Published in final edited form as:

Stat Med. 2016 October 15; 35(23): 4077–4092. doi:10.1002/sim.7005.

Measuring the Individual Benefit of a Medical or Behavioral Treatment Using Generalized Linear Mixed-Effects Models

Francisco J. Diaz

Department of Biostatistics, The University of Kansas Medical Center, Mail Stop 1026, 3901 Rainbow Blvd., Kansas City, KS 66160, United States

Francisco J. Diaz: fdiaz@kumc.edu

Abstract

We propose statistical definitions of the individual benefit of a medical or behavioral treatment and of the severity of a chronic illness. These definitions are used to develop a graphical method that can be used by statisticians and clinicians in the data analysis of clinical trials from the perspective of personalized medicine. The method focuses on assessing and comparing individual effects of treatments rather than average effects, and can be used with continuous and discrete responses, including dichotomous and count responses. The method is based on new developments in generalized linear mixed-effects models, which are introduced in this article. To illustrate, analyses of data from the STAR*D clinical trial of sequences of treatments for depression and data from a clinical trial of respiratory treatments are presented. The estimation of individual benefits is also explained.

Keywords

Chronic diseases; clinical trials; disease severity; empirical Bayesian prediction; generalized linear mixed-effects models; random effects linear models

1. Introduction

Health professionals and statisticians involved in personalized medicine (PM) for chronic disorders face many challenges, including the need to develop methods for assessing individual variations in clinical responses to medical or behavioral treatments (MBTs) [1]. Despite the fact that the individual benefit of a treatment is a central concept in PM, however, the analysis of data from clinical trials (CTs) usually focuses on average treatment effects. Although researchers may conduct subgroup and interaction analyses secondarily, they are usually limited by lack of power and by the difficulty in knowing or measuring all relevant variables. Moreover, current statistical theory and practice lack a statistical or mathematical definition of individual benefit and methods specifically designed for operationally measuring this important concept.

Conflicts of interest. The Author declares that there is no conflict of interest.

The main objectives of this article are to propose a statistical definition of the individual benefit of an MBT administered to a patient with a chronic disorder and to suggest a graphical method for assisting statisticians and clinicians in the data analysis of CTs from a PM perspective. Our method focuses on the analysis of individual rather than average effects of treatments. The definition and the graphical method that it implies are based on new developments in generalized linear mixed-effects models (GLMMs), which are described here and have not been published before. Our method can be used with continuous and discrete responses, including dichotomous and count responses. The estimation of individual benefits is also described.

The development of personalized medicine (PM) is spurred by two commonly accepted tenets: (1) that some MBTs may be detrimental rather than beneficial to some people, and (2) that the degrees of benefit provided by the same treatment may vary in different people [1]. These tenets are not always demonstrated empirically, however, and in many cases the evidence presented to support them is questionable [2, 3]. This is because researchers often overlook the importance of variance components in the design and interpretation of clinical trials from a PM frame of reference [3]. For instance, researchers frequently are unaware of the impossibility of detecting subject-by-treatment interactions when they do not take subject-level repeated measures within and across treatments or do not analyze them appropriately. This, despite the fact that such interaction is a central concept of PM [2, 3, 4]. As a result, repeated measures are underutilized in clinical research, and parallel-group designs are still the most common ones in placebo-controlled or comparative-effectiveness studies.

Some authors have proposed graphical methods for demonstrating the existence of differential treatment effects not necessarily explained by measured variables. For cross-over designs with two treatments and four periods, Senn [3] plots the difference between the second pair of responses against that of the first pair. A high correlation between differences suggests a strong element of personal response. In the same design, Schall *et al.* [5] plot studentized residuals based on differences between the averages of measures from the two treatments against a subject identifier. An outlying point indicates a strong subject-by-formulation interaction. In meta-analyses, several authors have graphed mortality or event rates from study arms under treatment against those from control arms [6, 7, 8]. They aver differential treatment effects if the treatment-control relationship is not parallel to the identity line. This latter approach, however, may be hampered by regression to the mean and ecological fallacy since it is based on aggregated data [9]. Thus, conclusions may not hold true at the individual level.

Boissel *et al.* [8] defines a treatment benefit as the difference in the rate of the disease outcome between untreated and treated subjects. Using this definition and a Cox regression model of time to cardiovascular events with fixed regression coefficients, Li *et al.* [10] computes confidence intervals for the individual benefit of some drugs based on the subject's prognostic factors. The prediction, though, does not take into account the potential variability of regression coefficients in the subject population.

GLMMs are useful tools for the analysis of repeated measures [11, 12]. Considerable research suggests that regression models with random effects (REs) can be used to establish a solid paradigm for the construction of the mathematics and statistics of PM research and practice, especially in the treatment of chronic diseases [4, 13-18]. The fact that GLMMs have concepts that allow describing patient populations as a whole (the fixed effects) and, simultaneously, concepts that allow describing patients as individuals (the REs) suggests that these models contain the key ideas for providing PM with a rigorous mathematical language [4, 16-19]. Underlying this is the belief that the variability of a random coefficient is not just a mathematical artifact to control for patients' heterogeneity but also the result of real variation in the biological and environmental factors that have made humans develop as individuals [2, 4, 16, 17, 20]. Moreover, solid empirical and theoretical work by the Sheiner School of Pharmacology [13-15] and others [17, 18, 21] shows that a combination of mixed models with empirical Bayesian feedback (EBF) can be employed successfully in pharmacotherapy individualization and that EBF is well anchored in standard decision theory. Thus, both biological and mathematical arguments support the development of methodological instruments for PM based on GLMMs.

In Section 2, we present the overarching statistical models of medical or psychological care for chronic disorders guiding this article. One can view this section as a useful description of GLMMs from a PM viewpoint in the context of clinical trials. Section 3 contains the main ideas of this article and introduces statistical definitions of the benefit function of an MBT and the illness severity function. It also has graphical examples of how these functions represent many clinical phenomena. The benefit function allows modeling the benefit of an MBT as it varies with (as a function of) the patient or subject. To illustrate how to use our graphical tools in practice, we analyze data from clinical trials of treatments for depression (Section 4) and respiratory illnesses (Section 5). Section 6 presents an approach to the estimation of an individual benefit. Stata code is in Appendices 1 and 2 (Tables 1-2) and in the online material (Table 3 and figures).

2. Personalized medicine models

Let Y be a (possibly transformed) variable measuring or indicating some aspect of the state of a particular chronic disorder of a subject (or patient). Assume the goal of an MBT in a CT or clinical practice is to modify the value of Y . Examples of Y are: 1) a dichotomous indicator of whether or not a child with autism has appropriate expressive language skills; 2) the positive or negative subscales of the PANSS syndrome scale, which assess schizophrenia severity; 3) blood glucose or cholesterol concentrations, which partially assess the metabolic syndrome; 4) the 17-item Hamilton Rating Scale for Depression (HAM-D17); or 5) the total number of abstemious days (NADs) in a month, which assesses chronic alcoholism. For some of these measurements, the MBT's goal is usually to reduce the value of the measurement to the minimum possible value (e.g. PANSS, HAM-D17), or to increase it to the maximum possible value (e.g. NADs). But for some others, the goal is that the value lies inside a prespecified range or set (e.g. glucose, cholesterol concentration, dichotomous language indicator).

Let Q symbolize a particular MBT or sequence of MBTs. Before a particular subject ω is administered Q , a baseline value of Y is measured $k_{0,\omega}$ times. A measured baseline value of Y is denoted by $Y_{0,\omega,i}$ and assumed to satisfy

$$E[Y_{0,\omega,i}|\Lambda_\omega] = g^{-1}(\Lambda_\omega), i=1, \dots, k_{0,\omega}, \quad (1)$$

where Λ_ω is a constant number that characterizes subject ω and g is a known link function. Λ_ω is viewed as a constant number representing the state of the disease of subject ω before the MBT is started. But, Λ_ω varies from subject to subject, that is, Λ_ω is a particular realization of a random variable Λ' whose distribution is completely determined by variability in the subject population. It is additionally assumed the conditional distribution of $Y_{0,\omega,i}$ given Λ' belongs to an exponential family $\mathcal{F}$. For instance, $\mathcal{F}$ may be the family of normal, Poisson or Bernoulli distributions.

After administering Q to subject ω and the subject's response is stabilized, Y is measured $k_{1,\omega}$ times. Each measure is denoted by $Y_{Q,\omega,j}$ and it is assumed there is a number $\beta_{Q,\omega}$ that satisfies

$$E[Y_{Q,\omega,j}|\Lambda_\omega, \beta_{Q,\omega}] = g^{-1}(\Lambda_\omega + \beta_{Q,\omega}), j=1, \dots, k_{1,\omega}. \quad (2)$$

We stress that both Λ_ω and $\beta_{Q,\omega}$ are *constant numbers* in subject ω (they do not change during the clinical trial). We also assume

$$\Lambda_\omega = \alpha_\omega + \boldsymbol{\lambda}^T \mathbf{X}_\omega, \quad (3)$$

where $\mathbf{X}_\omega$ is a vector of baseline clinical, environmental, genetic, biological or demographic covariates, $\boldsymbol{\lambda}$ has the same value for all subjects, but α_ω is a characteristic constant of subject ω which varies from subject to subject. Thus, α_ω is a realization of a population-level random variable α .

In some subject populations, a covariate vector $\mathbf{Z}_\omega$ moderates the effect of Q . This vector may contain some of the covariates in $\mathbf{X}_\omega$. In these cases, we additionally assume

$$\beta_{Q,\omega} = \gamma_\omega + \boldsymbol{\theta}^T \mathbf{Z}_\omega, \quad (4)$$

where $\boldsymbol{\theta}$ has the same value for all subjects, but γ_ω is a characteristic constant of subject ω which varies from subject to subject. That is, γ_ω is a realization of a population-level random variable γ . It is possible to assume that $\boldsymbol{\lambda}$, $\boldsymbol{\theta}$ or both vary from subject to subject, but the simpler model that assumes fixed $\boldsymbol{\lambda}$ and $\boldsymbol{\theta}$ will suffice to introduce our ideas.

Definition 2.1. If $\beta_{Q,\omega}$ has the same value for all subjects, that is, if $\beta_{Q,\omega}$ does not depend on ω , then Eqs. (1)-(4) together are called a *one-dimensional personalized medicine* (1-PM) model of treatment Q . And if $\beta_{Q,\omega}$ varies from subject to subject, then Eqs. (1)-(4) are a *two-dimensional personalized medicine* (2-PM) model of treatment Q . Here, $\beta_{Q,\omega}$ is viewed as a particular realization of a random variable β_Q^* that reflects variation at the subject population level. In the 1-PM model, β_Q^* is a constant. In both models, we additionally assume the conditional distribution of $Y_{Q,\omega,j}$ given (Λ', β_Q^*) belongs to $\mathcal{F}$.

For estimation purposes, we followed the classical formulation of GLMMs and assumed that Λ' in the 1-PM model, or (Λ', β_Q^*) in the 2-PM model, are normally distributed given X_ω and Z_ω . This assumption, however, is not essential for our proposed approach and can be relaxed if robust estimation methods are used (for instance, the methods in Wang *et al.* [22]).

In contrast to the traditional interpretation of random parameters presented in the literature of random-effects linear models (RELM) or GLMMs, the randomness of Λ' (or α), or that of β_Q^* (or γ) in a 2-PM model, are not viewed here as the result of a random selection of subjects for a CT, or of a randomization of subjects to experimental MBTs, but the result of real-life variation in the biological, environmental and clinical dimensions that shape subjects' individualities [2, 4, 16, 18]. That is, Λ' (and β_Q^* in the 2-PM model) are unobserved but real dimensions with probability distributions that reflect the underpinnings of the observed clinical phenomenon. The actual values of these dimensions in a particular individual, that is, Λ_ω and $\beta_{Q,\omega}$ can be predicted using empirical Bayes estimators which are also called best linear unbiased predictors (BLUPs) in the context of RELMs [11, 12]. Note also that the conditional expectations $E[\cdot | \Lambda_\omega]$ and $E[\cdot | \Lambda_\omega, \beta_{Q,\omega}]$ can be interpreted as expected values “given subject ω ”.

3. Severity and benefit functions

This section proposes a statistical definition of the benefit function of an MBT or sequence of MBTs, under the general assumption of a 1-PM or a 2-PM model. This function describes how the benefit varies depending on the subject (or patient) and enables statisticians and clinicians to interpret some results obtained with GLMMs from a PM perspective. Here, Λ_ω and $(\Lambda_\omega, \beta_{Q,\omega})$ are called the *basal trait* and *trait vector* of subject ω . We write Λ and β_Q in place of Λ_ω and $\beta_{Q,\omega}$ to refer to a generic subject from the population of subjects. We interpret (Λ, β_Q) as “a patient”. In 1-PM models, we also interpret Λ as “a patient” since β_Q is a population constant in those models. We write Y_0 and Y_Q in place of $Y_{0,\omega,i}$ and $Y_{Q,\omega,j}$ to refer to generic values of Y measured before or under Q , respectively.

Suppose the therapy's goal is that the subjects achieve the state $Y \in \mathcal{T} \subseteq \mathbb{R}$. The set $\mathcal{T}$ is the *therapeutic target*. The *basal severity function* (SF) s_0 is a function with domain $\mathbb{R}$ defined as $s_0(\Lambda) = P(Y_0 \in \mathcal{T}^C | \Lambda)$, $\Lambda \in \mathbb{R}$. For a subject with basal trait Λ , $s_0(\Lambda)$ is called *the subject's basal disease severity*. When y , y^* and $\delta \geq 0$ are fixed numbers, the targets, $[y, \infty)$, $(-\infty, y]$ and $[y^* - \delta, y^* + \delta]$ are called an incremental, a decremental, and a closed target, respectively, and y^* is a target response.

Now let Q be a particular MBT (or sequence of MBTs). If (Λ, β_Q) is the trait vector of a particular subject, the *post-treatment severity* of the subject's disease is defined as $P(Y_Q \in \mathcal{T}^C | \Lambda, \beta_Q)$. Also, under a 1-PM model, the *post-treatment severity function* s_1 is a function with domain $\mathbb{R}$ defined as $s_1(\Lambda; Q) = P(Y_Q \in \mathcal{T}^C | \Lambda, \beta_Q)$, $\Lambda \in \mathbb{R}$. Alternatively, under a 2-PM model, the *post-treatment severity function* s_2 is defined as $s_2(\Lambda, \beta_Q) = P(Y_Q \in \mathcal{T}^C | \Lambda, \beta_Q)$, $(\Lambda, \beta_Q) \in \mathbb{R}^2$ (the domain is $\mathbb{R}^2$).

For fixed Λ , $s_0(\Lambda)$ is interpreted as the severity of the disease of a subject with basal trait $\Lambda = \Lambda_\omega$ before the subject is administered any MBT. An incremental target occurs, for instance, in a treatment for chronic alcoholism with the goal that the patient achieves at least $y = 26$ abstemious days in a month. Closed targets are implemented in treatments for diabetes, at which, for instance, glucose levels between 70-100 mg/dL before meals are the treatment goal. And a decremental target is implemented in a treatment for major depression aiming at reducing the HAM-D17 score; for instance, aiming at a score $y = 7$. A closed target with $\delta = 0$ is of interest only with discrete responses. This case occurs particularly with dichotomous responses, for which reaching one of the two values of Y is the therapy goal.

For a treatment Q represented by a 1-PM model, the ideal situation is that $s_1(\Lambda; Q) < s_0(\Lambda)$ for all Λ . That is, Q should reduce all subjects' illness severity. However, it is possible that, for some treatments, $s_1(\Lambda; Q) \geq s_0(\Lambda)$ for some or all Λ (see Examples 3.1 and 3.2). Thus, our general formulation allows modeling MBTs that may be ineffective or even detrimental for some subjects.

Definition 3.1. Under a 1-PM model of treatment Q , the *benefit function (BF)* b_1 of Q is defined by $b_1(\Lambda; Q) = s_0(\Lambda) - s_1(\Lambda; Q)$, $\Lambda \in \mathbb{R}$. Alternatively, under a 2-PM model, the BF b_2 of Q is defined by $b_2(\Lambda, \beta_Q) = s_0(\Lambda) - s_2(\Lambda, \beta_Q)$, $(\Lambda, \beta_Q) \in \mathbb{R}^2$.

In practice, BFs are useful in the comparison of the benefits of two or more MBTs to a particular patient. Examples of SFs and BFs for 1-PM models with Gaussian response and identity link are provided below. Sections 4 and 5 present data analyses using a 2-PM model with Gaussian response and identity link and a 1-PM model with Bernoulli response and logit link.

Example 3.1. (*Incremental targets in 1-PM models with Gaussian response and identity link*). With a Gaussian response and identity link, the equations of a PM model for a subject ω can be written as

$$Y_{0,\omega,i} = \Lambda_\omega + \varepsilon_{0,\omega,i} \quad \text{and} \quad Y_{Q,\omega,j} = \Lambda_\omega + \beta_{Q,\omega} + \varepsilon'_{\omega,j},$$

where $\varepsilon_{0,\omega,i}$ is a random variable with mean 0 which represents the variation of Y about Λ_ω that is caused by interoccasion variability in either the subject's body or external environment or by measurement errors. Similarly, $\varepsilon'_{\omega,j}$ has mean 0 and represents variation of Y about $\Lambda_\omega + \beta_{Q,\omega}$ which may have the same causes as those of the variation of $\varepsilon_{0,\omega,i}$. Thus, for an incremental target and fixed $y \in \mathbb{R}$,

$$s_0(\Lambda) = \Phi\left(\frac{y - \Lambda}{\sigma_\varepsilon}\right) \text{ and } b_1(\Lambda; Q) = \Phi\left(\frac{y - \Lambda}{\sigma_\varepsilon}\right) - \Phi\left(\frac{y - \Lambda - \beta_Q}{\sigma'_\varepsilon}\right), \Lambda \in \mathbb{R},$$

where Φ is the standard normal cdf, and σ_ε and σ'_ε are the standard deviations of $\varepsilon_{0,\omega,i}$ and $\varepsilon'_{\omega,j}$, respectively. Next we explain the graphical interpretation of these functions from a PM perspective, using $y = 1$ and $\sigma_\varepsilon = \sigma'_\varepsilon = 1$. Figure 1 (A) shows s_0 . For incremental targets, $s_0(\Lambda)$ is a decreasing function of Λ since, by Eq. (1), untreated subjects with large Λ_ω tend to have higher values of $Y_{0,\omega,i}$ which reflect a less severe illness. Figure 1 (B) illustrates function $b_1(\cdot; Q_1)$ for a treatment Q_1 with $\beta_{Q_1} = 3$. Here, $b_1(\Lambda; Q_1) > 0$ for all Λ . However, in general, the illness severity is not reduced by Q_1 by the same amount in all subjects. For instance, the probability that $Y \in (-\infty, 1]$ is reduced by 0.5 or more only in subjects with basal trait Λ in $[-2, 1]$. In contrast, the effectiveness of Q_1 in subjects for whom $\Lambda < -5$ is very small or unimportant because $b_1(\Lambda; Q_1) \approx 0$ in these subjects. Figure 1 (C) depicts $b_1(\cdot; Q_2)$ for a treatment Q_2 with $\beta_{Q_2} = -2$. It illustrates that, with an incremental target, $b_1(\Lambda; Q) = 0$ for all Λ if and only if $\beta_Q = 0$. Therefore, the MBT is detrimental or ineffective for all subjects when $\beta_Q = 0$; that is, in such cases, the treatment undesirably increases the illness severity rather than decreasing it.

Figure 1 (D) superimposes 3 functions: the functions $s_0(\cdot)$ and $b_1(\cdot; Q_1)$ shown in (A) and (B), and $b_1(\cdot; Q_3)$ for a treatment Q_3 with $\beta_{Q_3} = 1.7$. For subjects with basal severities $s_0(\Lambda) \leq 0.6$, Q_1 and Q_3 are both very effective in that subjects' post-treatment severities reach values close to 0, which is indicated by the fact that, for these subjects, $s_0(\Lambda) \approx b_1(\Lambda; Q_1) \approx b_1(\Lambda; Q_3)$. But, for subjects with Λ satisfying $0.6 < s_0(\Lambda) < 0.87$, only Q_1 is able to produce post-treatment severities close to 0. Finally, for subjects with basal severities > 0.87 but not too close to 1, treatment Q_1 may be substantially more beneficial than Q_3 , although neither treatment will reduce completely these subjects' illness severity.

Figure 1 (D) illustrates the potential of BFs as tools that may help clinicians to compare and personalize MBTs. For instance, if treatment Q_3 is cheaper or expected to produce fewer side effects than Q_1 , then there is no doubt that Q_3 is a better choice than Q_1 for patients with severities ≤ 0.6 . Empirical Bayesian estimation of the value of Λ_ω for a particular patient will help make decisions once 1-PM model parameters are estimated with a training patient sample. This estimation of Λ_ω is made using baseline values $Y_{0,\omega,i}$ provided by the patient.

Observe that SFs and BFs allow modeling two well-known facts in clinical practice when incremental or decremental targets are used: (1) that two MBTs may have *essentially* the same benefits in a large group of patients, even if the treatments' effects (the β_Q 's) on the outcome variable Y are different; and (2) that even the best treatments may fail with very severe illnesses, that is, in patients for whom $s_0(\Lambda) \approx 1$.

BF and basal SF under a decremental target and a 2-PM model are illustrated in Section 4. With incremental or decremental targets, the value of y does not affect the graphs' shapes of BFs and basal SFs. It affects only the graphs' positions on the horizontal axis. So, if the goal

is to *compare* two MBTs in the same patient population, the actual value of y used is immaterial. Any clinical decision regarding therapy personalization is invariant with respect to the value of y , *provided patients are identified with their basal illness severities*, not with their values of Λ . To illustrate, all the above conclusions about treatments Q_1 and Q_3 obtained from Figure 1 (D) are valid regardless of the value of y chosen.

Example 3.2. (Closed targets in 1-PM models with Gaussian response and identity link).

More complex clinical phenomena occur when the therapeutic target is closed. In this case, for $\delta > 0$,

$$s_0(\Lambda) = 1 - \Phi\left(\frac{y^* + \delta - \Lambda}{\sigma_\varepsilon}\right) + \Phi\left(\frac{y^* - \delta - \Lambda}{\sigma_\varepsilon}\right), \Lambda \in \mathbb{R}, \text{ and} \quad (5)$$

$$b_1(\Lambda; Q) = \Phi\left(\frac{y^* - \delta - \Lambda}{\sigma_\varepsilon}\right) - \Phi\left(\frac{y^* + \delta - \Lambda}{\sigma_\varepsilon}\right) + \Phi\left(\frac{y^* + \delta - \Lambda - \beta_Q}{\sigma'_\varepsilon}\right) - \Phi\left(\frac{y^* - \delta - \Lambda - \beta_Q}{\sigma'_\varepsilon}\right), \Lambda \in \mathbb{R}. \quad (6)$$

Figure 2 shows examples of the functions in (5) and (6) for $\sigma_\varepsilon = \sigma'_\varepsilon = 1$ and therapeutic goals represented by $y^* = 1$ and $\delta = 1.64$. Figure 2 (A) shows the basal severity function s_0 . For a closed target, there exists a Λ at which $s_0(\Lambda)$ reaches its minimum value. This Λ is denoted as $\Lambda_m = \operatorname{argmin}_{\Lambda \in \mathbb{R}} s_0(\Lambda)$. In Figure 2 (A), $\Lambda_m = 1$ and $s_0(\Lambda_m) = 0.1$. In this case, there are no subjects with a basal severity lower than 0.1. Figure 2 (B) shows the benefit function $b_1(\cdot; Q_1)$ for a treatment Q_1 with $\beta_{Q_1} = 3$. Here, Q_1 is beneficial for subjects with $\Lambda < -0.5$, but detrimental if $\Lambda > -0.5$. That is, under treatment Q_1 , the health of subjects with $\Lambda > -0.5$, which is measured by Y , worsens. For subjects with $\Lambda = -0.5$, Q_1 does not produce any change in their illness severity. A somewhat opposite situation occurs for a treatment Q_2 for which $\beta_{Q_2} = -2$, whose BF is depicted in Figure 2 (C); in this case, Q_2 benefits only subjects with $\Lambda > 2$.

Figure 2 (D) superimposes the functions in A, B and C. It shows that Q_1 benefits only a subpopulation of subjects for whom $\Lambda < \Lambda_m$, and Q_2 benefits only a subpopulation of subjects for whom $\Lambda > \Lambda_m$, because $-0.5 < \Lambda_m = 1 < 2$. However, subjects with basal severities very close to 1 obtain meager benefits from both treatments.

Figure 2 (E) superimposes the functions $s_0(\cdot)$ in A, $b_1(\cdot; Q_1)$ in B, and an additional benefit function $b_1(\cdot; Q_3)$ with $\beta_{Q_3} = 1.7$. From this figure, the following can be said about subjects for whom $\Lambda < \Lambda_m = 1$: 1) subjects with basal severities < 0.22 will obtain undesirable effects (negative benefits) from both Q_1 and Q_3 ; 2) those with basal severities between 0.22 and 0.45 will obtain positive benefits from Q_3 but not from Q_1 ; 3) those with basal severities between 0.45 and 0.76 will obtain positive benefits from both Q_1 and Q_3 , but Q_3 is superior to Q_1 (i.e. Q_3 produces a larger reduction of the illness severity); and 4) Q_1 is more

beneficial than Q_3 when administered to subjects with severities > 0.76 , but the effects of Q_1 and Q_3 are negligible in subjects with basal severities that are very close to 1.

In general, the shapes of the functions in (5) and (6) do not depend on y^* , although they depend on δ . In particular, the lowest basal severity $s_0(\Lambda_m)$ does not depend on y^* but depends on δ . In contrast, Λ_m depends on y^* but does not depend on δ .

Importantly, with a closed target and a 1-PM model with Gaussian response and identity link, if treatments Q_1 and Q_2 have different average effects on the illness (that is, if $\beta_{Q_1} \neq \beta_{Q_2}$), there are always subjects for whom treatment Q_1 is better than Q_2 and subjects for whom Q_2 is better than Q_1 . Also, in general, for any treatment Q , if $\beta_Q > 0$, then Q is detrimental to subjects with trait $\Lambda > y^* - \beta_Q/2$ (that is, $b_1(\Lambda; Q) < 0$) but beneficial to the other subjects, although Q may not have the same detrimental or beneficial effect in two different subjects. And if $\beta_Q < 0$, then Q is detrimental for subjects with trait $\Lambda < y^* - \beta_Q/2$ but beneficial otherwise. In particular, if $\beta_Q > 0$, Q does not benefit subjects with $\Lambda > \Lambda_m$; and if $\beta_Q < 0$, Q does not benefit subjects with $\Lambda < \Lambda_m$.

Thus, many well-known clinical phenomena occurring under a closed target can be modeled using BFs and SFs. For instance, that MBTs with higher statistical effect sizes on Y are not necessarily better for all patients. In general, for two different MBTs Q and Q' , even if $\beta_Q > \beta_{Q'} > 0$, there may be a subset of patients with $\Lambda < \Lambda_m$ for whom Q' is more effective than Q . Moreover, for any treatment Q , there may be patients with a refractory illness for whom Q is neither beneficial nor detrimental, even if $|\beta_Q|$ is very large. These are the patients with very large values of $|\Lambda|$, for whom $b_1(\Lambda; Q)$ is close to 0. (See Figure 2.)

To further illustrate SFs' and BFs' versatility, note that, with a closed target and Gaussian response and identity link, and appropriate values of σ_e , σ'_e and δ , it is possible to build basal severities for which $s_0(\Lambda_m)$ is close to 0 in order to model subjects with low basal severities. Also, values of β_Q representing MBTs that reduce the illness severity to almost 0 in subjects with low basal severities can easily be exhibited. In summary, SFs and BFs provide a mathematical language that allows describing many phenomena observed in actual clinical settings.

4. Application to clinical trial of antidepressants

The STAR*D study investigated 4,041 adult subjects with non-psychotic major depressive disorder who were treated in outpatient settings [23]. The primary purpose of the study was to determine which alternative treatments work best if Citalopram (CIT) monotherapy does not produce an acceptable response. The study examined a large number of MBTs and MBT sequences and combined both standard randomization procedures and a detailed protocol with a naturalistic approach that attempted to mimic real-life clinical practice. Here, we focus only on 170 subjects who satisfied the following criteria: (1) the subjects did not have an acceptable response to CIT therapy but received either Bupropion (BUP, N=69) or Venlafaxine (VEN, N=101) as alternative treatments in the next study phase; and (2) the subjects were declared to have a successful response to these alternative treatments and were consequently assigned to follow-up, in which the subjects continued under the alternative

treatment. Thus, each of these subjects was considered to have benefited from BUP or VEN. We want to compare the amounts of benefit BUP and VEN provided.

Here, Y is the HAM-D17 total depression score. It produced values from 0 to 42 inclusive, and all integer numbers in this range were observed in the patient population. Thus, we treated the HAM-D17 score as a continuous variable, as most statisticians would do in practice. Each subject provided at most two baseline measures: $Y_{0,\omega,1}$, which was obtained right before CIT therapy started, and $Y_{0,\omega,2}$, obtained at the end of CIT therapy. Thus, $k_{0,\omega} = 0, 1$ or 2 . Since CIT therapy was not considered successful in these subjects, we assumed $Y_{0,\omega,1}$ and $Y_{0,\omega,2}$ were repeated measures of the same dimension Y_0 . Similarly, each subject provided at most two measures of Y under the alternative treatment: $Y_{Q,\omega,1}$, which was obtained right before the subject passed to follow-up, and $Y_{Q,\omega,2}$, obtained after about three months of follow-up, where $Q = \text{BUP or VEN}$. Thus, $k_{1,\omega} = 0, 1$ or 2 . In total, the 170 subjects provided 598 HAM-D17 scores for this analysis. Each subject provided one to four scores, and the average number of scores provided was 3.5. An assumption of missingness at random was made to include subjects who did not provide all four values of Y .

2-PM models for BUP and VEN were combined into a GLMM model, assuming a Gaussian response, identity link and independent homoscedastic errors (Table 1). Stata's `meglm` command was used to fit the model (StataCorp LP, College Station, TX, USA) (Appendix 1). Residual analyses showed both that the model fitted well and that the assumption of normality for the untransformed scores was reasonable. The model included a random intercept, and random effects for BUP and VEN treatments. An unstructured covariance matrix for these 3 random coefficients was estimated, except that the covariance between BUP and VEN was set to 0 because the two treatments were administered to independent samples. The database had one row per each available HAM-D17 score, and two dummy variables representing treatments were computed, namely X_{BUP} and X_{VEN} . The variable X_{BUP} was coded as 1 if the score was obtained under BUP, or 0 if the score was obtained at baseline or under VEN. Analogously, X_{VEN} was coded as 1 if the score was obtained under VEN, or 0 if the score was obtained at baseline or under BUP. In Table 1, baseline is the “reference point” with respect to which the effects of BUP and VEN are assessed. The effects of age, gender, race, smoking, alcohol and use of illegal drugs were examined as possible patient covariates, but the final model included only black race (Table 1). The black race variable was coded as 1 if the subject was self-declared black or African American, 0 otherwise.

There was a significant interaction between BUP treatment and black race (Table 1). Using the estimates in Table 1, we computed the benefit functions in Figure 3. As in the STAR*D protocol, we used HAM-D17 ≥ 7 as an indicator of remission, which is a decremental target with $y = 7$. With a decremental target, we have $\mathcal{T}^C = (y, \infty)$. Then, the formulas for basal SF and for BF under decremental targets in 2-PM models with Gaussian response and identity link are respectively

$$s_0(\Lambda) = 1 - \Phi\left(\frac{y - \Lambda}{\sigma_\varepsilon}\right), \Lambda \in \mathbb{R}, \text{ and } b_2(\Lambda, \beta_Q) = \Phi\left(\frac{y - \Lambda - \beta_Q}{\sigma_\varepsilon}\right) - \Phi\left(\frac{y - \Lambda}{\sigma_\varepsilon}\right), (\Lambda, \beta_Q) \in \mathbb{R}. \quad (7)$$

Figure 3 shows s_0 , and profiles of b_2 for different values of β_Q , for blacks (A) and non-blacks (B). In Figure 3 (A), $b_2(\Lambda, \beta_Q)$ is plotted as a function of only Λ for four different values of β_Q , namely

$$\beta_{Q,1} = \hat{E}[\beta_{\text{BUP}}], \quad \beta_{Q,2} = \hat{E}[\beta_{\text{VEN}}], \\ \beta_{Q,3} = \hat{E}[\beta_{\text{BUP}}] - 1 \times (\hat{V}(\beta_{\text{BUP}}))^{1/2} \text{ and } \beta_{Q,4} = \hat{E}[\beta_{\text{VEN}}] + 1 \times (\hat{V}(\beta_{\text{BUP}}))^{1/2},$$

where, $\hat{E}[\beta_{\text{BUP}}] = -10.0 + 5.24$, $\hat{E}[\beta_{\text{VEN}}] = -9.85$, $\hat{V}(\beta_{\text{BUP}}) = 10.6$, and $\hat{V}(\beta_{\text{VEN}}) = 8.49$. Each of these four curves represents a subpopulation of patients. For instance, the curve corresponding to $\beta_{Q,1}$ represents black subjects for whom BUP had an average effect; the curve for $\beta_{Q,3}$ represents black subjects for whom BUP had an effect one standard deviation lower than BUP average effect; and so on.

Similarly, in Figure 3 (B), in addition to the curves for $\beta_{Q,i}$, $i = 1, \dots, 4$, we also plotted curves with

$$\beta_{Q,5} = \hat{E}[\beta_{\text{BUP}}] + 1.5 \times (\hat{V}(\beta_{\text{BUP}}))^{1/2} \text{ and } \beta_{Q,6} = \hat{E}[\beta_{\text{VEN}}] - 1.5 \times (\hat{V}(\beta_{\text{VEN}}))^{1/2}.$$

However, for non-black subjects, $\hat{E}[\beta_{\text{BUP}}] = -10.0$.

Figure 3 also shows vertical grid lines representing subjects with different values of the basal trait Λ . For instance, the number -2 in the top horizontal axis represents the value of Λ given by $\hat{E}[\Lambda] - 2 \times (\hat{V}(\Lambda))^{1/2}$, where $\hat{V}(\Lambda) = 20.9$, and $\hat{E}[\Lambda] = 17.0 + 3.54$ for blacks and $\hat{E}[\Lambda] = 17.0$ for non-blacks (Table 1). In this case, the vertical line -2 represents subjects for whom Λ is two standard deviations lower than the average Λ of their race group. Similarly, the vertical line with 0 at the top corresponds to the average Λ of the corresponding race group ($17.0 + 3.54$ for blacks, 17.0 for non-blacks).

Here, the basal severity is interpreted as the probability that a particular subject with basal trait Λ has a HAM-D17 score ≥ 7 at a particular moment before taking BUP or VEN, and the benefit of a treatment for that subject is the reduction of this severity in probability units.

In black subjects, the average effect of BUP (estimated as $-10.0 + 5.24$) was statistically different from the average effect of VEN (-9.85) (Wald $\chi^2 = 9.7$, $\text{df} = 1$, $p = 0.002$). Comparing the curves for $\beta_{Q,1}$ and $\beta_{Q,2}$ in blacks (Figure 3A), we observe that, in general, after controlling for basal illness severity, average black subjects benefitted substantially

more from VEN than BUP; that is, fixing the value of Λ , the curve for $\beta_{Q,2}$ is substantially higher than that for $\beta_{Q,1}$, especially in subjects whose Λ was between -2 and -1 standard deviations from the average Λ . However, when comparing the curves for $\beta_{Q,3}$ and $\beta_{Q,4}$, we observe that there were subpopulations of black subjects who obtained similar or more benefit from BUP than from VEN (the curve for $\beta_{Q,3}$ is higher than that for $\beta_{Q,4}$).

In non-black subjects, the average effect of BUP was not statistically different from the average effect of VEN (-10 versus -9.85) ($p = 0.82$). In Figure 3 (B), we observe that, after controlling for basal severity, average non-black subjects obtained the same benefit from BUP and VEN, which can be inferred from the fact that the curves for $\beta_{Q,1}$ and $\beta_{Q,2}$ are hardly distinguishable. However, we can exhibit subpopulations of non-black subjects for whom BUP provided substantially more benefit than VEN (compare the curves for $\beta_{Q,3}$ and $\beta_{Q,4}$), as well as subpopulations for whom VEN provided substantially more benefit than BUP (compare the curves for $\beta_{Q,5}$ and $\beta_{Q,6}$). These subpopulations are predicted to exist because $V(\beta_{\text{BUP}})$ and $V(\beta_{\text{VEN}})$ are estimated to be larger than 0 (for BUP, the estimate is 10.6 [3.54, 31.8]; for VEN, 8.49 [2.77, 26.0] ; Table 1).

Interestingly, both black and non-black subjects with Λ higher than the average Λ of their race had a very severe illness before trying BUP or VEN. This can be inferred from the fact that the basal severities in both Figures 3 (A) and (B) are almost flat and close to 1 at the right of the vertical lines with 0 at the top. However, the curves for $\beta_{Q,2}$ show that the maximum benefit of VEN was 0.2 for these black subjects (Figure 3A), but 0.44 for the non-black subjects (Figure 3B). In general, VEN benefitted more non-black than black severely ill subjects with comparable values of β_{VEN} . This illustrates the important point that differences in the basal trait Λ may contribute to differences in the benefit of a treatment between two populations, not only to the basal severity.

5. Application to clinical trial of respiratory treatment

Next, we analyze data from a clinical trial comparing the effects of a drug (called Drug A) versus Placebo on a chronic respiratory condition [24]. Here, only data from Center 1 is analyzed. Fifty-six subjects were randomized to either Drug A ($N=27$) or Placebo ($N=29$). The dichotomous response Y was coded as 1 if the subject had a good respiratory status (status 3 or 4), or 0 otherwise. Only one baseline (pre-treatment) measure was available from each subject: $Y_{0,\omega,1}$. (Of the 56 subjects, only seven had a good baseline respiratory status.) Four repeated measures were obtained during treatment: $Y_{Q,\omega,j}$, $j = 1, \dots, 4$, where $Q = \text{Drug A or Placebo}$.

1-PM models for Drug A and Placebo were combined into a random intercept logistic regression model of Y , using Stata's `meglm` command (Table 2; Appendix 2). Each subject occupied 5 rows of the database, one per each value of Y . Two dummy variables were computed and used as independent variables: X_A , which was coded as 1 if Y was obtained under Drug A, or 0 if Y was obtained at baseline or under Placebo; and X_B , coded as 1 if Y was obtained under Placebo, or 0 if Y was obtained at baseline or under Drug A.

Since the therapeutic target is to achieve a good respiratory status (i.e. $\mathcal{T} = \{1\}$) and we are modeling a Bernoulli response with a logit link, the basal SF and BF are respectively

$$s_0(\Lambda) = 1 - (1 + \exp(-\Lambda))^{-1} \text{ and } b_1(\Lambda; Q) = (1 + \exp(-\Lambda - \beta_Q))^{-1} - (1 + \exp(-\Lambda))^{-1}, \Lambda \in \mathbb{R},$$

where $\beta_Q = 1.81$ if $Q = \text{Drug A}$, or $\beta_Q = 0.49$ if $Q = \text{Placebo}$ (Table 2).

Figure 4 depicts $s_0(\cdot)$, and $b_1(\cdot; Q)$ for both $Q = \text{Drug A}$ and $Q = \text{Placebo}$. A particular value on the Λ axis can be viewed as a composite of a particular subject's age and unobserved variables that also shaped the subject's individuality. That is, one value on the Λ axis represents a particular individual. Here, the basal severity for a particular subject is interpreted as the probability that he/she had a poor respiratory status at a particular time point before treatment started.

The vertical lines on Figure 4 mark average values of Λ for particular ages observed in the patient sample (the youngest and oldest observed ages were 11 and 63 years old). For instance, the predicted average value of Λ for subjects of age 28 was $0.65 - 0.073 \times 28 = -1.39$. Thus, for an average subject of age 28, the basal severity was 0.8, and the benefit this subject obtained from Drug A was 0.4. Thus, Drug A reduced by half the probability that this subject exhibited a poor respiratory condition.

Observe that subjects of age 47 or older tended to have a very severe illness, with basal severities > 0.9 . For these subjects, Drug A provided a very modest benefit, reducing their illness severity just by 0.2 or less probability units. In contrast, children of age 15 or younger tended to have basal severities < 0.6 , and Drug A reduced their illness severity by about 0.4 probability units. Interestingly, Drug A provided a similar benefit for subjects of age between 11 and the observed median age (28). In these particular subjects, Drug A reduced the illness severity by about 0.4. In general, for subjects of age > 28 , older age was not only associated with increased disease severity, but also with a reduced benefit from Drug A. Note the above conclusions cannot be easily inferred by just browsing the parameter estimates in Table 2. This reinforces our claim that benefit functions help understand the clinical phenomenon and provide valuable additional information. In fact, the above complex interaction between age and Drug A could not be detected by including an interaction term in the model.

6. Estimation of the benefit to an individual patient

To illustrate how our approach translates into clinical practice, we examine how well a non-black patient with depression responded to bupropion using empirical Bayesian estimation of treatment benefit. The patient can be a STAR*D subject or a new patient. The empirical Bayes predictor and the credibility interval for the benefit to a patient ω depend on the

averages $\bar{Y}_{0,\omega} = \left(\sum_{i=1}^{k_{0,\omega}} Y_{0,\omega,i} \right) / k_{0,\omega}$ and $\bar{Y}_{Q,\omega} = \left(\sum_{j=1}^{k_{1,\omega}} Y_{Q,\omega,j} \right) / k_{1,\omega}$, the amount of data collected from the patient ($k_{0,\omega}$ and $k_{1,\omega}$), covariate values, and the estimates of population parameters (Appendix 3). For instance, Table 3 shows that if bupropion lowers the patient's

HAM-D17 average score from 20 to 10, and $k_{0,\omega} = k_{1,\omega} = 2$, then the estimated bupropion benefit is 31.5% [95% credibility interval, (7.0, 68.8)].

In Table 3, estimated benefits increase from right to left. Indeed, larger score reductions from baseline reflect higher bupropion benefits, and the benefit scale is sensitive to score deterioration—the larger the deterioration, the lower our degree of belief that the treatment is beneficial. Also, other things being equal, the greater the $k_{1,\omega}$ the shorter the interval length and, therefore, the higher the precision of the benefit estimate. An analogous trend is observed for $k_{0,\omega}$.

7. Discussion

In this paper, we propose definitions for the severity functions of chronic diseases and the benefit functions of MBTs. We also describe a graphical method for examining the severity of a chronic disease and the individual benefit of an MBT. BF's can be a good addition to the statistician's tool box. In fact, rather than just reporting that the effects of two or more MBTs on an outcome variable are significantly different, the statistician may provide a better picture of the clinical reality by reporting the treatments' BF's along the basal SF and by describing how the benefits of the treatments differ in specific patient subpopulations. In particular, reporting that the effects of two MBTs are significantly different is not enough since an average difference does not necessarily imply that the two MBTs will provide substantially dissimilar epidemiological benefits to a specific patient subpopulation although they can provide differing benefits across the patients of other subpopulations.

Our methodology allows demonstrating graphically the existence of subjects with differential responses to a treatment, comparing the benefits of a variety of treatments, and quantifying the effectiveness of treatments in specific patients. It utilizes repeated measures to separate from noise the components of variability that are relevant to PM; i.e., between-subject variability and subject-by formulation interaction variability [2, 3]. Since conclusions based on our methodology are model-dependent, however, we must carefully evaluate the goodness-of-fit of the model before drawing conclusions.

In 1-PM models, the effect β_Q of an MBT on Y is the same for all patients; whereas, in 2-PM models, it differs across patients. This may wrongly suggest that only 2-PM models are of interest in PM, whose main tenet is that some MBTs may act differentially across patients. As Figure 1 suggests, however, the fact that an MBT has the same effect in all patients does not imply that the MBT benefits all patients in the same fashion. (Actually, most statisticians seem to believe incorrectly that same effect size implies same MBT benefits.) Furthermore, since the basal trait Λ is not only shaped by covariates but also by other unobserved factors represented by α , the presence of the same effect in all patients does not imply similar benefits for two patients with the same covariate values. Thus, 1-PM models and their associated BF's also allow modeling important clinical phenomena.

Empirical Bayes predictors of Λ and β_Q can be used to estimate both the basal severity of a specific patient and the benefit that the patient obtained from a treatment. To estimate parameters of 1-PM and 2-PM models, we can use standard software for fitting GLMMs (or

RELMs in the case of Gaussian response with identity link). For example, we can use SAS GLIMMIX or MIXED procedures (SAS Institute Inc, Cary, NC, USA) or Stata's meglm or mixed commands (StataCorp LP, College Station, TX, USA). These computer packages can easily provide empirical Bayes estimators.

The potential for using REs models as tools for separating genetic and environmental causes of human and animal responses was openly recognized by some of the first researchers who explored these models' mathematics and applications [25-27]. In fact, empirical and theoretical studies suggest a working hypothesis that can be named The Mixed-Model Hypothesis: The random coefficients of a mixed-regression model of a pharmacokinetic or pharmacodynamic response incorporate all the variability of the response that patients' uncontrolled genetic heterogeneity causes as well as some inter-individual variability that other factors generate. In contrast, the variability of the error term of the model does not reflect genetic variability across patients [16]. In other words, the total variability of random coefficients is an "upper bound" of the genetic variability that was not controlled for by model covariates [2, 4, 28]. Thus, statisticians and clinicians involved in PM research must not overlook the potential of the concept of REs to develop statistical tools for PM.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The STAR*D study was supported by NIMH Contract #N01MH90003, and has the ClinicalTrials.gov Identifier, NCT00021528. The Author was supported by an Institutional Clinical and Translational Science Award, NIH/NCATS Grant Number UL1TR000001 (awarded to the University of Kansas Medical Center). The author did not receive any salary or payment from the STAR*D grant or its investigators. This manuscript may not reflect the opinions or views of the STAR*D study investigators or the NIH. The author thanks two reviewers whose suggestions led to substantial improvements to the paper. The author also thanks Mr. Douglas Eikermann for editorial assistance.

Appendix 1: Stata code for STAR*D data

First, we created a matrix that set the covariance between the random effects of BUP and VEN to 0:

```
matrix input ReCovMat=( . . . . \ 0 . . . . \ . . . . )
```

```
matrix rownames ReCovMat=BUP VEN Intercept
```

```
matrix colnames ReCovMat= BUP VEN Intercept
```

The random-effects linear model was fitted with the following code:


```
meglm HRSD17_Total black01 BUP_01 VEN_01 interBlackBUP ///
```

```
// id: BUP_01 VEN_01, covariance(fixed(ReCovMat))
```

where HRSD17_Total = Y , black01 is the race variable, BUP_01 = X_{BUP} , VEN_01 = X_{VEN} , interBlackBUP is the product of black01 and BUP_01, and id is a variable that identified the subjects.

Appendix 2: Stata code for respiratory data

The following code fitted the random intercept logistic regression model:

```
meglm GoodStat01 Age Placebo_01 Drug_01 ||Patient: , ///
```

```
family(bernoulli) link(logit)
```

where GoodStat01 = Y , Placebo_01 = X_P , Drug_01 = X_A , and Patient is a variable that identified the subjects.

Appendix 3: Formulas for the empirical Bayes predictor and credibility interval for an individual benefit

Consider the 2-PM model in Section 4. Using formulas in Frees [12] (pages 149, 420), it can be shown that, for a particular patient ω , the posterior mean vector and variance-covariance matrix of $(\Lambda', \beta_Q^*)^T$ given the patient's HAM-D17 scores are respectively

$$\mu_Y = \sigma_\varepsilon^{-2} D (I_2 - \sigma_\varepsilon^{-2} Z^T Z (D^{-1} + \sigma_\varepsilon^{-2} Z^T Z)^{-1}) Y^* + (E[\Lambda'], E[\beta_Q^*])^T \text{ and } \Sigma = D - D Z^T V^{-1} Z D,$$

where $D = \text{Var}((\Lambda', \beta_Q^*)^T)$, $V = \sigma_\varepsilon^2 I_k + Z D Z^T$, $k = k_{0,\omega} + k_{1,\omega}$ and I_2 and I_k are 2×2 and $k \times k$ identity matrices. $Z = (Z_{ij})$ is a $k \times 2$ matrix such that $Z_{i1} = 1$ for all $1 \leq i \leq k$; $Z_{i2} = 0$ for $1 \leq i \leq k_{0,\omega}$; and $Z_{i2} = 1$ for $k_{0,\omega} + 1 \leq i \leq k$. Finally, $Y^* = (b_1, b_2)^T$ where

$$b_1 = k_{0,\omega} (\bar{Y}_{0,\omega} - E[\Lambda']) + k_{1,\omega} (\bar{Y}_{Q,\omega} - E[\Lambda'] - E[\beta_Q^*]) \text{ and } b_2 = k_{1,\omega} (\bar{Y}_{Q,\omega} - E[\Lambda'] - E[\beta_Q^*]).$$

We substituted parameters in μ_Y and Σ with estimates obtained with code in Appendix 1. Using the estimates of μ_Y and Σ , we simulated 20,000 pairs of the form (Λ, β_Q) from the corresponding bivariate normal distribution. These pairs were replaced into b_2 in formula (7). The median of the 20,000 Monte Carlo benefits was the empirical Bayes predictor of the

patient's benefit, and the 2.5% and 97.5% percentiles constituted the 95% credibility interval (Table 3). The author wrote a Stata ado program called preben, which does these computations (online material).

References

1. Food and Drug Administration. Paving the way for personalized medicine: FDA's role in a new era of medical product development. U.S. Department of Health and Human Services. Food and Drug Administration; Silver Spring, MD: 2013.
2. Senn S. Individual therapy: new dawn or false dawn. *Drug Information Journal*. 2001; 35:1479–1494.
3. Senn S. Mastering variation: variance components and personalised medicine. *Statistics in Medicine*. 2015; in press. doi: 10.1002/sim.6739
4. Diaz FJ, Berg MJ, Krebill R, Welty T, Gidal BE, Alloway R, Privitera M. Random-effects linear modeling and sample size tables for two special cross-over designs of average bioequivalence studies: the 4-period, 2-sequence, 2-formulation and 6-period, 3-sequence, 3-formulation designs. *Clinical Pharmacokinetics*. 2013; 52:1033–1043. [PubMed: 24085600]
5. Schall R, Endrenyi L, Ring A. Residuals and outliers in replicate design crossover studies. *Journal of Biopharmaceutical Statistics*. 2010; 20:835–849. [PubMed: 20496209]
6. L'Abbé KA, Detsky AS, O'Rourke K. Meta-analysis in clinical research. *Annals of Internal Medicine*. 1987; 107:224–233. [PubMed: 3300460]
7. Lubsen J, Tijssen JG. Large trials with simple protocols: indications and contraindications. *Controlled clinical trials*. 1989; 10:151–160.
8. Boissel JP, Kahoul R, Marin D, Boissel FH. Effect model law: An approach for the implementation of personalized medicine. *Journal of Personalized Medicine*. 2013; 3:177–190. [PubMed: 25562651]
9. Sharp SJ, Thompson SG, Altman DG. The relation between treatment benefit and underlying risk in meta-analysis. *British Medical Journal*. 1996; 313:735–738. [PubMed: 8819447]
10. Li W, Girard P, Boissel JP, Gueyffier F. The calculation of a confidence interval on the absolute estimated benefit for an individual patient. *Computers and Biomedical Research*. 1998; 31:244–256. [PubMed: 9731267]
11. Rabe-Hesketh, S.; Skrondal, A. Generalized linear mixed-effects models. In: Fitzmaurice, G.; Davidian, M.; Verbeke, G., et al., editors. *Longitudinal Data Analysis*. London: Chapman & Hall/CRC; 2009.
12. Frees, EW. *Longitudinal and panel data*. Cambridge: Cambridge University Press; 2004.
13. Sheiner LB, Rosenberg B, Melmon KL. Modelling of individual pharmacokinetics for computer-aided drug dosage. *Computers and Biomedical Research*. 1972; 5:411–459. [PubMed: 4634367]
14. Whiting B, Kelman AW, Grevel J. Population pharmacokinetics. Theory and clinical application. *Clinical Pharmacokinetics*. 1986; 11:387–401. [PubMed: 3536257]
15. Pillai G, Mentré F, Steimer JL. Non-linear mixed effects modeling – from methodology and software development to driving implementation in drug development science. *Journal of Pharmacokinetics and Pharmacodynamics*. 2005; 32:161–183. [PubMed: 16283536]
16. Diaz FJ, Yeh HW, de Leon J. Role of statistical random-effects linear models in personalized medicine. *Current Pharmacogenomics and Personalized Medicine*. 2012; 10:22–32. [PubMed: 23467392]
17. Diaz FJ, Cogollo M, Spina E, Santoro V, Rendon DM, de Leon J. Drug dosage individualization based on a random-effects linear model. *Journal of Biopharmaceutical Statistics*. 2012; 22:463–484. [PubMed: 22416835]
18. Diaz FJ, de Leon J. The mathematics of drug dose individualization should be built with random effects linear models. *Therapeutic Drug Monitoring*. 2013; 35:276–277. [PubMed: 23503456]
19. Diaz FJ, Eap CB, Ansermot N, Crettol S, Spina E, de Leon J. Can valproic acid be an inducer of clozapine metabolism? *Pharmacopsychiatry*. 2014; 47:89–96. [PubMed: 24764199]

20. Botts S, Diaz FJ, Santoro V, Spina E, Muscatello MR, Cogollo M, Castro FE, de Leon J. Estimating the effects of co-medications on plasma olanzapine concentrations by using a mixed model. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*. 2008; 32:1453–1458. [PubMed: 18555573]
21. Diaz FJ, Rivera TE, Josiassen RC, de Leon J. Individualizing drug dosage by using a random intercept linear model. *Statistics in Medicine*. 2007; 26:2052–2073. [PubMed: 16847902]
22. Wang P, Tsai GF, Qu A. Conditional inference functions for mixed-effects models with unspecified random-effects distribution. *Journal of the American Statistical Association*. 2012; 107:725–736.
23. Fava M, Rush AJ, Trivedi MH, Nierenberg AA, et al. Background and rationale for the sequenced treatment alternatives to relieve depression (STAR*D) study. *Psychiatric clinics of North America*. 2003; 26:457–494. [PubMed: 12778843]
24. Stokes, ME.; Davis, CS.; Koch, GG. *Categorical Data Analysis Using the SAS System*. 2nd. Cary, NC: SAS Institute Inc.; 2000.
25. Henderson CR. Estimation of variance and covariance components. *Biometrics*. 1953; 9:226–252.
26. Kalow W, Endrenyi L, Tang BK. Repeat administration of drugs as a means to assess the genetic component in pharmacological variability. *Pharmacology*. 1999; 58:281–284. [PubMed: 10325572]
27. Neale, MC.; Cardon, LR. *Methodology for genetic studies of twins and families*. London: Kluwer; 1992.
28. Ozdemir V, Kalow W, Tothfalusi L, Bertilsson L, Endrenyi L, Graham JE. Multigenic control of drug response and regulatory decision-making in Pharmacogenomics: The need for an upper-bound estimate of genetic contributions. *Current Pharmacogenomics and Personalized Medicine*. 2005; 3:53–71.

Figure 1.

Illustration of a basal severity function and benefit functions for a 1-PM model with a Gaussian response with identity link, using an incremental therapeutic target with $y = 1$ and $\sigma_\varepsilon = \sigma'_\varepsilon = 1$. (A) Basal severity function. (B) Benefit function for $\beta_{Q_1} = 3$. (C) Benefit function for $\beta_{Q_2} = -2$. (D) Superimposition of the functions in A and B and another benefit function for which $\beta_{Q_3} = 1.7$.

Figure 2. Illustration of a basal severity function and benefit functions for a 1-PM model with a Gaussian response with identity link, using a closed target with $y^* = 1$, $\delta = 1.64$ and

$\sigma_\varepsilon = \sigma'_\varepsilon = 1$. (A) Basal severity function. (B) Benefit function for $\beta_{Q_1} = 3$. (C) Benefit function for $\beta_{Q_2} = -2$. (D) Superimposition of the functions in A, B and C. (E) Superimposition of the functions in A and B and an additional benefit function with $\beta_{Q_3} = 1.7$.

Figure 3. Comparison of benefit functions of Bupropion and Venlafaxine in patients refractory to Citalopram treatment. (A) Black patients. (B) Non-black patients.

Figure 4.

Benefit functions for Drug A and Placebo in a clinical trial aiming at improving respiratory status. The vertical lines mark average values of Λ for subjects of particular ages. The illustrated ages are the youngest age in the patient sample (11), the oldest (63), the median age (28), and the 10% and 90% percentiles (15 and 47).

Table 1

Random effects linear model of HAM-D17 scores from 170 subjects refractory to Citalopram monotherapy who received either Bupropion or Venlafaxine as alternative treatments. The effects of Bupropion and Venlafaxine are measured with respect to observations obtained before administering these treatments.^a

Parameter name	Parameter estimate	P-value	95% CI
<i>Fixed effects</i>			
Black race ^b	3.54	0.001	[1.48, 5.59]
BUP ^c	-10.0	<0.001	[-11.3, -8.71]
VEN ^d	-9.85	<0.001	[-10.9, -8.82]
Interaction Black race*BUP	5.24	0.002	[1.89, 8.58]
Intercept	17.0	<0.001	[16.1, 17.9]
<i>Variances of random effects</i>			
BUP	10.6	--	[3.54, 31.8]
VEN	8.49	--	[2.77, 26.0]
Intercept	20.9	--	[14.9, 29.4]
<i>Covariances between random effects^e</i>			
Cov(BUP, Intercept)	-8.23	0.035	[-15.9, -0.59]
Cov(VEN, Intercept)	-8.06	0.022	[-15.0, -1.15]
Residual variance	19.9	--	[16.7, 23.6]

CI: Confidence interval; BUP: Bupropion; VEN: Venlafaxine.

^aBefore fitting the models, a few inconsistencies in the STAR*D data were found and corrected by the author. These inconsistencies and corrections were documented in a text file that is available from the author on request.

^bThe dichotomous variable was defined as 1 if the subject was self-declared black or African American, 0 otherwise.

^cThe dichotomous variable was defined as 1 if the HAM-D17 score was obtained under BUP treatment, or 0 if it was obtained at baseline or under VEN treatment.

^dThe dichotomous variable was defined as 1 if the HAM-D17 score was obtained under VEN treatment, or 0 if it was obtained at baseline or under BUP treatment.

^eSince subjects on BUP and VEN constituted two independent samples, the covariance between the random effects of BUP and VEN was fixed to 0 during parameter estimation.

Table 2

Random intercept logistic regression model of good respiratory status (N=56). The effects of Drug A and Placebo are measured with respect to baseline respiratory status.^a

Parameter name	Parameter estimate	P-value	95% CI
<i>Fixed effects</i>			
Age	−0.073	0.05	[−0.15, 0.0004]
Drug A ^b	1.81	0.002	[0.65, 2.98]
Placebo ^c	0.49	0.39	[−0.63, 1.62]
Intercept	0.65	0.60	[−1.76, 3.05]
<i>Variance of intercept</i>	7.93		[3.55, 17.7]

CI: Confidence interval.

^aThe dichotomous respiratory status was defined as 1 if the subject had a good respiratory status (status 3 or 4), 0 otherwise.

^bThe dichotomous variable was defined as 1 if the respiratory status was obtained under Drug A, or 0 if it was obtained at baseline or under Placebo.

^cThe dichotomous variable was defined as 1 if the respiratory status was obtained under Placebo, or 0 if it was obtained at baseline or under Drug A.

Table 3

Empirical Bayes predictors and 95% credibility intervals for the benefit of bupropion to specific non-black patients. Estimation of the benefit for a particular patient requires knowing the patient's average baseline HAM-D17 score ($\bar{Y}_{0,\omega}$), the average response score ($\bar{Y}_{Q,\omega}$), and the numbers of repeated measures ($k_{0,\omega}$ and $k_{1,\omega}$). Benefits were multiplied by 100.

(A) $k_{0,\omega} = 1$					
		$\bar{Y}_{Q,\omega}=3$	$\bar{Y}_{Q,\omega}=10$	$\bar{Y}_{Q,\omega}=20$	$\bar{Y}_{Q,\omega}=30$
$\bar{Y}_{0,\omega}=30$	$k_{1,\omega}=1$	39.6 (7.9, 81.2)	20.4 (2.2, 63.9)	5.2 (0.3, 32.0)	0.8 (0.0, 10.4)
	$k_{1,\omega}=2$	51.8 (16.8, 84.7)	21.8 (3.7, 58.5)	2.4 (0.1, 16.7)	0.1 (0.0, 1.6)
	$k_{1,\omega}=3$	58.4 (25.0, 85.6)	22.4 (4.9, 55.0)	1.5 (0.1, 10.4)	0.0 (0.0, 0.4)
$\bar{Y}_{0,\omega}=20$	$k_{1,\omega}=1$	55.4 (15.6, 86.8)	34.3 (5.8, 75.3)	11.4 (0.9, 48.5)	2.3 (0.1, 19.9)
	$k_{1,\omega}=2$	61.8 (25.5, 87.6)	31.5 (7.0, 68.8)	5.0 (0.3, 25.4)	0.2 (-0.0, 3.4)
	$k_{1,\omega}=3$	65.4 (32.4, 87.6)	29.9 (7.7, 63.1)	2.7 (0.2, 15.1)	0.0 (-0.0, 0.7)
$\bar{Y}_{0,\omega}=10$	$k_{1,\omega}=1$	59.7 (20.5, 86.1)	44.5 (9.0, 78.7)	19.7 (0.9, 59.4)	4.9 (-1.3, 31.7)
	$k_{1,\omega}=2$	62.0 (24.9, 86.1)	38.7 (8.6, 72.1)	7.8 (-0.8, 34.3)	0.4 (-1.6, 6.0)
	$k_{1,\omega}=3$	63.0 (28.2, 85.7)	34.7 (8.4, 66.6)	4.0 (-1.7, 20.6)	0.0 (-1.1, 1.3)
(B) $k_{0,\omega} = 2$					
		$\bar{Y}_{Q,\omega}=3$	$\bar{Y}_{Q,\omega}=10$	$\bar{Y}_{Q,\omega}=20$	$\bar{Y}_{Q,\omega}=30$
$\bar{Y}_{0,\omega}=30$	$k_{1,\omega}=1$	30.8 (5.3, 73.3)	15.3 (1.5, 54.8)	3.9 (0.2, 25.5)	0.6 (0.0, 8.3)
	$k_{1,\omega}=2$	44.1 (12.9, 79.7)	17.6 (2.8, 51.9)	2.0 (0.1, 14.1)	0.1 (0.0, 1.4)
	$k_{1,\omega}=3$	52.0 (20.5, 81.9)	19.0 (4.0, 49.6)	1.3 (0.1, 9.1)	0.0 (0.0, 0.4)
$\bar{Y}_{0,\omega}=20$	$k_{1,\omega}=1$	53.6 (15.2, 86.6)	33.5 (6.0, 74.6)	12.1 (1.1, 48.6)	2.8 (0.1, 21.2)
	$k_{1,\omega}=2$	60.8 (24.7, 87.5)	31.1 (7.2, 68.2)	5.5 (0.4, 26.1)	0.3 (-0.0, 4.1)
	$k_{1,\omega}=3$	64.8 (31.7, 87.7)	29.8 (7.9, 62.6)	3.1 (0.2, 15.9)	0.1 (-0.1, 0.9)
$\bar{Y}_{0,\omega}=10$	$k_{1,\omega}=1$	59.8 (21.2, 84.9)	46.3 (9.0, 78.7)	22.9 (-1.0, 61.6)	6.5 (-5.5, 37.0)
	$k_{1,\omega}=2$	61.3 (24.9, 84.9)	39.8 (7.6, 72.1)	8.9 (-5.0, 37.5)	0.2 (-6.4, 7.8)
	$k_{1,\omega}=3$	61.9 (28.0, 84.5)	35.3 (6.8, 66.5)	4.1 (-6.7, 22.6)	-0.1 (-5.2, 1.5)