4 shows the BFs for NAT2 affinity for isoniazid, or \( K_m \). A child’s height was the primary predictor and had the highest variable importance score. Height however could be substituted by age, given the tight link between height and age (Pearson \( r = 0.917; p < 0.001 \)), a tightly linked relationship frequently used in the clinic. Thus age was the main predictor of \( K_m \) as well; in fact, the coefficient of determination \( (R^2) \) value of age alone was 31%. However, MARS also identified isoniazid peak concentration and genotype as important predictors (Table 4). Combined, these factors explained 80% of the variance in NAT2 enzyme affinity, as given by equation for \( K_m \) and its predictors:

\[
K_m (\text{mg/L}) = 1.721 - 0.001^*\text{BF}_1 + 0.044^*\text{BF}_2 - 0.001^*\text{BF}_4 + 0.045^*\text{BF}_6; \quad R^2 = 0.80
\]

Table 4 shows that \( \text{BF}_4 \) is a special type of a basis function, described by an interaction of \( \text{BF}_1 \) (age or height) and genotype. In other words, age constrained the effect of genotype on enzyme affinity.

Table 5 shows the BFs for isoniazid elimination constant (\( K_{\text{INH}} \)). The elimination rate constant is a composite measure reflecting both acetylation and elimination of parent compound via loss in urine. The main predictors were isoniazid dose and \( \text{NAT2} \) genotype. In this case, however \( \text{BF}_2 \) was an interaction based on the modification of effect of genotype on the effect of isoniazid dose (i.e., isoniazid concentration), an example.
of higher order interaction among predictors. This relationship is shown in Eq. (4):

$$K_{\text{HHI}} = 10.297 + 0.00033'B_{F_1} - 0.0002'B_{F_2}; R^2 = 0.64$$ (4)

This was confirmed by standard regression methods, which showed that for each unit increase in isoniazid dose above 40 mg, $K_{\text{HHI}}$ increased by an additional $2.82 \times 10^{-4}$ per hour for subset 1 versus $1.05 \times 10^{-4}$ per hour for subset 2 genotypes, a 2.69-fold change in slope.

As regards to the ratio $V_{\text{AINH}}/V_{\text{FINH}}$, the $R^2$ for the relationship was only 0.55 in the best model. We thus considered that our potential predictors such as clinical and demographic factors were not as explanatory of this parameter, and thus no further explorations were made.

4. Discussion

In pharmacometrics, primary pharmacokinetic parameters are the observed values such as elimination rate constant ($K_{\text{HHI}}$), volume of distribution (both central and peripheral), and in the current study also $K_m$ and $V_{\text{max}}$. These observations quantify rates and extents of underlying physicochemical processes. In other words, these parameters, including the enzyme reaction constants, are the biological phenotypes. On the other hand, in pharmacogenomics, inferred acetylation status is common terms of the “acetylation phenotype.” Acetylation phenotype is a grouping of drug clearances (into 3 overlapping Gaussian curves). Moreover, commonly in clinical practice, acetylation status is assigned using 2-h post-dose isoniazid concentrations. However, the isoniazid elimination rate itself is a product of at least two physicochemical processes, acetylation and the loss of unmetabolised (parent) compound in urine. “Phenotyping” methods such as the 2-h concentration or concentration-time profiles or $AUC_{0-24}$ are even further removed from the primary process of N-acetylation: they are products of several primary biological processes, namely absorption rate constant (and closely related time to maximum concentration), volume of distribution, and elimination rate constants. Only the last of these is related to acetylation, and only partially. The primary biological process of acetylation itself (i.e., the phenotype), sensu strictior, is defined by $K_m$ and $V_{\text{max}}$ values, which are the “observed” values based on calculations using the Michaelis-Menten Eq. (1). These values are thus not inferred from genotypes, but are the “observed”. Thus, effect of genotype should be inferred based on these phenotypes, and not vice-versa. Here, we simply asked: what are the physiological and clinical factors that could affect these quantitative measures of phenotype, if at all, and which factors affect the most? In terms of the genes, what groupings of NAT2 genes categorize the phenotypes of isoniazid NAT2 affinity and maximum velocity in children, during periods of highest growth? MARS, in which predictors do not take a pre-specified form or groupings (e.g. inferred acetylation status), is excellent at such an agnostic function, and utilizes only the presented data for its own categorization.

To answer these questions, we identified the $K_m$ and $V_{\text{max}}$ values for each child from clinical data, based on a pharmacometric approach. The mean population estimates of 0.012 $\mu M$ for $K_m$ for isoniazid. In biochemical assays for measurement of $K_m$ that examined isoniazid acetylation by human NAT2 from 2-h post-mortem liver homogenates of a rapid acetylator in the laboratory, the $K_m$ values were 0.018 $\pm 0.004$ $\mu M$ (Weber and Cohen, 1968). This means that the pharmacometric approach of using the rate of production of N-acetylsisoniazid and disappearance of isoniazid is a robust enough approach to calculate these values for each child, and gives values similar to those from enzyme kinetic experiments of liver biopsy specimens. In other words, the measurement of the phenotype was likely accurate. In the case of the NAT2 alleles we identified in the children, genotypes such as NAT2*4, NAT2*12A, NAT2*12B, NAT2*12C, and NAT2*13 alleles, and the predominance of the NAT2*5 allele have been shown to be common in Black South African adults (Dandara et al., 2003; Loktionov et al., 2002). Thus, the distribution of the common alleles in our small pediatric cohort mirrors that in adults. Application of AI methods to these parameter estimates and genotypes identified several important patterns between the predictors and enzyme reaction kinetic constants. Our approach is tractable and follows steps that are easy to apply to any pediatric drug whose metabolite is known for each child, and for a population of children. Thus, our approach could be used to examine the effects of maturation and drug doses on both compartmental pharmacokinetic parameters and Michaelis-Menten relationships for other phase II metabolism reactions, an unexplored pediatric space (Kearn et al., 2003).

Second, a current problem in systems pharmacology is the difficulty in scaling mathematical models from the level of chemical reactions
Fig. 4. Contribution of dose and age on maximum velocity. (A) Effect of dose on $V_{\text{max}}$ shows lack of effect of dose below 190 mg, after which dose has a positive relationship with $V_{\text{max}}$. In this case, dose is a surrogate of drug concentration. (B). The figure summarizes the hinge function $\max (0, 1934.5-\text{Age})$, a perfect example of the flexibility of MARS in modeling change in relationship between predictor and outcome at a hinge, and thus non-linear. The figure shows that there is a large effect of age until 5.3 years when the contribution gets to a plateau, suggesting the point of full maturation.

inside the cell to the level of the individual patient, and to populations of patients. The MARS outcome could allow this to be accomplished, starting with reaction kinetics at the level of a Michaelis-Menten reaction to whole body physiological parameters in the child (isoniazid elimination rate), based on relevant clinical and demographic predictors automatically selected by the algorithm. Moreover, MARS handles both linear and non-linear interactions simultaneously, allowing for construction of non-linear models. Furthermore, MARS is flexible enough to identify interactions without a prior falsifiable hypothesis to link reactions constants to a predictor, as is the case of isoniazid concentration and $K_m$. The findings led to new hypotheses being generated, which can now be tested in separate studies using standard statistical approaches. Thus, the mathematical relationships are identified first, and then physiological meanings can be investigated further to improve the precision of the estimates for better reproducibility.

Third, we found that the relationships between predictors such as age, dose or isoniazid concentration, and NAT2 enzyme kinetics were non-linear. In other words, several clinical and physiological factors affected speed of isoniazid acetylation and affinity, and interacted with each other. Of major importance was the effect of age on both speed of acetylation and enzyme affinity. Age played a critical role in modulating these enzyme activities, and age's contribution changed with child's age, with the maximal adult type of activity encountered after 5.3 years old. This time period likely represents the time period for maturation of the NAT2 enzyme. On the other hand, the effect of dose and drug concentration always superseded that of NAT2 genotype. The effect of dose and drug concentration was persistent regardless of the enzyme reaction constant examined, and was often the primary predictor, except in the case of $K_m$. The physiological and pharmacological basis for this remain to be worked out, but the fact that the hinge was always encountered beyond a certain concentration suggests a real concentration-dependent effect, consistent with many pharmacological processes. The GCV step identified $R^2 \geq 80\%$ on post-test sets, which means there is a high likelihood of the same observations with future data-sets, so that the findings are likely valid. However, clearly nothing in our current understanding of $K_m$ and $V_{\text{max}}$ can explain how the pharmacological and clinical factors we identified as predictors could affect both enzyme affinity and maximum velocity for isoniazid acetylation. However, the historical disaster of gray baby syndrome in children treated with the drug chloramphenicol suggests that maturation has a major effect on enzyme function (Glazer et al., 1980; Mulhall et al., 1983; Sutherland, 1959). The functional deficiency in the UDP-glucuronyl transferase enzyme system encountered in the first few weeks of life, especially in premature neonates, led to cardiovascular collapse when high concentrations of the antibiotic were achieved in the children (Glazer et al., 1980; Mulhall et al., 1983; Sutherland, 1959). However, we reiterate that, the philosophy behind machine learning methods such as MARS is prediction and pattern recognition, and not identification of causal pathways. Thus, the next step is to test the effect of age and isoniazid concentration on NAT2 $K_m$ and $V_{\text{max}}$ using standard hypothesis testing approaches, and if confirmed then a new understanding of NAT2 catalyzed reaction kinetics should be sought.

Fourth, our findings could impact the most widely used pharmacokinetic models of maturation as commonly recommended by regulatory authorities (European Medicines Agency, 2007). Several methods are used to examine the effect of maturation on pharmacokinetic parameters, with the two most common being allometric scaling based on the ¾ fractal geometry law and physiologically-based pharmacokinetic modeling (Hope et al., 2007). These methods use specific drug clearance rates derived in adults to estimate drug clearances in children, with ontogeny effects scaled based on dry tissue weight of the organ clearing the drug. The central assumption is that xenobiotic metabolism enzyme reaction rates are the same between adults and children, and that what varies is the quantity of the enzymes proportional to organ weight and size. It is recognized in some cases, as for example sirolimus, that there could be different fetal isoform enzymes that get replaced by others after birth, thereby leading to altered clearance (Emoto et al., 2015); nevertheless the reaction kinetic constants of each of the different isoforms are assumed fixed. Moreover, the development of clearance

### Table 4

Factors predictive of enzyme affinity ($K_m$) on MARS analysis.

| Function                      | Units     | Variable importance score (%) |
|-------------------------------|-----------|------------------------------|
| BF$_1$ & BF$_3$                | max (0, height-54) | 100 |
| BF$_2$                        | Max (0, isoniazid peak concentration-11.7) | 53  |
| BF$_4$                        | Genotypes subset 1*BF$_1$ | 26  |
| BF$_5$                        | Genotypes in subset 2 | 26  |

*BF$_1$ Genotypes subset 1: *SD/*1A, *5C/*12C, *4/*14, *4/*5, *4/*5C, *5B/*5B, *5B/*5C, *5C/*5K, *5C/*6N.

*BF$_2$ Genotypes subset 2: *12A/*13A, *6A/*12A, *4/*4.
with maturation is described by a single mathematical function. Our data suggests that non-linear functions and analyses, with no prior assumptions, could be more accurate.

Finally, we found no quantitative relationship between several measures of malnutrition such as Z-scores and isoniazid metabolism. Indeed, no single model included Z-scores and MUAC as a predictor. However, most children had moderate malnutrition, and it could be that effects are encountered only with severe malnutrition. Thus, this question remains unresolved.

Our study has several limitations. The first is the small sample of children enrolled. However, we deliberately optimized experimental design to identify information rich pharmacokinetic sampling so that the smaller numbers were deliberate. Moreover, MARS has successfully identified predictors in smaller populations than these ones in several studies by us and by others. Second, our findings lack physiological explanations. These will be subject of future studies. Third, the intent of AI approaches we used is not hypothesis testing, but rather pattern recognition. Thus, future studies in which our findings are tested as a falsifiable hypothesis are recommended.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ejmb.2016.07.031.

Author Contributions

Conception and design: T. Gumbo, B.M., J. Pasipanodya, P. Jeena, W. Bishai.

Development of methodology: Z. Rogers, H. Hiruy, J. G. Pasipanodya, J. Adamson, P. Jeena, W. Bishai, T. Gumbo

Acquisition of data: Z. Rogers, H. Hiruy, C. Mbowane, J. Adamson, L. Ngotho, F. Karim, P. Jeena, W. Bishai

Analysis and interpretation of data: Z. Rogers, J. Pasipanodya, P. Jeena, W. Bishai, T. Gumbo

Writing, review and/or revision of the manuscript: Z. Rogers, H. Hiruy, J. Pasipanodya, P. Jeena, W. Bishai, T. Gumbo

Study supervision: P. Jeena, W. Bishai

Conflict of Interest

All authors declare no conflict of interest.

Acknowledgements

The support of the Howard Hughes Medical Institute and NIH grants AI37856, AI36973, 97138, R56AI111985 is gratefully acknowledged.

References

Blum, M., Demierre, A., Grant, D.M., Heim, M., Meyer, U.A., 1991. Molecular mechanism of slow acetylation of drugs and carcinogens in humans. Proc. Natl. Acad. Sci. U.S.A. 88, 5237–5241.

Brasier, A.R., Garcia, J., Wiktorowicz, J.E., Spratt, H.M., Comach, G., Ju, H., Recinos, A., Soman, K., Forshey, B.M., Halsey, E.S., Blair, P.J., Rocha, C., Bazer, F., Victor, S.S., Wu, Z., Stafford, S., Watts, D., Morrison, A.C., Scott, T.W., Kochel, T.J., 2012. Venezuelan Dengue Fever Working Group. Discovery proteomics and nonparametric modeling pipeline in the development of a candidate biomarker panel for dengue hemorrhagic fever. Clin. Transl. Sci. 5, 8–20.

Breiman, L., 2001. Statistical modeling: the two cultures. Stat. Sci. 16, 199–231.

Campbell, D.K., 1987. Nonlinear science: from paradigms to practicalities. Los Alamos Science Special Issue, pp. 218–262.
Lin, H.Y., Wang, W., Liu, Y.H., Soong, S.J., York, T.P., Myers, L., Hu, J.J., 2008. Comparison of multivariate adaptive regression splines and logistic regression in detecting SNP-SNP interactions and their application in prostate cancer. J. Hum. Genet 53, 802–811.

Liu, R., Li, X., Zhang, W., Zhou, H.H., 2015. Comparison of nine statistical model based warfarin pharmacogenetic dosing algorithms using the racially diverse international warfarin pharmacogenetic consortium cohort database. PLoS One 10, e0135784.

Loktonov, A., Moore, W., Spencer, S.P., Vorster, H., Neil, T., O’Neill, I.K., Bingham, S.A., Cummings, J.H., 2002. Differences in N-acetylation genotypes between Caucasians and Black South Africans: implications for cancer prevention. Cancer Detect. Prev 26, 15–22.

McDonagh, E.M., Boukouvala, S., Akilli, E., Hein, D.W., Altman, R.B., Klein, T.E., 2014. PharmGKB summary: very important pharmacogene information for N-acetyltransferase 2. Pharmacogenet. Genomics 24, 409–425.

Michaelis, L., Menten, M.L., 1913. Die Kinetik der Invertinwirkung. Biochem. Z 49, 333–369.

Modongo, C., Pasipanodya, J.G., Zetola, N.M., Williams, S.M., Sirugo, G., Gumbo, T., 2015. Amikacin concentrations predictive of ototoxicity in multidrug-resistant tuberculosis patients. Antimicrob. Agents Chemother 59, 6337–6343.

Modongo, C., Pasipanodya, J.G., Magazi, B.T., Srivastava, S., Zetola, N., Williams, S., Sirugo, G., Gumbo, T., 2016. Artificial intelligence and amikacin exposures predictive of outcome in multidrug-resistant tuberculosis patients. Antimicrob. Agents Chemother accepted for publication.

Mulhall, A., de Louvois, J., Hurley, R., 1983. Chloramphenicol toxicity in neonates: its incidence and prevention. Br. Med. J. (Clin. Res. Ed.) 287, 1424–1427.

Pasipanodya, J., Gumbo, T., 2011. An oracle: antituberculosis pharmacokinetics-pharmacodynamics, clinical correlation, and clinical trial simulations to predict the future. Antimicrob. Agents Chemother 55, 24–34.

Pasipanodya, J.G., Srivastava, S., Gumbo, T., 2012. Meta-analysis of clinical studies supports the pharmacokinetic variability hypothesis for acquired drug resistance and failure of antituberculosis therapy. Clin. Infect. Dis 55, 169–177.

Pasipanodya, J.G., McIlrroe, H., Burger, A., Wash, P.A., Smith, P., Gumbo, T., 2013. Serum drug concentrations predictive of pulmonary tuberculosis outcomes. J. Infect. Dis 208, 1464–1473.

Pasipanodya, J.G., Mubanga, M., Nskekhe, M., Pandie, S., Magazi, B.T., Gumedze, F., Myer, L., Gumbo, T., Mayosi, B.M., 2015. Tuberculous pericarditis is multibacillary and bacterial burden drives high mortality. ELBioMedicine 2, 1634–1639.

Reed, M.D., 1999. Optimal sampling theory: an overview of its application to pharmacokinetic studies in infants and children. Pediatrics 104, 627–632.

Rey, E., Gendrel, D., Treluyer, J.M., Tran, A., Pariente-Khayat, A., d’Athis, P., Pons, G., 2001. Isoniazid pharmacokinetics in children according to acetylator phenotype. Fundam. Clin. Pharmacol 15, 355–359.

Sabbagh, A., Darlu, P., Crouau-Roy, B., Poloni, E.S., 2011. Arylamine N-acetyltransferase 2 (NAT2) genetic diversity and traditional subsistence: a worldwide population survey. PLoS One 6, e18507.

Sutherland, J.M., 1959. Fatal cardiovascular collapse of infants receiving large amounts of chloramphenicol. Am. J. Dis. Child 97, 761–767.

Wang, J., Endrenyi, L., 1992. A computationally efficient approach for the design of population pharmacokinetic studies. J. Pharmacokin. Biopharm 20, 279–294.

Weber, W.W., Cohen, S.N., 1968. The mechanism of isoniazid acetylation by human N-acetyltransferase. Biochim. Biophys. Acta 151, 276–278.

World Health Organization, 2010. Treatment of Tuberculosis in Children.