Citation: CPT Pharmacometrics Syst. Pharmacol. (2017) 6, 552–559; doi:10.1002/psp4.12208

ORIGINAL ARTICLE

Supervised Machine-Learning Reveals That Old and Obese People Achieve Low Dapsone Concentrations

RG Hall II[1], JG Pasipanodya[2], MA Swancutt[3], C Meek[1], R Leff[1] and T Gumbo[2,4]*

The human species is becoming increasingly obese. Dapsone, which is extensively used across the globe for dermatological disorders, arachnid bites, and for treatment of several bacterial, fungal, and parasitic diseases, could be affected by obesity. We performed a clinical experiment, using optimal design, in volunteers weighing 44–150 kg, to identify the effect of obesity on dapsone pharmacokinetic parameters based on maximum-likelihood solution via the expectation-maximization algorithm. Artificial intelligence-based multivariate adaptive regression splines were used for covariate selection, and identified weight and/or age as predictors of absorption, systemic clearance, and volume of distribution. These relationships occurred only between certain patient weight and age ranges, delimited by multiple hinges and regions of discontinuity, not identified by standard pharmacometric approaches. Older and obese people have lower drug concentrations after standard dosing, but with complex patterns. Given that efficacy is concentration-dependent, optimal dapsone doses need to be personalized for obese patients.

CPT Pharmacometrics Syst. Pharmacol. (2017) **6**, 552–559; doi:10.1002/psp4.12208; published online 13 July 2017.

Study Highlights

WHAT IS THE CURRENT KNOWLEDGE ON THE TOPIC?

☑ The pharmacokinetics of dapsone have not been characterized in overweight or obese people, a problem since the world population has become increasingly obese.

WHAT QUESTION DOES THIS STUDY ADDRESS?

☑ This study characterizes the population pharmacokinetics, including covariate effects, of dapsone in overweight and obese adults from the United States.

WHAT THIS STUDY ADDS TO OUR KNOWLEDGE

☑ Weight and age were identified as having an impact on dapsone pharmacokinetics, based on the use of Multivariate adaptive regression splines (MARS), an artificial intelligence algorithm that enables delineation of nonlinear relationships and high-order interactions. The effect was nonlinear and discontinuous at some values.

HOW MIGHT THIS CHANGE DRUG DISCOVERY, DEVELOPMENT, AND/OR THERAPEUTICS?

☑ MARS allows a simpler approach to identify pharmacokinetic covariates and their nonlinear interactions using smaller sample sizes, compared to standard approaches. The complex relationships identified for dapsone suggest the need to redefine doses in obese and older patients.

Dapsone [4–4′-diaminodiphenyl sulphone] was first synthesized by Fromm and Wittman in 1908.[1] It is one of the most commonly used pharmacological compounds on the globe, given its mode of activity as a competitive inhibitor of para-amino benzoic acid in the evolutionarily conserved folate synthesis pathway.[1] This mechanism gives it a broad spectrum of activity against *Mycobacterium leprae*, fungi, and parasites of global health significance such as *Plasmodium falciparum* and *Toxoplasma gondii*. In addition, dapsone inhibits neutrophil activity, and is prescribed for this immunomodulatory role in several dermatological disorders, as well as brown recluse spider bite wounds.[1] Given the important role of the relationship between drug concentrations achieved in patients and therapy failure as well as acquired drug resistance and other pharmacodynamics outcomes, it is imperative to identify the factors responsible for dapsone pharmacokinetic variability.[2–6] One factor that has emerged as an important pharmacokinetic covariate for many antiinfective agents is patient weight; specifically, weight increases beyond 66.3 kg.[7–10] This pharmacokinetic covariate is a concern since the human species is becoming obese, and most adults have weight above 66.3 kg. High-level obesity is now encountered across the globe, with an estimated 2 billion people now considered obese, who also live in the very countries where dapsone is used extensively.[11–14]

Obesity is defined using the interaction parameter, body mass index (BMI), with weight (kg) and height in meters (m) interacting. Overweight is a BMI $\geq$25 kg/m^2 and obesity a BMI $\geq$30 kg/m^2. In the case of dapsone, the impact of weight on drug concentrations has been assumed to be linear through all ranges, such that both systemic clearance and volume of distribution have been considered to increase linearly for each 1 kg increase above body mass

[1]Dose Optimization and Outcomes Research (DOOR) Program, School of Pharmacy, Texas Tech University Health Sciences Center, Dallas, Texas, USA; [2]Center for Infectious Diseases Research and Experimental Therapeutics, Baylor Research Institute, Baylor University Medical Center, Dallas, Texas, USA; [3]Department of Medicine, University of Texas Southwestern Medical Center, Dallas, Texas, USA; [4]Department of Medicine, University of Cape Town, Observatory, Cape Town, South Africa. *Correspondence: T Gumbo (tawanda.gumbo@bswhealth.org)
Received 24 January 2017; accepted 18 May 2017; published online on 13 July 2017. doi:10.1002/psp4.12208

(colloquial term "weight"), not BMI.[15,16] However, based on some of our prior studies, such linear relationships in biologic systems over a wide range of values should be suspect.[8,9,17–20] Moreover, the effect of weight on pharmacokinetics in the ranges of up to 150 kg, now increasingly encountered in patients, has hitherto not been investigated for dapsone. Prior studies have been of relatively underweight patients and mixtures of babies and adults.[15,16,21] Here, we optimized a clinical experimental study design to more precisely identify the effects of obesity on dapsone pharmacokinetic parameters.

Using the example of trimethoprim-sulfamethoxazole, we found that if experimental design is optimized and sampling times are optimized, machine-learning algorithms such as multivariate adaptive regression splines (MARS) can robustly identify covariates of pharmacokinetic parameters with as few as 36 patients.[22] Here we utilized the same strategy, design, and sample size for dapsone. We recruited subjects of different weights by stratifying by BMI groups and gender in the study design, so that enough subjects of different weight ranges of up to 150 kg could be studied. Second, we performed intensive pharmacokinetic sampling at eight timepoints in each patient over 72 h, to allow greater precision and minimum bias in identifying pharmacokinetic parameter estimates. Third, we utilized MARS to simultaneously identify linear and nonlinear interactions between potential covariates and pharmacokinetic parameters. All potential covariates were included *in toto* for a single implementation of the model, without presupposing the nature of any interactions. MARS, introduced by Jerome Freidman in 1991, is a machine-learning method that does not have linearity assumptions and does not need the data to follow a specific distribution.[23–25]

METHODS

Regulatory approval and ethics

The study was submitted and approved by Institutional Review Boards (IRBs) of the University of Texas Southwestern Medical Center (#032011-121) and the Texas Tech University Health Sciences Center (#A10-3591). The study was also registered on ClinicalTrials.gov (NCT01165840). All participants provided informed written consent prior to voluntary participation.

Recruitment

We recruited three groups of 12 people (six men, six women) based on BMI (normal weight, <25 kg/m^2; overweight and class I/II obesity, 25–40 kg/m^2; class III obesity, >40 kg/m^2) who were ≥ 18 years old. Participants were recruited from July 20th 2010 to May 2nd, 2012. Participants were ineligible if they 1) were pregnant or nursing, 2) had a medical contraindication or history of allergy to dapsone, sulfones, or sulfonamides, 3) had abnormal liver function tests (transaminases >10 times upper limit of normal, alkaline phosphatase >5 times upper limit of normal, or a total bilirubin >5 times upper limit of normal), or had glucose-6-phosphate dehydrogenase deficiency. With the same design, a sample size of 36 has been shown to be adequate in identifying the role of weight as pharmacokinetic covariate in the past.[8–10,17,22]

Study procedures

Participants were first screened in an outpatient clinic, where they had a medical history obtained and a physical examination performed. An informed consent was then obtained from participants who fulfilled the study criteria, and blood drawn for comprehensive metabolic profile, complete blood count, and glucose-6-phosphate dehydrogenase deficiency. The participants were then admitted to the University of Texas Southwestern Clinical Trials and Research Center (CTRC). Participants received a single oral dapsone dose of 100 mg under directly observed therapy. Vitals signs were examined and recorded 0.5, 1, 2, and 24 h after the dapsone administration. Blood draws (10 mL each) to determine dapsone concentrations were obtained at the following times: 0 h (predose), 1, 2, 4, 8, 16, and 24 h following dapsone administration, timepoints and number of samples identified in D-optimality in ADAPT (see below). Patients were discharged from the CTRC after 24 h. Two additional blood draws were performed 48 and 72 h postdose in the CTRC outpatient clinic, for measurement of dapsone concentration as well as for complete blood count and comprehensive metabolic profile. At each timepoint, blood samples were centrifuged for plasma separation with plasma stored at -80°C for transport to the Center for Pharmacology and Experimental Therapeutics to measure dapsone concentrations.

Adverse events were ascertained through patient interview and review of laboratory data. The IRB was notified of serious adverse events within 24 h. Data safety monitoring meetings included study coordinators and investigators (R.G.H., M.A.S., T.G.). These meetings occurred after the enrollment of 12 participants (September 2010), at the time of continuing review (April 2011, April 2012) and study enrollment closure (May 2012).

Drug concentration assay

Plasma samples were quantified for dapsone content using a validated, liquid chromatography/tandem mass spectrometry (LC-MS/MS) analytical method that included a Shimadzu (Columbia, MD) liquid chromatograph and AB Sciex (Foster City, CA) 5500 QTRAP tandem mass spectrometer. Unknown plasma samples (100 μL) were transferred into a 96-well plate with glass inserts to which 100 μL sulfamethizole (internal standard, IS) was added. The plate was removed and vortex mixed for 5 min. Then 500 μL of 80:20 ACN:MeOH was transferred to each insert, after which the samples were vortex mixed for 10 min then centrifuged for 10 min, 4°C at 3,300*g*. Next, 500 μL of the supernatant was transferred to a well plate with clean inserts and evaporated to dryness using a Labconco Centrivap. The dried samples were reconstituted with 100 μL of 50:50 MeOH:DIW and analyzed by reverse phase LC-MS/MS. Dapsone and IS were quantitated using positive electrospray ionization (+ESI) combined with multiple reaction monitoring (MRM) for the respective precursor→product ion combinations of 249.0→155.9 m/z for dapsone and 271.0→155.9 m/z for IS. The standard curve was linear ($r^2 = 0.9942$) and ranged from 5 ng/mL to 1,800 ng/mL.

Pharmacokinetic modeling

In all, 280 dapsone concentrations were comodeled using the ADAPT 5 program of D'Argenio et al.[26] Steps utilized have been detailed in the past.[3,8–10,17,22] Briefly, initial pharmacokinetic parameters were estimated using the standard two-stage approach, for one-, two-, and three-compartment models, with and without lag, and with first-order input and elimination. Estimates were then used for further analysis of the same compartment models using the maximum-likelihood solution via the expectation-maximization (MLEM) algorithm. The best compartment model was chosen based on Akaike information criteria, Bayesian information criteria (which penalizes for complexity), negative log-likelihoods, and parsimony, as described in our prior studies.[2–4,7–10,17,22]

Nonlinear system analyses to identify covariates

MARS was used to examine the nonlinear behavior of pharmacokinetic factors and their interactions with potential covariates. Potential covariates examined included anthropometric measures such as weight and height, demographic factors such as age and self-identified "racial" grouping, clinical factors, and laboratory test values for comprehensive metabolic profile and complete blood count. The potential covariates were also examined for linear and nonlinear interactions between themselves. All these variables were used to independently predict systemic clearance, apparent volume of distribution, and absorption constant for each participant who had been identified using MLEM. MARS uses basis functions (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. Individual estimates for the pharmacokinetic parameters were incorporated in the MARS algorithm. Overfitting procedures were used to grow large models with up to 15 BFs and up to two-way interactions between them, which were then pruned back using generalized cross-validation (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. Briefly, in view of our small sample size (35 patients), MARS built a sequence of models using all available data for learning purposes, after which the optimal model was determined based on GCV, which penalized model size and degrees of freedom parameter. We set the degree of freedom at 3 and also used naïve R-squared values to assess goodness and compare models. Change in both R^2 and mean squared error (MSE) values were used to examine for overfit in training and test samples. Overfit thresholds were identified by large difference between training MSE and test MSE, as well as flattening or U-curvature in test MSE with increase in model complexity. SALFORD predictive modeling software (v. 7, San Diego, CA) was used to run MARS models.

MARS is designed for situations in which the number of variables of interest (p) is relatively large when compared to the number of observations (n): colloquially defined as $p \gg n$.[25,27,28] MARS is a nonparametric modeling technique that builds piecewise regression models using BFs and then ranks the most influential predictors in each model using variance-bias tradeoffs. BF are data-driven and locally defined (not on the global data), such that the relationships between the dependent variable and the independent variables, i.e., predictors, is a set of discontinuous regression lines, each with its own slope, but joint end-to-end via knots or hinges. These knots represent local points in the data where the relationships or slope changes, thereby reflecting nonlinear patterns within the data. The output for MARS are given as BF to account for their nonlinear behavior of potential predictors (i.e., covariates) with the dependent variable, as well as the tendency to interact among themselves. The BFs take on several forms, which help explain what they mean. First, a BF can be a simple hinge function that takes the form "max (0, covariate – constant)" which means that the value of the expression is that of covariate (e.g., weight) minus a constant for all values of covariate-constant less than zero (i.e., for all values of covariate $<$ constant), otherwise it is zero value. The constant is at the hinge. Similar to its predecessor, a classification and regression tree that is commonly used for categorical dependent variables, MARS is used for a continuous dependent variable but also transforms data, groups, and then ranks predictors based on information theory. Thus, MARS can also be used for classification purposes and detection of high-order interactions.[28] Since MARS is a forward stepping spline regression model, it takes the functional form:

$$f(x) = \beta_0 + \Sigma \beta_m h_m x \quad (1)$$

where $h_m(x)$ is a function of the BF.[25] Lastly, MARS uses a generalized adjusted penalty and generalized cross-validation to select the best set of BF to include in the models. The net effect is to produce smooth curves and surfaces such that residuals sum of squares are minimized and then demonstrating up to three-way interactions between the dependent variable and predictors in 3D plots.

Statistical analysis

The D'Agostino–Pearson omnibus normality test was used to check whether data were normally distributed. Comparison of clinical and demographic factors between groups, shown in **Table 1**, was made using either the Student's *t*-test or Kruskal–Wallis test, while the Spearman rank correlation was used to demonstrate correlation between weight and age in all patients and then separately in males and females. Some continuous variables that were found to be important predictors in MARS were converted into categorical variables based on the BF obtained and then comparison between groups including gender made accordingly. Fisher's exact test or the χ^2 test was used to compare proportions between groups, when appropriate. All tests were two-sided; statistical significance was set at 0.05 and analyses were performed using STATA (v. 12, College Station, TX) and Prism (GraphPad v. 6.0, La Jolla, CA) software packages.

RESULTS

We recruited 35 subjects; 24 (69%) were overweight or obese. Ten of these 24 people (42%) had metabolic syndrome. The distribution of BMI and weight are shown in

Table 1 Demographic and clinical characteristics of 35 study participants[a]

Demographic variable	All	Female	Male	*P*-value
Study participants	35 (100)	18 (51)	17 (49)	
Self-identified race *n* (%)				
White	29 (83)	15 (83)	14 (82)	0.94
Non-white	6 (17)	3 (17)	3 (18)	
Co-morbid conditions *n* (%)	24 (69)	12 (67)	12 (71)	0.80
Metabolic syndrome *n* (%)	10 (29)			
Overweight or obese (BMI>=25) *n* (%)	24 (69)	12 (67)	12 (71)	0.80
Age (median [range]) years	36 (21–77)	31 (21–57)	43 (23–77)	0.10
Weight (median [range])	85.1 (70.8–130.5)	79.2 (58–112.9)	89.2 (74.8–138.4)	0.18
BMI (median [range])	30.2 (23.6–43.9)	30.05 (22.9–43.7)	30.2 (24.4–43.9	0.70
Serum creatine (median [range])	0.83 (0.68–0.99)	0.71 (0.63–0.76)	0.99 (0.9–1.06)	<0.001
Blood urea nitrogen (median [range]) mg/dl	13.0 (7–28)	12 (9–20)	13 (7–28)	0.07

[a]Figures in parenthesis denote *n* = absolute number, () are percent, unless stated otherwise. Laboratory reference values: serum creatinine: men, 0.7–1.3 mg/dL; women 0.6–1.1 mg/dL; blood urea nitrogen: 6–20mg/dL.

Figure 1; these were not normally distributed, as demonstrated by a $P < 0.05$ on D'Agostino K^2 normality test, fulfilling our experimental design intent. On the other hand, height (in meters), utilized in calculating the BMI metric, was normally distributed. **Table 1** shows the demographic characteristics of study participants; 51% were women and 49% men. There was no significant difference in weight, BMI, or the prevalence of comorbid conditions between men and women.

The concentration–time profiles in all subjects are shown in **Figure 2a**. The highest to lowest peak concentrations had a ratio of ~10; all patients had received the same 100 mg dose of dapsone, illustrating the adage that a drug's dose is a poor surrogate for drug concentration. MLEM revealed results for model selection shown in **Supplementary Table S1**. Based on Akaike information criteria, Bayesian information criteria, and parsimony, a one-compartment model was found to best describe dapsone pharmacokinetics. Observed vs. predicted concentrations for the 280 serum concentrations are shown in **Figure 2b** for the one-compartment model. The base model parameters mean value, percent relative standard error (RSE), and standard deviation as a percent covariance were 1.85 L/h (37.3%) and 37.6% for systemic clearance, 64.0 L (23.4%) and 50.5% for volume, and 2.57/hr (102%) and 94.7% for absorption constant, respectively. The half-life ± standard deviation was 25.2 ± 7.7 h, similar to observations in patients treated for a variety of infectious diseases.[1,15,21]

Implementation of MARS identified the results shown in **Table 2.** The table shows that patient weight, age, and blood urea nitrogen (BUN) were identified and ranked as important covariates. **Table 2** also shows the final MARS model equations describing the effects of the three covariates on each of three one-compartment model pharmacokinetic parameters. The most important predictor in MARS model in **Table 2** was set at 100% and the predictors remaining in the model after backward elimination weighed relative to the primary or most important predictor. Multiple variables that scored at 100% mean that they were equally important. Thus, in **Table 2**, for the absorption constant (K_a), BUN is the primary predictor but weight contributes to the variance of dapsone K_a by 44% relative to the BUN. With regard to systemic clearance, patient's weight and age equally contributed to the clearance variance. On the other hand, for volume, only patient weight contributes to the variability.

In the final equations (numbers 2–4) shown in **Table 2**, the covariates are given as BFs. **Figure 3** shows the

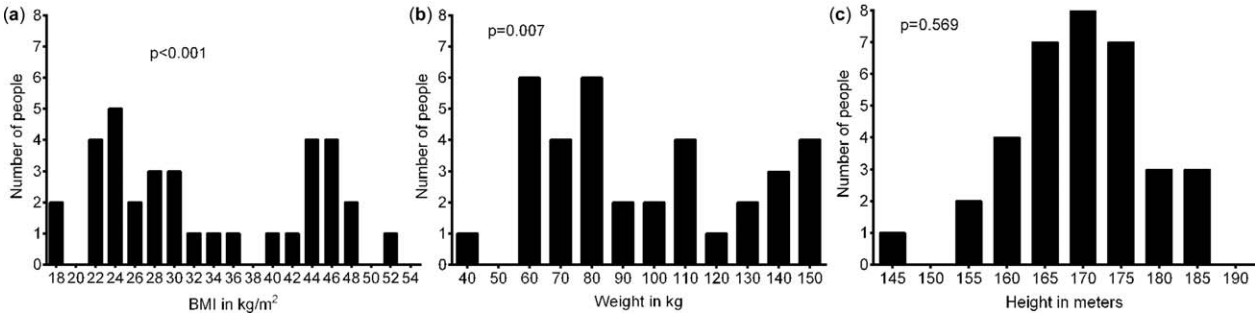

Figure 1 Distribution of weight, BMI, and height in study participants. We tested the demographic factors that are part of the definition of overweight and obesity status for normality using the D'Agostino K^2 normality test. (**a**) BMI was not normally distributed, based on $P < 0.05$. This reflected that we had succeeded in the experimental study design of recruiting participants whose BMIs were equally spread out between the 40 and 150 kg range. (**b**) Similarly, weight was not normally distributed, based on $P = 0.007$. (**c**) However, height, which is part of the definition of BMI, was found to be normally distributed.

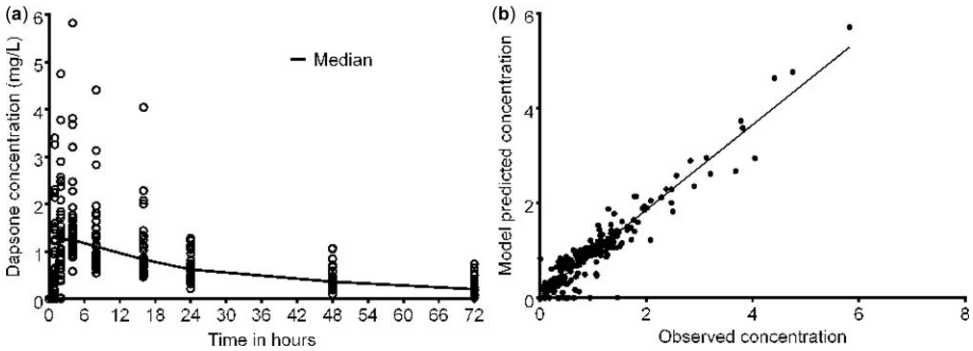

Figure 2 Concentration–time profile of dapsone in 35 participants. **(a)** All patients received the same dose of 100 mg a day. However, the concentrations achieved at each timepoint varied widely, so that the peak concentrations varied 1,000% between the lowest and highest concentrations. The median is shown, and illustrates why measures of central tendency such as median and mean ± standard deviation (which was 1.63 ± 1.03 mg/L) are poor descriptors of drug concentration estimates and their variability. Indeed, at no timepoint (except 24 h timepoint; $P = 0.08$) were the concentrations even normally distributed based on the D'Agostino and Pearson omnibus normality test. **(b)** Observed vs. predicted dapsone concentrations for a one-compartment model, showing an $r^2 = 0.91$. The slope was 0.90 ± 0.02, and thus differed slightly from 1.00.

relationship between weight and volume of distribution, reflecting BF_8 [max(0, Weight – 69.8)] and BF_{10} [max (0, Weight – 74.8)]. The figure shows multiple hinges delimiting different slopes in the relationship, a perfect example the nonlinearity. A positive coefficient for a BF means that higher values of the covariate leads to higher values for the pharmacokinetic parameter, as shown for example in **Figure 4**. However, a negative coefficient means two things: either there is a reduction in the slope between the covariate and the dependent variable or there is a negative correlation between covariate and dependent variable (pharmacokinetic parameter).

The second form of a BF is shown in **Figure 4**, which depicts BF_3 in **Table 2** for absorption constant (K_a), whereby the BF is a product of hinge functions or of hinge functions and a BF. BF_3 is the product of a weight vs. K_a hinge function (max (0, 63.7 – Weight)) and BF_1 (i.e., max (0, BUN – 7)). This BF is shown in the 3D covariate interaction plot of **Figure 4**. The expressions are conditional on the hinges from the parent BF. The interaction detection and selection of predictors is further shown in **Table 3**. The r^2 reported in **Table 3** were obtained after general cross-validation and thus give both reliability of the parameter estimates and robustness of the final model. As shown, weight explained 64% of the variance in dapsone volume.

In bivariate analysis, the median (interquartile) dapsone volume was significantly lower in women (66.43L; 39.89, 84.77) than men (86.60L; 71.7, 92.45), $P = 0.027$. However, this gender difference was lost in multivariate analyses of dapsone volume and therefore is not shown in the BF shown in **Table 3**. In addition, gender differences were not revealed when clearance and absorption constant values were examined in bivariate analyses. Median (interquartile) clearance was 1.92 (1.59, 2.26) L/h in females and 2.07 (1.95, 2.14) L/h in males.

DISCUSSION

Artificial intelligence algorithms, especially machine learning, are being increasingly employed to examine biological data.[29–39] We started utilizing these for pharmacometric analyses almost 10 years ago.[3,7,10] One of the algorithms, MARS, was used here to provide an agnostic (i.e., nonprespecified) identification of covariates responsible for interindividual variability of dapsone absorption, volume,

Table 2 Basis functions used to model the population pharmacokinetics of dapsone in adults.

	Variable importance			
Outcome (Y)	**Variable**	**Rank**	**Basis functions definitions**	**Model equations***
K_a (hr^{-1})	BUN (mg/dL)	100%	BF_1 = max (0, BUN – 7);	$Y = 3.967 - 0.062*BF_3 + 0.162*BF_5$ (2)
	Weight	44%	BF_3 = max (0, 63.7 – Weight)*BF_1;	
			BF_5 = max (0, 74.8 – Weight)	
Clearance (L/hr)	Weight	100%	BF_2 = max (0, 77.2 – Weight);	$Y = 2.048 - 0.002BF_3 - 0.005BF_4$ (3)
	Age	100%	BF_3 = max (0, Age – 27)*BF_2;	
			BF_4 = max (0, 27 – Age)*BF_2;	
Volume (L)	Weight	100%	BF_8 = (0, Weight – 69.8);	$Y = 36.648 + 8.894BF_8 - 8.818BF_{10}$ (4)
			BF_{10} = max (0, Weight – 74.8);	

BUN, blood urea nitrogen levels in mg/dl on initial blood sampling; weight, patient's weight in kg; variable importance denotes the weighted value of each variable in predicting the given dapsone pharmacokinetic parameter; the model equation is given by linear combination of basis functions. Up to a maximum of 15 basis functions were derived from the data allowing for a minimum of 3 (~10%) of observations per knots and a maximum of 2nd order interactions; then a backward step elimination selection process was used in the regression with moderate penalty for model complexity.

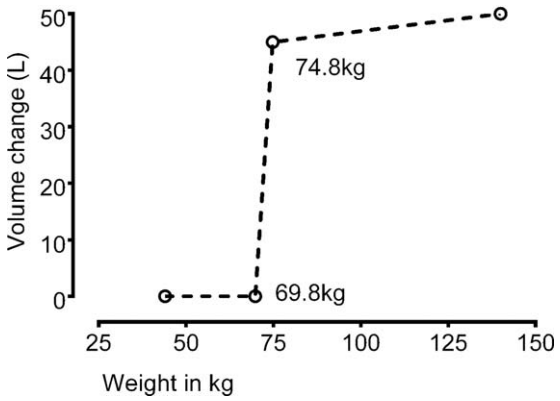

Figure 3 Depiction of basis functions of weight vs. volume of distribution. The figure illustrates two basis functions, BF_8 and BF_{10}, in Table 3. Depicted are hinges at patient weights of ~70 kg and 75 kg, illustrating different slopes for the relationship between weight and volume.

and clearance, as well as their slopes. MARS evaluates both linear and nonlinear relationships simultaneously, as well as high-order interactions of covariates among themselves and with primary pharmacokinetic parameters. In addition, it allows for regions where the relationships are discontinuous, or do not exist. Thus, MARS is able to identify multiple hinges delineating different slopes of the relationship between pharmacokinetic parameters and covariates for monotonic and nonmonotonic functions. Moreover, the main goal of machine-learning methods such as MARS is predictive accuracy, and its use of cross-validation techniques is highly reproducible in independent samples.[25] We propose its routine use as a simplification of methods to identify significant covariates in general population pharmacokinetic analyses in a nonbiased and nonlinear fashion. This approach has important implications for the design of early phase clinical studies. Sample size, and thus risk of adverse events to patients, can be limited by application of optimal experimental design strategies, and the application of methods such as MARS for covariate selection. Further precision in estimating pharmacokinetic parameters would come from applying optimal sampling theory, which minimizes the number of samples to be drawn while improving accuracy, which could also reduce the patient sample size.

Using this approach, we found weight to be a significant covariate for dapsone absorption, volume, and systemic clearance. In the past, the role of weight has been examined assuming specific fractal geometry considerations for several antiinfective agents.[8–10,17–20,22] Our current approach with MARS bypasses the need to fix particular power laws with weight, by virtue of the fact that the final MARS equation derives its own coefficients, similar to letting exponents in fractal geometry "float." In other words, we did not fit any specific power law. Thus, it was our aim to be able to interpret both the model coefficients for inference and for predictions. Up to now, the understanding had been that dapsone systemic clearance increases linearly by 0.03 L/hr and volume by 0.7 L for each kg increase above 62 kg.[15] Here we show that there were multiple hinges for the relationship between weight and systemic clearance (77.2 kg) and with volume (69.8 kg and 74.8 kg), which relationships were modified (i.e., interacted) by different hinges of age; the effect of weight on absorption constant was conditional on BUN. Interestingly, the weight hinges are all within close range of the value of 66.3 kg; the reasons are unclear. Thus, weight is an important covariate of primary dapsone pharmacokinetic parameters, but the relationships are complex and modified by the effect of age and renal function. To optimally treat obese patients with dapsone, therefore, we will need to personalize dosing recommendations by the patient's weight in order to account for their unique ability to absorb, distribute, metabolize, and eliminate this compound. One approach, which utilizes engineering simulations to achieve that, has been detailed before with other antiinfective agents.[40,41]

Dapsone systemic clearance had a hinge at the age of 27 years. This illustrates one of the advantages of the agnostic approach we took. First, BF_3 and BF_4 interact (modify) the effect of weight and are shown as multiplied with BF_2. Second, BF_3 and BF_4 are termed mirror BFs: there is a reversal of slope below the hinge age of 27 years compared to above, and are thus nonlinear. Third, the hinge age of 27 years differs from the usual, and somewhat arbitrary, categorization of children as below 18 years old, and old age as beginning in the 60s. In pharmacokinetic terms, MARS avoids an arbitrary "typical" value for covariate, and instead identifies it from the data. Thus, one of the advantages of MARS is identification of such threshold values for covariates, without prior assumptions based on social and political categorization. Thus, the age at which a pharmacokinetic parameter changes is driven by the data for that pharmaceutical compound for that patient group, without the need for prior assumptions that categorize patients into convenient but arbitrary age groups.

Our study has strength and several limitations. The machine-learning algorithms we used do not infer causality or mechanism. However, when combined with an experimental design, the approach gives insights into important nonlinear interactions that can be explored further as

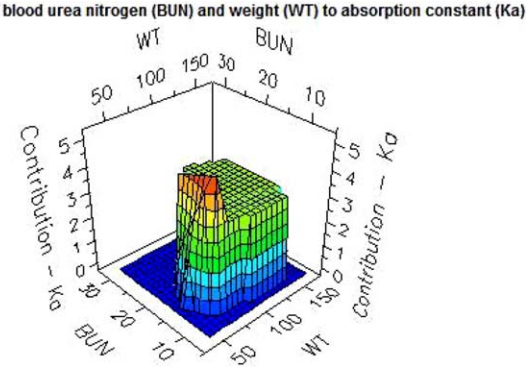

Figure 4 The joint contribution of blood urea nitrogen and weight to dapsone absorption constant. The 3D plot demonstrates the interaction of hinge function and basis function for interaction of weight (WT) and blood urea nitrogen (BUN) to affect the absorption constant (K_a).

Table 3 Interactions detected and selection of predictors for dapsone pharmacokinetics based on MARS

Outcome (Y)	r^2	Basis function[a]	Coefficient	Variable	Sign	Parent		Knot
						Sign	Variable	
K_a (hr^{-1})	0.36	0	3.967					
		3	−0.062	Weight	+		BUN	63.7
		5	0.162	Weight	+			74.8
Clearance (L/hr)	0.48	0	2.049					
		3	−0.002	Age	+	–	Weight	27
		4	−0.005	Age	–-	–	Weight	27
Volume (L)	0.64	0	36.648					
		8	8.894	Weight	+			69.8
		10	−8.818	Weight	– -			74.8

K_a, absorption constant, BUN, blood urea nitrogen.
[a]The basis functions are defined above; knot denotes the hinge or level at which values of the coefficient or slope of the regression changes.

primary hypotheses. In this prospective clinical experiment, by design we recruited equal numbers of patients into different weight bands (as opposed to recruitment of all comers), deliberately recruited equal numbers for either gender, and the pharmacokinetic sample number and sampling times were identified using optimal sampling theory and D-optimality in ADAPT. Another limitation was that in our study design, a single dapsone dose was administered to persons without active disease. The pharmacokinetic parameters of dapsone could be altered by the disease process it is administered to treat, by disease severity, and at steady state. However, the pharmacokinetic parameter estimates in our base model are similar to those observed by other investigators using multiple doses and/or evaluating patients with active disease.[15,16] We also did not evaluate routes of disposition or the effect of dapsone metabolites, so that we did not evaluate the physiological reasons why obese patients have higher drug metabolism.

Acknowledgments. R.G.H. was supported by Grant KL2RR024983 (“North and Central Texas Clinical and Translational Science Initiative”) to UT Southwestern Medical Center from the National Center for Research Resources of the National Institutes of Health (NIH). Recruitment, admission of volunteers for the study, and collection of blood samples was performed by UT Southwestern Medical Center’s Clinical and Translational Research Center personnel within the Center, which is supported by the NIH CTSA grant UL1 RR024982. J.G.P. and T.G. were supported by R01AI079497 from the National Institute of Allergy and Infectious Diseases of the NIH.

Conflict of Interest. T.G. is a scientific advisor for LuminaCare Solutions and LEAF Pharmaceuticals, and is on the board of directors for Kiara Health. T.G. also founded Jacaranda Biomed Inc., and an investor in Marula Cosmetics Inc. R.G.H. is an advisory board member for Genetech and has received research funding from Merck. The other authors have no conflicts of interest to disclose.

Author Contributions. R.G.H., J.G.P., M.A.S., C.M., R.L., and T.G. wrote the article; R.G.H., R.L., and T.G. designed the research; R.G.H., M.A.S., and T.G. performed the research; J.G.P., C.M., R.L., and T.G. analyzed the data.

Accession Numbers. ClinicalTrials.gov #NCT01165840; other study ID number: AMAIRB97, 5UL1RR024982–02.

1. Gumbo, T. Chemotherapy of tuberculosis, *Mycobacterium avium* complex disease, and leprosy. In: Brunton LL, Chabner B, Knollmann B, editors. *Goodman & Gilman’s The Pharmacological Basis of Therapeutics*. New York: McGraw Hill Medical, 2011.
2. Srivastava. S., Pasipanodya, J.G., Meek, C., Leff, R. & Gumbo, T. Multidrug-resistant tuberculosis not due to noncompliance but to between-patient pharmacokinetic variability. *J. Infect. Dis.* **204**, 1951–1959 (2011).
3. Pasipanodya, J.G, McIlleron, H., Burger, A., Wash, P.A., Smith, P. & Gumbo, T. Serum drug concentrations predictive of pulmonary tuberculosis outcomes. *J. Infect. Dis.* **208**, 1464–1473 (2013).
4. Pasipanodya, J.G., Srivastava, S. & Gumbo, T. 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 (2012).
5. Mosha. D. *et al.* Population pharmacokinetics and clinical response for artemether-lumefantrine in pregnant and nonpregnant women with uncomplicated *Plasmodium falciparum* malaria in Tanzania. *Antimicrob. Agents Chemother.* **58**, 4583–4592 (2014).
6. Dorlo, T.P. *et* al. Failure of miltefosine in visceral leishmaniasis is associated with low drug exposure. *J. Infect. Dis.* **210**, 146–153 (2014).
7. Gumbo, T., Hiemenz, J., Ma, L., Keirns, J.J., Buell, D.N. & Drusano, G.L. Population pharmacokinetics of micafungin in adult patients. *Diagn. Microbiol. Infect. Dis.* **60**, 329–331 (2008).
8. Hall, R.G., Swancutt, M.A & Gumbo, T. Fractal geometry and the pharmacometrics of micafungin in overweight, obese, and extremely obese people. *Antimicrob. Agents Chemother.* **55**, 5107–5112 (2011).
9. Hall, R.G., Swancutt, M.A., Meek, C., Leff, R. & Gumbo, T. Weight drives caspofungin pharmacokinetic variability in overweight and obese people: fractal power signatures beyond two-thirds or three-fourths. *Antimicrob. Agents Chemother.* **57**, 2259–2264 (2013).
10. Jain, M.K., Pasipanodya, J.G., Alder, L, Lee, W.M. & Gumbo, T. Pegylated interferon fractal pharmacokinetics: individualized dosing for hepatitis C virus infection. *Antimicrob. Agents Chemother.* **57**, 1115–1120 (2013).
11. Flegal, K.M., Carroll, M.D., Ogden, C.L. & Curtin, L.R. Prevalence and trends in obesity among US adults, 1999–2008. *JAMA* **303**, 235–241 (2010).
12. Filozof, C., Gonzalez, C., Sereday, M., Mazza, C., & Braguinsky, J. Obesity prevalence and trends in Latin-American countries. *Obes. Rev.* **2**, 99–106 (2001).
13. Kruger, H.S., Puoane, T., Senekal, M. & van der Merwe, M.T. Obesity in South Africa: challenges for government and health professionals. *Public Health Nutr.* **8**, 491–500 (2005).
14. Swinburn, B.A. *et al.* The global obesity pandemic: shaped by global drivers and local environments. *Lancet* **378**, 804–814 (2011).
15. Simpson, J.A. *et al.* Population pharmacokinetic and pharmacodynamic modelling of the antimalarial chemotherapy chlorproguanil/dapsone. *Br. J. Clin. Pharmacol* **61**, 289–300 (2006).
16. Moura, F.M., Dias, R.M., Araujo, E.C., Brasil, L.M., Ferreira, M.V. & Vieira, J.L. Dapsone and body mass index in subjects with multibacillary leprosy. *Ther. Drug Monit.* **36**, 261–263 (2014).
17. Hall, R.G., Swancutt, M.A., Meek, C., Leff, R.D. & Gumbo, T. Ethambutol pharmacokinetic variability is linked to body mass in overweight, obese, and extremely obese people. *Antimicrob. Agents Chemother.* **56**, 1502–1507 (2012).
18. West, G.B., Brown, J.H. & Enquist, B.J. The fourth dimension of life: fractal geometry and allometric scaling of organisms. *Science* **284**, 1677–1679 (1999).
19. West, G.B., Savage, V.M., Gillooly, .J, Enquist, B.J., Woodruff, W.H. & Brown, J.H. Physiology: Why does metabolic rate scale with body size? *Nature* **421**, 713 (2003).

20. Mandelbrot, B.B. *The Fractal Geometry of Nature*. New York: W.H. Freeman and Company; 1982.
21. Gatti, G. *et al.* Population pharmacokinetics of dapsone administered biweekly to human immunodeficiency virus-infected patients. *Antimicrob. Agents Chemother* **40**, 2743–2748 (1996).
22. Hall, R.G., Pasipanodya, J.G., Meek, C., Leff, R.D., Swancutt, M. & Gumbo, T. Fractal geometry-based decrease in trimethoprim-sulfamethoxazole concentrations in overweight and obese people. *CPT Pharmacometrics Syst. Pharmacol.* **5**, 674–681 (2016).
23. Friedman, J.H. Multivariate adaptive regression splines. *Ann. Statist.* **19**, 1–68 (1991).
24. Friedman, J.H. & Roosen, C.B. An introduction to multivariate adaptive regression splines. *Stat. Methods Med. Res.* **4**, 197–217 (1995).
25. Hastie, T., Tibshirani, R. & Friedman, J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. New York: Springer; 2001.
26. D'Argenio DZ, Schumitzky A, Wang X. *ADAPT 5 User's Guide: Pharmacokinetic/ Pharmacodynamic Systems Analysis Software*. 2009. Los Angeles: Biomedical Simulations Resource.
27. Hastie T, Tibshirani R, Friedman J. Note on "Comparison of model selection for regression" by Vladimir Cherkassky and Yunqian Ma. *Neural. Comput.* **15**, 1477–1480 (2003).
28. Chigutsa, E. *et al.* Impact of nonlinear interactions of pharmacokinetics and MICs on sputum bacillary kill rates as a marker of sterilizing effect in tuberculosis. *Antimicrob. Agents. Chemother.* **59**, 38–45 (2015).
29. Nesbeth, D.N. *et al.* Synthetic biology routes to bio-artificial intelligence. *Essays Biochem.* **30**, 381–391 (2016).
30. Beam, A.L. & Kohane, I.S. Translating artificial intelligence into clinical care. *JAMA* **316**, 2368–2369 (2016).
31. Chuang, L.C. & Kuo, P.H. Building a genetic risk model for bipolar disorder from genome-wide association data with random forest algorithm. *Sci. Rep.* 2017 **7**, 39943 (2017).
32. Lin, H.Y. *et al.* 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 (2008)
33. Uppu, S., Krishna, A. & Gopalan, R. A review of machine learning and statistical approaches for detecting SNP interactions in high-dimensional genomic data. *IEEE/ ACM Trans Comput Biol Bioinform.* doi: 10.1109/TCBB.2016.2635125 (2016) [Epub ahead of print].
34. Ju, H., Brasier, A.R., Kurosky, A., Xu, B., Reyes, V.E., & Graham, D.Y. Diagnostics for statistical variable selection methods for prediction of peptic ulcer disease in *Helicobacter pylori* infection. *J. Proteomics Bioinform.* **7**, 1000307 (2014).
35. Swaminathan, S. *et al.* Drug concentration thresholds predictive of therapy failure and death in children with tuberculosis: bread crumb trails in random forests. *Clin. Infect. Dis.* **63**(suppl 3), S63–S74 (2016).
36. Rogers, Z. *et al.* The non-linear child: ontogeny, isoniazid concentration, and NAT2 genotype modulate enzyme reaction kinetics and metabolism. *EBioMedicine.* **11**, 118–126 (2016).
37. Modongo, C. *et al.* Artificial intelligence and amikacin exposures predictive of outcomes in multidrug-resistant tuberculosis patients. *Antimicrob. Agents Chemother.* **60**, 5928–5932 (2016).
38. Cho, H., Berger, B. & Peng, J. Compact integration of multi-network topology for functional analysis of genes. *Cell Syst.* **3**, 540–548 (2016).
39. Montaño-Gutierrez, L.F., Ohta, S., Kustatscher, G., Earnshaw, W,C. & Rappsilber J. Nano random forests to mine protein complexes and their relationships in quantitative proteomics data. *Mol Biol Cell.* pii: mbc.E16-06-0370. doi: 10.1091/mbc.E16-06-0370 (2017) [Epub ahead of print].
40. Pasipanodya, J., Hall, R. & Gumbo, T. In silico-derived bedside formula for individualized micafungin dosing for obese patients in the age of deterministic chaos. *Clin. Pharmacol. Ther.* **97**, 292–297 (2015).
41. Pasipanodya, J. & Gumbo, T. An oracle: antituberculosis pharmacokinetics-pharmacodynamics, clinical correlation, and clinical trial simulations to predict the future. *Antimicrob. Agents. Chemother* **55**, 24–34 (2011).

Supplementary information accompanies this paper on the *CPT: Pharmacometrics & Systems Pharmacology* website (http://psp-journal.com)