OPEN

Determination of Phenotypic Resistance Cutoffs From Routine Clinical Data

Alejandro Pironti, MSc,* Hauke Walter, Dr med,†‡ Nico Pfeifer, Dr rer nat,* Elena Knops, Dr rer nat,§ Nadine Lübke, Dr rer nat,‖ Joachim Büch, Dr phil,* Simona Di Giambenedetto, MD,¶ Rolf Kaiser, Dr rer nat,§ and Thomas Lengauer, Dr phil, PhD,* The EuResist Network Study Group

Background: HIV-1 drug resistance can be measured with phenotypic drug-resistance tests. However, the output of these tests, the resistance factor (RF), requires interpretation with respect to the in vivo activity of the tested variant. Specifically, the dynamic range of the RF for each drug has to be divided into a suitable number of clinically meaningful intervals.

Methods: We calculated a susceptible-to-intermediate and an intermediate-to-resistant cutoff per drug for RFs predicted by geno2pheno[resistance]. Probability densities for therapeutic success and failure were estimated from 10,444 treatment episodes. The density estimation procedure corrects for the activity of the backbone drug compounds and for therapy failure without drug resistance. For estimating the probability of therapeutic success given an RF, we fit a sigmoid function. The cutoffs are given by the roots of the third derivative of the sigmoid function.

Results: For performance assessment, we used geno2pheno[resistance] RF predictions and the cutoffs for predicting therapeutic success in 2 independent sets of therapy episodes. HIVdb was used for performance comparison. On one test set (n = 807), our cutoffs and HIVdb performed equally well receiver operating characteristic curve [(ROC)–area under the curve (AUC): 0.68]. On the other test set (n = 917), our cutoffs (ROC–AUC: 0.63) and HIVdb (ROC–AUC: 0.65) performed comparatively well.

Conclusions: Our method can be used for calculating clinically relevant cutoffs for (predicted) RFs. The method corrects for the activity of the backbone drug compounds and for therapy failure without drug resistance. Our method's performance is comparable with that of HIVdb. RF cutoffs for the latest version of geno2pheno[resistance] have been estimated with this method.

Key Words: phenotypic drug-resistance testing, clinically relevant cutoffs, antiretroviral therapy, human immunodeficiency virus type 1, statistical methods

(J Acquir Immune Defic Syndr 2017;74:e129–e137)

INTRODUCTION

Drug resistance of HIV type 1 (HIV) can be measured in vitro with phenotypic resistance tests.[1–3] The output of these tests is the fold-change in the median inhibitory concentration between a certain HIV variant and a reference strain. This quantity is called the resistance factor (RF) and is also referred to as phenotype, in this context. Statistical models trained on genotype–phenotype pairs afford accurate prediction of the RF, given a genotype.[4,5] The utility of phenotypes in optimizing cART has been established.[6–9] However, phenotypes require interpretation with respect to the in vivo activity of the tested variant, ie, the range of possible RFs for a drug (dynamic range) has to be divided into a suitable number of clinically meaningful intervals.[8–13] These intervals are drug dependent and defined in cutoffs of the RF. Historically, refinement of the interpretation of the RF has undergone several iterations. Initially, a cutoff per drug was defined in terms of the reproducibility of the phenotypic tests (technical cutoffs).[14] Later, the distribution of RFs of (samples from) therapy-naive patients was used for defining a susceptible-to-resistant RF cutoff for each drug (biological cutoffs).[15,16] Finally, it was recognized that the RF, an in vitro measurement, requires explicit transformation for its intended in vivo application, the prediction of the suppression of the viral load (VL). Efforts in defining clinically relevant cutoffs (short: clinical cutoffs) gave rise to the notion of the division of the dynamic range into the idealized categories susceptible, intermediate, and resistant (SIR), using 2 cutoffs per drug. Susceptible indicates full drug activity, intermediate decreased drug activity, and resistant no drug activity.[8,11,14]

Although drug compounds in cART act in concert, the RF quantifies the activity of a drug in vitro and in isolation. Thus, RF cutoff determination with clinical data requires correcting for the activity of the backbone compounds.[8,11] This correction can be achieved using hard-to-obtain data from (pseudo-) monotherapies for cutoff determination. In pseudo-monotherapies, an examined drug compound is added to a failing regimen. This allows for observing the activity of the drug with reduced influence of the backbone.[8,9,17] Correction for the backbone activity can be also achieved mathematically.[12,18,19] However, no methodology exists for calculating cutoffs that produce clinically meaningful SIR categories without requiring the expert selection of thresholds for drug activity.

In this analysis, we establish a novel methodology for calculating clinically relevant phenotypic resistance cutoffs from routine clinical data. Although drug resistance is a continuum, we aim at calculating cutoffs with which the dynamic range can be divided into intervals that best approximate the SIR categories, as defined above. We correct for the activity of the backbone compounds mathematically, as well as for lack of therapeutic success despite full susceptibility to the drug compounds in a cART, as measured in a phenotypic test or predicted by a genotypic interpretation system.

MATERIALS AND METHODS

Drugs Considered in This Study

The following antiretroviral drugs are considered in this work: lamivudine (3TC), abacavir (ABC), zidovudine (AZT), stavudine (d4T), didanosine (ddI), emtricitabine (FTC), tenofovir (TDF), efavirenz (EFV), nevirapine (NVP), amprenavir (APV), atazanavir (ATV), darunavir (DRV), fosamprenavir (FPV), indinavir (IDV), lopinavir (LPV), nelfinavir (NFV), saquinavir (SQV), and tipranavir (TPV). Other antiretroviral drugs were excluded because of insufficient representation in the EuResist Integrated Database (EIDB)[20] or lack of a model for genotypic drug-resistance interpretation in geno2pheno[resistance] (g2p; http://www.geno2pheno.org; see Ref. 4).

Data Sources

A total of 36,744 protease and reverse-transcriptase sequences from treatment-naive patients were obtained from the Los Alamos National Laboratory Sequence Database (LANLSD[21]; http://www.hiv.lanl.gov/; downloaded on March 31, 2015).

HIV data from routine clinical practice were obtained from 2 sources: the EIDB (http://www.euresist.org; downloaded April 11, 2014) and the HIVdb treatment-change episode (TCE) repository[22] (http://hivdb.stanford.edu; downloaded November 21, 2013). The EIDB contains data from 66,254 patients, including HIV genotypes, VL measurements, CD4 counts, and compounds used in antiretroviral therapies. The HIVdb TCE repository stores 1527 TCEs from 4 data sources, including 58 TCEs from the EuResist database. In the context of the HIVdb TCE repository, a TCE documents relevant clinical parameters concerning a change in the drug compounds of cART.

Distribution of Fold-Change for Therapy-Naive Patients

Nucleotide sequences from LANLSD were aligned, translated, and interpreted with g2p v. 3.4, a data-driven genotypic drug-resistance interpretation system that predicts the RF from the genotype for various protease inhibitors (PIs) and reverse-transcriptase inhibitors (RTIs). Sequences resulting in an alignment error or producing a warning due to missing important sequence regions were discarded. Because we require an RF distribution for susceptible virus variants, sequences containing at least one major drug-resistance mutation[23] were excluded as well, as therapy-naive patients may carry drug-resistant variants.[24,25] The remaining sequences were used for calculating a therapy-naive RF distribution for the drugs considered in this study.

Calculation of Probabilities of Susceptibility (POS)

g2p is a genotypic drug-resistance interpretation system that predicts RFs for PIs and RTIs, given a nucleotide sequence of the protease and reverse-transcriptase genes. The number of genotype-phenotype pairs per drug on which g2p v. 3.4 is trained is given in Table 1, Supplemental Digital Content, http://links.lww.com/QAI/A934. Using g2p v. 3.4, all nucleotide sequences in EIDB were aligned. Sequences resulting in an alignment error or producing a warning due to missing important sequence regions were discarded. The remaining protease and reverse-transcriptase sequences were translated and interpreted. The resulting RF predictions were used for fitting a sigmoid function for approximating the POS given a predicted RF, as described before (probability of resistance[4]). Briefly, a 2-component Gaussian-mixture model is fitted to the RF predictions for each drug. These 2 Gaussians represent the susceptible and resistant viral populations, respectively. A sigmoid function is then fitted to approximate the probability of resistance, given an RF, that can be calculated with these 2 Gaussians. We define the POS as one minus the probability of resistance.

Definition of Treatment Episode (TE)

TEs differ from TCEs in that no treatment change is required, ie, TEs encompass first-line therapies as well. In this analysis, a TE consists of a baseline protease and reverse-transcriptase genotype, a list of drug compounds used in therapy, a follow-up VL and, optionally, a baseline VL. The baseline genotype and the baseline VL must have been obtained no earlier than 90 days before treatment initiation, in line with previously developed definitions of the standard datum.[20,26] However, baseline genotypes for first-line therapies are exempt from this requirement, as the virus has not been subject to selective pressure by drug therapy. In the presence of multiple data points, baseline measurements obtained at a date closest to therapy initiation are preferred. Follow-up VLs must have been obtained between 4 and 12 weeks after therapy initiation, preferring VL measurements whose measurement date is closest to 8 weeks after therapy start. TEs with a VL below 400 HIV-1 RNA copies per milliliter of blood serum (short: cp/mL) or presenting at least 100-fold reduction in the VL are labeled as

successes. TEs for which a baseline VL is available, the follow-up VL is above 400 copies per milliliter, and the VL reduction is less than 100-fold are labeled as failures.[20,26] TEs containing drugs not considered in this analysis or with unboosted PIs (except for nelfinavir) do not satisfy the TE definition.

Definition of Treatment-Failure Episode (TFE)

TEs with baseline sequences predicted by g2p to have high RFs for some of the therapy's drug compounds are rare in EIDB. Therefore, we resort to nucleotide sequences measured during therapy, which imply therapeutic failure, as a certain degree of viral replication is required for sequencing. Toward this end, we define a TFE as follows. A TFE consists of a list of drugs used during therapy and an HIV sequence obtained during the therapy. The therapy is required to have lasted at least 4 weeks, and the genotype is required to have been obtained no earlier than 4 weeks before therapy stop. In TFEs, we assume that the treating clinician changed the therapy based on the results of the genotypic drug-resistance test. TFEs-containing drugs not considered in this analysis or with unboosted PIs (except for nelfinavir) do not satisfy the TFE definition.

Calculation of Probabilities of Success

After application of the TE definition to the EIDB, approximately 10% of the resulting TEs were selected at random and set aside for testing purposes. The remaining TEs were merged with TFE obtained from EIDB, labeled as failures. The resulting data set was used for estimating RF-conditional probabilities of success for each drug (see Methods, Supplemental Digital Content, http://links.lww.com/QAI/A934). Summarizing, TE baseline genotypes were interpreted with g2p, resulting in predicted RFs for the drugs in each TE. For each drug considered in this analysis, RF-conditional success and failure probability densities were estimated using weighted kernel density estimation [27]. When estimating the success density, we correct for the activity of the backbone compounds of therapy by down-weighting RFs from success-labeled TEs with the POS of the backbone compounds. When estimating the failure density, we correct for lack of success in the absence of resistance by down-weighting failures with the POS of their drug compounds. The resulting success and failure densities were used for calculating empirical probabilities of success. Aiming at noise reduction and analytical determination of cutoffs, we fit the following sigmoid function to the probabilities of success. Let $a, b, c, d \in \mathbb{R}_0^+$ be the parameters of the sigmoid function.

$$\hat{P}_i(x) = \frac{a-d}{1 + \exp[-b(x-c)]} + d. \quad (1)$$

For determining the susceptible-to-intermediate (lower) and the intermediate-to-resistant (upper) cutoffs, we use the

TABLE 1. Therapy Characteristics

	EuResist TE Development Set	EuResist TFE Development Set	EuResist TE Test Set	HIVdbTCE TE Test Set
N	8855	1589	807	917
Log baseline VL				
Available	7016 (82%)	NA	765 (95%)	917 (100%)
Mean	4.4	NA	4.5	4.3
Median	4.6	NA	4.6	4.2
IQR	5.2–3.8	NA	5.2–3.8	4.7–3.7
Baseline CD4 count				
Available	6849 (77%)	NA	670 (83%)	917 (100%)
Mean	283.4	NA	298.1	290.9
Median	257	NA	281	258
IQR	410–83	NA	418–120	398–141
No. recorded past treatment lines				
Mean	2.8	4.4	2.5	3.5
Median	1	3	0	3
IQR	4–0	6–1	4–0	5–2
Compound frequency				
NRTI	8703 (98%)	1539 (97%)	792 (98%)	905 (99%)
NNRTI	2933 (34%)	602 (38%)	264 (33%)	398 (43%)
PI	5531 (62%)	754 (47%)	532 (66%)	714 (78%)
First-line therapies	3315 (37%)	0 (0%)	379 (47%)	0 (0%)
Genotypic resistance				
NRTI	3420 (39%)	1169 (74%)	243 (30%)	843 (92%)
NNRTI	2691 (30%)	809 (51%)	195 (24%)	461 (50%)
PI	1558 (18%)	529 (33%)	111 (14%)	553 (60%)
No resistance	4499 (51%)	252 (16%)	485 (60%)	26 (3%)

For therapies in the development and test sets, the following characteristics are summarized above: baseline VL, baseline CD4 count, number of recorded past treatment lines, number of first-line therapies, drug-class composition, and genotypic resistance by drug class.

IQR, interquartile range; NA, not available; NNRTI, nonnucleoside RTI; NRTI, nucleotide/nucleoside RTI; TE, therapy episode; TFE, therapy-failure episode.

TABLE 2. Drug-Compound Distributions

	EuResist TE Development Set	EuResist TFE Development Set	EuResist TE Test Set	HIVdbTCE TE Test Set
3TC or FTC	6069	1093	674	484
ABC	1445	313	157	265
AZT	1756	525	181	169
D4T	833	463	90	407
ddI	928	396	99	338
TDF	3669	442	411	307
EFV	1831	357	204	287
NVP	502	246	60	111
APV/r or FPV/r	336	64	31	117
ATV/r	926	98	117	76
DRV/r	569	27	63	6
IDV/r	136	65	14	103
LPV/r	2306	258	244	223
NFV	344	188	39	72
SQV/r	205	70	27	115
TPV/r	55	13	7	3

The numbers of drug compounds in the development and test sets are tabulated above.

TE, therapy episode; TFE, therapy-failure episode.

roots of the third derivative of the right-hand side of Equation 1, as these are located at the extrema of the curvature of $\hat{P}_i(x)$. Thus, they are located at RF values at which the probability of success is significantly reduced or marginal, respectively. Cutoff determination for each drug was performed with 1000 bootstrap replicates[28] of each subset of TEs including that drug. However, for some drugs, this procedure selected the lower cutoff at an RF below the 95th percentile of the RF distribution of therapy-naive patients. To avoid what could be interpreted as overcalling of intermediate resistance, the lower cutoff for a drug was selected either at the smaller root of the third derivative of the right-hand side of Equation 1 or at the 95th percentile of the RF distribution of therapy-naive patients for that drug, whichever is larger.[17]

Performance Assessment and Comparison

We assess the performance of the cutoffs calculated with the procedures we used on 2 test data sets. The first test data set includes TEs from the EIDB which were not used for cutoff determination (approximately 10% of the total). The second test data set contains TEs extracted from the HIVdb TCE Repository by applying the TE definition to it. For each TE, a genetic susceptibility score (GSS) was calculated. Specifically, RFs predicted for the drug compounds in the TEs were obtained with g2p, subsequently producing SIR labels through the cutoffs. Each drug was assigned an integer score, depending on its SIR label: S → 1, I → 0.5, R → 0. The GSS for a TE amounted to the sum of its integer scores. Performance of the GSS in predicting therapeutic success was calculated in

terms of area under the receiver operating characteristic curve (AUC[29]). For performance comparison, the baseline genotypes of the TEs in the test data sets were interpreted with HIVdb v.6.0.6,[30] resulting in SIR labels for the drug compounds in the TEs. GSS was calculated as described above, and the performance of the GSS in predicting therapeutic success was quantified in terms of area under the curve (AUC) as well. *P*-values were calculated with a 2-sided Wilcoxon signed-rank test.[31] Whenever multiple testing was performed, *P* values were corrected using the Benjamini–Hochberg method.[32]

RESULTS

Among 36,744 nucleotide sequences of the protease and reverse-transcriptase genes downloaded from LANLSD, 43 (<1%) protease and 75 (<1%) reverse-transcriptase sequences were discarded because of alignment problems or sequence-quality warnings. Furthermore, 860 (2.3%) protease and 680 (1.8%) reverse-transcriptase sequences were discarded because they contained some major drug-resistance mutations. In addition, a total of 74,764 nucleotide sequences from EIDB were submitted to g2p for interpretation, some of which did not correspond to either the protease or reverse-transcriptase genes. Of these, 21,199 (28%) were discarded because of alignment problems or because they triggered a sequence-quality warning. RF percentiles for the resulting RF distributions for therapy-naive patients, along with the corresponding POS, can be found in Table 2, Supplemental Digital Content, http://links.lww.com/QAI/A934. On average, RFs at the 95th percentile of the RF distribution of therapy-naive patients have a POS of 0.87 with a SD of 0.13. Baseline characteristics for the TEs and TFEs extracted from EIDB and the HIVdb TCE repository are shown in Table 1, whereas the drug-compound distribution for these data sets is shown in Table 2. On average (SD), follow-up VLs used for labeling TEs were obtained 7.04 (2.48) weeks after treatment initiation, with a median time of 7 weeks after treatment initiation (interquartile range: 9–5 (Figure 1, Supplemental Digital Content, http://links.lww.com/QAI/A934)). Successes accounted for 79% or 83%, respectively, of TEs that were labeled using a follow-up VL obtained 4 or 12 weeks after treatment initiation, respectively. Upper and lower RF cutoffs obtained with the procedure described in Methods are displayed in Table 3. Lower cutoffs for the drugs ddI, TDF, NVP, and NFV were replaced by the 95th percentile of the RF distribution of therapy-naive patients, as they were smaller than this percentile. For performance assessment and comparison, GSS for the TEs in 2 test sets were computed with predicted RFs, discretized with the obtained cutoffs, as well as with HIVdb discrete predictions. On the EIDB test set, our method and HIVdb performed equally well (AUC = 0.68). However, our method ($\mu = 2.55$; $\sigma = 0.88$) produced lower GSS than HIVdb ($\mu = 2.66$; $\sigma = 0.75$; $P < 10^{-16}$). Our method's integer scores for NRTIs ($\mu = 0.83$; $\sigma = 0.35$; vs. $\mu = 0.86$; $\sigma = 0.30$), NNRTIs ($\mu = 0.92$; $\sigma = 0.24$; vs. $\mu = 0.95$; $\sigma = 0.20$), and PIs ($\mu = 0.88$; $\sigma = 0.30$; vs. $\mu = 0.92$; $\sigma = 0.23$) were lower than those produced with HIVdb (corrected $P < 0.0043$). On the HIVdbTCE test set, our method (AUC = 0.63) and HIVdb (AUC = 0.65) performed comparably well. On average, our

TABLE 3. Clinically Relevant Phenotypic Resistance Cutoffs

	Lower Sigmoid Cutoff (SD)	Lower Sigmoid Cutoff Percentile	Selected Lower Cutoff	Upper Sigmoid Cutoff (SD)	Upper Sigmoid Cutoff Percentile
3TC or FTC	3.17 (1.02)	0.99	3.17	5.46 (1.04)	1
ABC	1.78 (1.03)	0.98	1.78	3.03 (1.09)	1
AZT	3.34 (1.05)	0.97	3.34	6.89 (1.08)	1
D4T	1.53 (1.02)	0.99	1.53	2.16 (1.09)	1
ddI	**1.51 (1.1)**	**0.75**	**1.82**	3.8 (1.21)	1
TDF	**1.6 (1.03)**	**0.92**	**1.71**	2.32 (1.03)	1
EFV	3.61 (1.08)	0.98	3.61	7.74 (1.1)	1
NVP	**4.88 (1.08)**	**0.94**	**5.34**	13.35 (1.09)	0.99
APV/r or FPV/r	2.64 (1.08)	0.99	2.64	5.33 (1.17)	1
ATV/r	2.77 (1.11)	0.98	2.77	5.15 (1.19)	1
DRV/r	3.94 (6.89)	1	3.94	14.68 (20.53)	1
IDV/r	3.09 (1.16)	0.99	3.09	6.56 (1.5)	1
LPV/r	2.29 (1.09)	0.99	2.29	7.45 (1.27)	1
NFV	**2.32 (1.09)**	**0.91**	**2.6**	5.78 (1.2)	1
SQV/r	2.64 (1.12)	0.99	2.64	6.13 (1.22)	1
TPV/r	2.36 (3.63)	0.96	2.36	13.65 (5.61)	1

For every drug, sigmoid functions were fitted to empirical probabilities of success. Lower and upper clinically relevant fold-change cutoffs for each drug were selected at the roots of the third derivative of the corresponding sigmoid function (sigmoid cutoffs). Above, these cutoffs are tabulated along with the percentile of the RF distribution of therapy-naive patients to which they correspond. Lower sigmoid cutoffs corresponding to a percentile of the RF distribution of therapy-naive patients below the 95th percentile were replaced by the 95th percentile of this distribution (marked in bold).

method ($\mu = 1.82$; $\sigma = 0.97$) produced lower GSS than HIVdb ($\mu = 2.18$; $\sigma = 0.88$; $P < 10^{-16}$) on the HIVdbTCE test set. Integer scores of our method for NRTIs ($\mu = 0.16$; $\sigma = 0.33$; vs. $\mu = 0.21$; $\sigma = 0.37$), NNRTIs ($\mu = 0.19$; $\sigma = 0.39$; vs. $\mu = 0.19$; $\sigma = 0.39$), and PIs ($\mu = 0.06$; $\sigma = 0.23$; vs. $\mu = 0.07$; $\sigma = 0.24$) were also lower than those produced with HIVdb (corrected $P < 0.043$).

DISCUSSION

In this work, we present a method for calculating clinically relevant cutoffs for the RF that does not depend on expert interpretation. Our method offers the following advantages. (1) The cutoffs produced with this method are tightly coupled with the clinically meaningful definition of lower and upper RF cutoffs for producing SIR categories (see Introduction). We achieve this by translating this definition into an adequate mathematical counterpart. Specifically, our method chooses cutoffs at the curvature extrema of a sigmoid function approximating the probability of success (Fig. 1). By doing so, we circumvent the need for manual selection of (arbitrary) drug activity thresholds.[12] Nevertheless, cutoffs can be chosen manually based on other geometric features of the sigmoid function (see Figure 2 and Table 3, Supplemental Digital Content, http://links.lww.com/QAI/A934) or at specific probabilities of success. (2) Our method corrects for the activity of backbone drug compounds, but does not achieve this using coarse, discrete weights, as with GSS. Instead, we correct for backbone activity in success-labeled TEs with fine-grained RF weights that depend on the POS of all therapy compounds. (3) In failure-labeled TEs, we correct for failures not associated with

resistance, which is especially important because we use data from routine clinical practice for calculating the cutoffs.

We have observed that empirical probabilities of success decrease as log RF values reach their lower limit (Fig. 1). This artifact results from the use of the probability density functions of the normal distribution for estimating success and failure densities (see Methods, Supplemental Digital Content, http://links.lww.com/QAI/A934). However, this artifact is corrected for by subsequent fitting of sigmoid functions to the empirical probabilities of success (Fig. 1).

We exemplify the computation of RF weights for an antiretroviral therapy with 3 drug compounds (see Table 4, Supplemental Digital Content, http://links.lww.com/QAI/A934). For failure-labeled TEs, weights decrease with the sum of the POS for all compounds in the therapy. The maximum weight, 1, is awarded to RFs from TEs for which the sum of POS is equal to 0. Thus, the influence of RFs from failure-labeled TEs is greater the less the susceptibility of the virus to the administered drug compounds, which corrects for failure in the absence of resistance. In contrast, for calculating RF weights for success-labeled TEs, the POS of the drug in question is subtracted from the sum of POS for the backbone drug compounds. The maximum weight, 1, is awarded to RFs of the drug in question indicating full drug susceptibility, whereas the virus is not susceptible to the backbone drug compounds. Hence, RFs from success-labeled TEs have a greater influence, the higher the susceptibility of the drug in question and the lower the susceptibility of the backbone drug compounds. This corrects for the activity of the backbone compounds. For selected drug compounds, a comparison of the densities obtained with and without RF weighting, respectively, is shown in Figure 2. When weighted
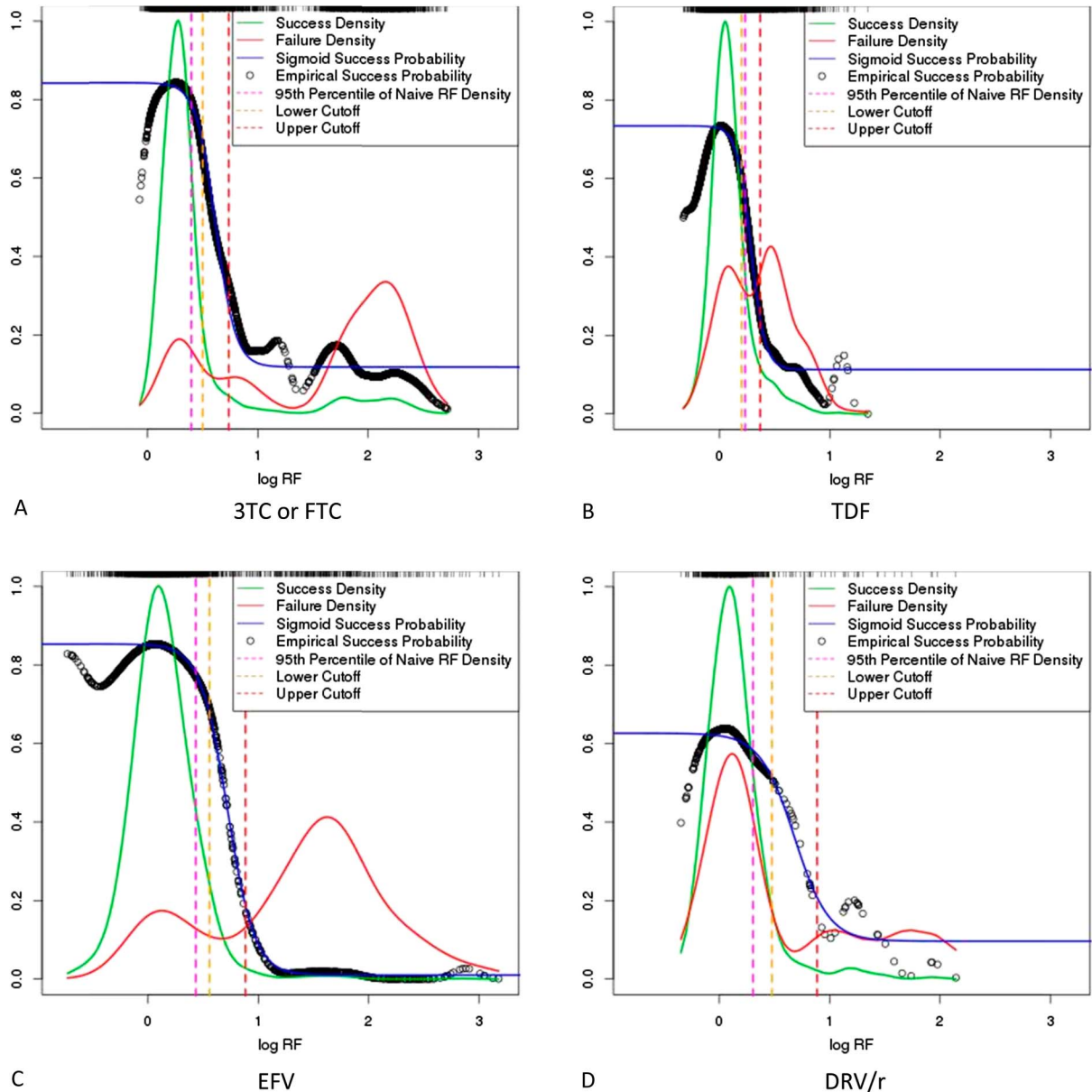
FIGURE 1. Density and probability plots (selected compounds). Weighted kernel density estimation was used for estimating a success and a failure density for each drug compound. These densities were used for calculating RF-dependent empirical success probabilities, to which a sigmoid function was fitted. A lower and an upper RF cutoff were chosen at the roots of the third derivative of the sigmoid function. However, the lower cutoff was replaced by the 95th percentile of the RF distribution of therapy-naive patients, if it was lower than this percentile. Success and failure densities, as well as empirical and sigmoid success probabilities are plotted above for the drug compounds 3TC or FTC (A), TDF (B), EFV (C), and DRV/r (D). Ninety-fifth percentiles of the RF factor distribution, as well as lower and upper cutoffs determined with the roots of the third derivative are indicated by vertical dashed lines. Resistance-factor density is also depicted as a rug plot at the top of each individual plot. Note that for producing these joint depictions of probabilities of success and of success and failure densities, success and failure densities were rescaled to the interval between zero and one. This was accomplished by dividing the values of the densities by the largest value of both densities.

failure densities are compared with their unweighted counterparts, one can observe a shift of the probability mass from lower RF ranges to higher ones. This effect is especially pronounced in the failure densities of ddI and IDV because their side-effect profiles are less favorable than those of 3TC/

FTC and DRV.[33–37] Thus, for cutoff calculation, our method downweights therapy failures that are not caused by drug resistance. Comparative examination of the weighted and unweighted success densities for ddI and IDV makes the effects of the weighting procedure apparent. Specifically, the
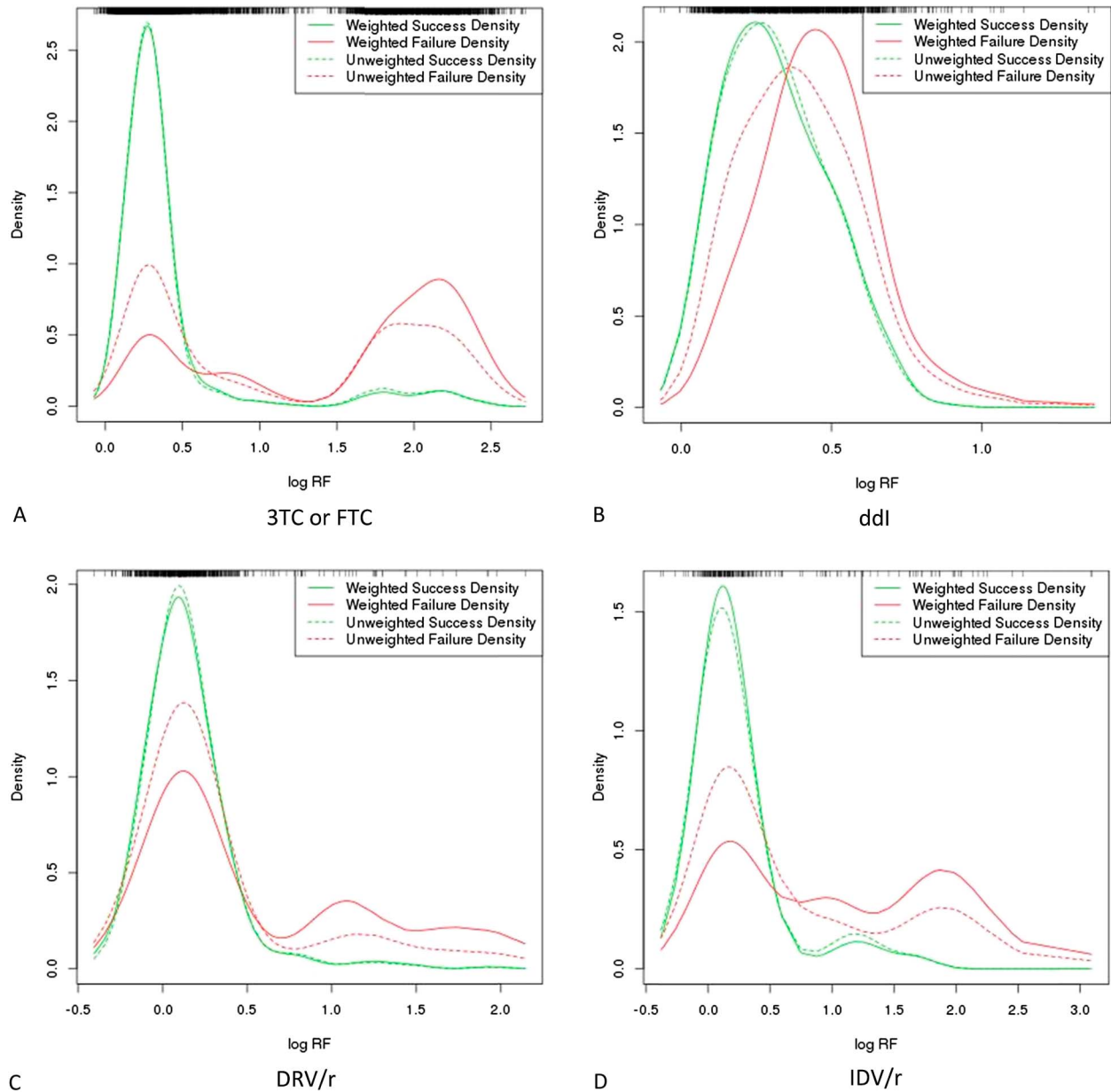
FIGURE 2. Comparison of weighted and unweighted densities (selected compounds). Empirical success and failure densities were estimated with weighted kernel density estimation, as described in Methods. Above, the weighted densities for the drug compounds 3TC or FTC (A), ddI (B), DRV/r (C), and IDV/r (D) are plotted along with their unweighted versions. Resistance-factor density is also depicted as a rug plot at the top of each individual plot.

weighting procedure compensates for the effect of the backbone drug compounds by down weighting the success densities at higher RF values.

The definition of SIR categories takes the activity of a drug into account, and (intentionally) does not account for the differential capacities of antiretroviral drugs to elicit virological success. For instance, at the same probability of success, the virus may be classified as susceptible to one drug whereas it is classified as intermediate for another drug. Taking this into consideration, cutoff determination through mathematical optimization is hampered by the difficulty of finding an adequate objective function for optimization that is universal for all drugs. In the past, cutoff calculation has been attempted by maximizing the AUC quantifying the performance of prediction of therapeutic success.[19] Briefly, cutoffs are iteratively optimized by producing a GSS with a given set of cutoffs and subsequently modifying these cutoffs such that there is an increase in the AUC. This procedure involves the calculation of cutoffs for all drugs in the data set simultaneously, and may be less effective in separating the effects of individual drugs. Furthermore, the use of such an objective function is bound to maximize predictive performance in the

GSS while producing inconsistencies of the applied SIR categories. We avoid this undesired effect by estimating cutoffs for each drug separately. Nonetheless, the separation of the contributions of individual drugs to therapeutic success remains challenging, as antiretroviral drugs may interact with each other.[38,39]

Except for DRV/r and TPV/r, the SD of the calculated cutoffs is at most 1.5. The higher variability of the cutoffs for DRV/r can be explained by the drug's comparatively high barrier to resistance.[33] Our data set contains only few DRV/r-containing TEs with an RF above the upper cutoff (Fig. 1). Therefore, significantly different numbers of TEs with a high resistance to DRV/r at baseline will be selected across bootstrap replicates, which results in increased variability. The increased variability for the cutoffs for TPV/r is due to the low numbers of TEs containing this drug (Table 2). We tested the calculated cutoffs on 2 test sets. The performance of cutoffs was equal or comparable with that of HIVdb. However, the GSS calculated with the cutoffs were significantly smaller than those calculated with HIVdb. Therefore, our cutoffs have a higher tendency to label an RF with nonsusceptible SIR categories than HIVdb. We interpret this to be a result of the strict adherence to SIR categories in the cutoff estimation procedure. Specifically, genotypic drug-resistance interpretation systems produced by experts may take the propensities of individual drugs to elicit therapeutic success into account more strongly. This results in fewer resistant predictions for drugs with a high potency and a high barrier to resistance.

In this work, we present a method for calculating clinically relevant cutoffs for (predicted) RFs. The method corrects for the activity of the backbone drug compounds and for therapy failure without drug resistance. A sigmoid function is used for estimating the probability of therapeutic success, given an RF. Therefore, RF cutoff selection can be based on geometric features of the sigmoid function or on thresholds of the probability of success. Validation with 2 clinical data sets showed that the performance of our method was comparable with that of a popular rules-based drug-resistance interpretation method. The method has been applied for estimating RF cutoffs for the latest version of g2p, v. 3.4.

REFERENCES

1. Hertogs K, de Béthune MP, Miller V, et al. A rapid method for simultaneous detection of phenotypic resistance to inhibitors of protease and reverse transcriptase in recombinant human immunodeficiency virus type 1 isolates from patients treated with antiretroviral drugs. *Antimicrob Agents Chemother.* 1998;42:269–276.
2. Walter H, Schmidt B, Korn K, et al. Rapid, phenotypic HIV-1 drug sensitivity assay for protease and reverse transcriptase inhibitors. *J Clin Virol.* 1999;13:71–80.
3. Petropoulos CJ, Parkin NT, Limoli KL, et al. A novel phenotypic drug susceptibility assay for human immunodeficiency virus type 1. *Antimicrob Agents Chemother.* 2000;44:920–928.
4. Beerenwinkel N, Däumer M, Oette M, et al. Geno2pheno: estimating phenotypic drug resistance from HIV-1 genotypes. *Nucleic Acids Res.* 2003;31:3850–3855.
5. Vermeiren H, Van Craenenbroeck E, Alen P, et al. Prediction of HIV-1 drug susceptibility phenotype from the viral genotype using linear regression modeling. *J Virol Methods.* 2007;145:47–55.
6. Altmann A, Sing T, Vermeiren H, et al. Advantages of predicted phenotypes and statistical learning models in inferring virological response to antiretroviral therapy from HIV genotype. *Antivir Ther.* 2009;14:273–283.
7. Flandre P, Chappey C, Marcelin AG, et al. Phenotypic susceptibility to didanosine is associated with antiviral activity in treatment-experienced patients with HIV-1 infection. *J Infect Dis.* 2007;195:392–398.
8. Zolopa AR. Incorporating drug-resistance measurements into the clinical management of HIV-1 infection. *J Infect Dis.* 2006;194 (suppl 1):S59–S64.
9. Miller MD, Margot N, Lu B, et al. Genotypic and phenotypic predictors of the magnitude of response to tenofovir disoproxil fumarate treatment in antiretroviral-experienced patients. *J Infect Dis.* 2004;189:837–846.
10. Brun-Vézinet F, Costagliola D, Khaled MA, et al. Clinically validated genotype analysis: guiding principles and statistical concerns. *Antivir Ther.* 2004;9:465–478.
11. Geretti AM. Clinical implications of HIV drug resistance to nucleoside and nucleotide reverse transcriptase inhibitors. *AIDS Rev.* 2006;8:210–220.
12. Winters B, Montaner J, Harrigan PR, et al. Determination of clinically relevant cutoffs for HIV-1 phenotypic resistance estimates through a combined analysis of clinical trial and cohort data. *J Acquir Immune Defic Syndr.* 2008;48:26–34.
13. Vingerhoets J, Tambuyzer L, Azijn H, et al. Resistance profile of etravirine: combined analysis of baseline genotypic and phenotypic data from the randomized, controlled phase III clinical studies. *AIDS.* 2010;24:503–514.
14. Perno C-F, Bertoli A. Clinical cut-offs in the interpretation of phenotypic resistance. In: Geretti AM, ed. *Antiretroviral Resistance in Clinical Practice.* London, United Kingdom: Mediscript; 2006. Available at: http://www.ncbi.nlm.nih.gov/books/NBK2254/. Accessed October 27, 2015.
15. Harrigan PR, Montaner JS, Wegner SA, et al. World-wide variation in HIV-1 phenotypic susceptibility in untreated individuals: biologically relevant values for resistance testing. *AIDS.* 2001;15:1671–1677.
16. Parkin NT, Hellmann NS, Whitcomb JM, et al. Natural variation of drug susceptibility in wild-type human immunodeficiency virus type 1. *Antimicrob Agents Chemother.* 2004;48:437–443.
17. Shulman NS, Hughes MD, Winters MA, et al. Subtle decreases in stavudine phenotypic susceptibility predict poor virologic response to stavudine monotherapy in zidovudine-experienced patients. *J Acquir Immune Defic Syndr.* 2002;31:121–127.
18. Winters B, Van Craenenbroeck E, Van der Borght K, et al. Clinical cut-offs for HIV-1 phenotypic resistance estimates: update based on recent pivotal clinical trial data and a revised approach to viral mixtures. *J Virol Methods.* 2009;162:101–108.
19. Altmann A, Däumer MP, Pironti A, et al. Application of a probabilistic optimization method for deriving clinical cutoffs for (predicted) drug resistance phenotypes. In: *Reviews in Antiviral Therapy and Infectious Diseases.* Virology Education; Sorrento, Italy: 2010.
20. Zazzi M, Kaiser R, Sönnerborg A, et al. Prediction of response to antiretroviral therapy by human experts and by the EuResist data-driven expert system (the EVE study). *HIV Med.* 2011;12:211–218.
21. Foley B, Leitner T, Apetrei C, et al. *HIV Sequence Compendium 2013.* Los Alamos, NM: Theoretical Biology and Biophysics Group, Los Alamos National Laboratory; 2013. Available at: http://www.hiv.lanl.gov/. Accessed March 31, 2015.
22. Rhee S-Y, Blanco J, Liu TF, et al. Standardized representation, visualization and searchable repository of antiretroviral treatment-change episodes. *AIDS Res Ther.* 2012;9:13.
23. Johnson VA, Calvez V, Gunthard HF, et al. Update of the drug resistance mutations in HIV-1: March 2013. *Top Antivir Med.* 2013;21:6–14.
24. Garcia-Lerma JG, MacInnes H, Bennett D, et al. Transmitted human immunodeficiency virus type 1 carrying the D67N or K219Q/E mutation evolves rapidly to zidovudine resistance in vitro and shows a high replicative fitness in the presence of zidovudine. *J Virol.* 2004;78:7545–7552.
25. Bennett DE, Camacho RJ, Otelea D, et al. Drug resistance mutations for surveillance of transmitted HIV-1 drug-resistance: 2009 update. *PLoS One.* 2009;4:e4724.
26. Zazzi M, Incardona F, Rosen-Zvi M, et al. Predicting response to antiretroviral treatment by machine learning: the EuResist project. *Intervirology.* 2012;55:123–127.
27. Duong T. Ks: kernel density estimation and kernel discriminant analysis for multivariate data in R. *J Stat Softw.* 2007;21:1–16.
28. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* 2nd ed. New York, NY: Springer; 2011.

29. Sing T, Sander O, Beerenwinkel N, et al. Visualizing classifier performance in R. *Bioinformatics*. 2005;21:3940–3941.
30. Liu TF, Shafer RW. Web resources for HIV type 1 genotypic-resistance test interpretation. *Clin Infect Dis.* 2006;42:1608–1618.
31. Wilcoxon F. Individual comparisons by ranking methods. *Biom Bull.* 1945;1:80.
32. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol.* 1995;57:289–300.
33. Fenton C, Perry CM. Darunavir: in the treatment of HIV-1 infection. *Drugs.* 2007;67:2791–2801.
34. Perry CM, Balfour JA. Didanosine: an update on its antiviral activity, pharmacokinetic properties and therapeutic efficacy in the management of HIV disease. *Drugs.* 1996;52:928–962.
35. Plosker GL, Noble S. Indinavir: a review of its use in the management of HIV infection. *Drugs.* 1999;58:1165–1203.
36. Perry CM, Faulds D. Lamivudine: a review of its antiviral activity, pharmacokinetic properties and therapeutic efficacy in the management of HIV infection. *Drugs.* 1997;53:657–680.
37. Frampton JE, Perry CM. Emtricitabine: a review of its use in the management of HIV infection. *Drugs.* 2005;65:1427–1448.
38. Wainberg MA, Brenner BG, Turner D. Changing patterns in the selection of viral mutations among patients receiving nucleoside and nucleotide drug combinations directed against human immunodeficiency virus type 1 reverse transcriptase. *Antimicrob Agents Chemother.* 2005;49:1671–1678.
39. Zeldin RK. Pharmacological and therapeutic properties of ritonavir-boosted protease inhibitor therapy in HIV-infected patients. *J Antimicrob Chemother.* 2003;53:4–9.