Title: Improving the role and contribution of pharmacokinetic analyses in antimalarial drug clinical trials

Running Title: Pharmacokinetics in antimalarial drug trials

Katherine Kay*, Eva Maria Hodel and Ian M. Hastings

Authors affiliation: Liverpool School of Tropical Medicine, Pembroke Place, Liverpool, L3 5QA, UK

*Corresponding author. Mailing address: Liverpool School of Tropical Medicine, Pembroke Place, Liverpool L5 3QA, UK; Telephone: 0151 705 2547; Fax: 0151 705 3371; E-mail: kwinter@liverpool.ac.uk
Abstract

It is now World Health Organization (WHO) policy that drug concentrations on day 7 be measured as part of routine assessment in antimalarial drug efficacy trials. The rationale is that this single pharmacological measure serves as a simple and practical predictor of treatment outcome for long half-life antimalarial drugs. Herein we review theoretical data and field studies and conclude that the day 7 drug concentration (d7c) actually appears to be a poor predictor of therapeutic outcome. This poor predictive capability combined with the fact that many routine antimalarial trials will have few or no failures means there appears little justification for this WHO recommendation. Pharmacological studies have a huge potential to improve antimalarial dosing and we propose study designs that use more focussed, sophisticated and cost-effective ways of generating these data than the mass collection of single d7c concentrations.

(139 words)
Introduction

The provision of effective antimalarial drugs is a cornerstone of public health policy in the majority of developing countries. Historically, the evolution of drug resistance undermined the effectiveness of first line therapies (e.g., (1-3)) and failing drugs were retained for much too long (4). This lead to a surveillance strategy of using regular monitoring to confirm the continued efficacy and effectiveness of local first line therapies (5). Resistance is conventionally regarded as a binary trait where infections may be classified as ‘resistant’ or ‘susceptible’ depending on patient therapeutic outcome. However, it is becoming more widely recognised that drug resistance in malaria is not a strictly binary trait of resistance/susceptible but is a probabilistic trait with therapeutic outcome depending on the interaction between three critical factors: (i) the level of parasite resistance (described by its pharmacodynamics, PD, such as half maximal inhibitory concentration (IC50)) (ii) the amount of drug the patient takes and how s/he subsequently processes it (the pharmacokinetics, PK, such as drug distribution and elimination rate), and (iii) the levels of human immunity. The latter is usually ignored in antimalarial drug deliberations (but see (6, 7)) on the basis that drugs should work even in non-immune patients; this makes the prediction of therapeutic outcome a function solely of PK and PD. In principle, this approach should allow us to (crudely) distinguish drug ‘failure’ from drug ‘resistance’. Treatment ‘failures’ are the result of human factors such as low drug concentrations due to, for example, inadequate drug dosing, unnoticed vomiting, or natural human PK variation. Drug ‘resistance’ occurs when infections survive treatment due to genetically-encoded parasite factors such as reduced sensitivity. This realisation led to suggestions (8, 9) that drug
concentrations measured seven days after treatment (day 7 drug concentration; \(d_{7c}\)) could be used to distinguish drug ‘failures’ from drug ‘resistance’. Day 7 was justified for three main reasons (outlined in (10)) (i) feasibility, as day 7 is one of the several days on which routine patient follow-up should be performed in antimalarial drug trials, (ii) pharmacodynamics, because, in theory, if \(d_{7c}\) of slowly eliminated antimalarials are at least twice the minimum parasiticidal concentration, all the infecting parasites should be eliminated, (iii) pharmacokinetics, as in theory day 7 drug exposure is determined only by variation in the elimination rate constant. Measurement of \(d_{7c}\) in clinical trials of antimalarial drugs has been widely promoted and is now supported by well-resourced reference laboratories (11). The World Health Organisation (10) have also repeated these assertions stating that “Measurement of concentrations of longer-acting antimalarial drugs on day 7 following initiation of treatment should be considered a routine part of trials” (page 70) because “The drug concentration on day 7 is predictive of the outcome” (page 73). They further assert “Measurement of the blood, serum or plasma concentration of slowly eliminated antimalarial drugs (i.e. terminal elimination half-life > 2 days) at a single time is simple and might be a better determinant of therapeutic response than the total AUC” (page 73).

We will argue that these assertions regarding the predictive ability of \(d_{7c}\) are clearly contradicted by PK/PD simulations, see below, and by re-analysis of field data (Supplementary Information) which suggest exactly the opposite i.e. that \(d_{7c}\) are actually rather poor predictors of therapeutic outcome.
What is a ROC curve?

Receiver-operator characteristic (ROC) curves originated in radio-technology but are now widely used in medical research to quantify the predictive value of a measurement, in our case, the ability of d7c to predict therapeutic outcome (12). The X-axis is 1-specificity and the Y-axis is sensitivity. It is clearer to re-label the X-axis simply as ‘specificity’ and reverse the axis, as in apparent in the Figure 1. Sensitivity is defined as the number of patients with ‘low’ d7c who fail treatment, divided by the total number of patients who fail treatment. Specificity is defined as the number of patients with ‘normal’ d7c cured, divided by the total number of patients cured.

The ROC curve plotted as continuous blue line on Figure 1 is actually a dot plot rather than a true algebraic function. Each d7c cut-off value is assessed for sensitivity and specificity and plotted onto the co-ordinates; the software then links these points with a line. In this example, we assume d7c lies between 0 and 100 ng/ml and the assumption is that d7c is used to predict therapeutic failure. Each hypothetical d7c cut-off value is assessed and included on the plot (blue numbers): high cut-off values, such as 60 and 80, result in most patients being classified with ‘low’ d7c, resulting in high sensitivity (the large group of ‘low’ d7c patients includes most failures) but poor specificity (many cured patients will be in this class of ‘low’ d7c), while low cut-off values, such as 10 and 30, have low sensitivity but higher specificity. In this example, the hypothetical cut-off value of 45 ng/ml seems the best compromise but the choice of cut-off is an objective choice that depends on the weighting given to consequences of wrong
classifications. Irrespective of the choice of cut-off value, the closer the ROC curve approaches the upper left-hand corner of the graph (i.e. high sensitivity and high specificity), the better its diagnostic capability. For future reference, the enumerated black points on Figure 1 are the sensitivity and specificity of the cut-off values of d7c reported or extracted from the literature (Supplementary Information); none of these points are remotely near the upper left-hand corner of the plot which would have indicated good predictive capability.

ROC curves allow an objective measure of the predictive capability of a diagnostic test (in this case, the ability of low d7c to predict treatment failure). In practice, d7c are often measured in large laboratory batches and only become available after the follow-up period. In this case d7c serves as an explanation, rather than a predictor, of therapeutic outcome. As would be expected from a statistical analysis, ROC curve analysis is unaffected by these semantic differences and properly quantifies both d7c predictive and explanatory roles. The closer the area under the ROC curve (auROC) is to 1, the better the diagnostic test performs (an auROC of 1 implies the test is perfectly accurate). An area under the ROC curve (auROC) of 0.5 indicates the diagnostic test has no predictive value (i.e. the test is equivalent to relying on pure chance) and is represented by the solid black diagonal line in Figure 1. The consensus for classifying the accuracy of a diagnostic test is the use of the “traditional” academic point system with 0.90–1 = excellent, 0.80–0.90 = good, 0.70–0.80 = fair, 0.60–0.70 = poor, 0.50–0.60 = fail (we have been unable to find an academic citation for this ‘tradition’ but it can be found on websites e.g. http://gim.unmc.edu/dxtests/roc3.htm). The red line in Figure 1 represents our simulated ROC analysis (6) for lumefantrine with an auROC of 0.615 (0.596-0.633; 95% CI). Using the
How good is day 7 concentration as a predictor of therapeutic outcome?

Intuitively, d7c can act as a good predictor of therapeutic outcome only if it is the dominant parameter determining outcome. There are a large number of interacting factors that ultimately determine therapeutic outcome of treatment with a typical artemisinin combination therapy (ACT; the currently recommended class of first-line antimalarials). ACTs typically contain two or three distinct drugs: the artemisinin parent drug (if given as artesunate or artemether), the artemisinin active metabolite DHA, and the partner drug (which may also have an active metabolite e.g. amodiaquine). The degree of parasite sensitivity to each drug (its PD profile) is typically described by Michaelis-Menten dynamics defined by three factors: IC50, maximal kill rate and slope of the dose response curve. The PK profile of each drug is described by three main factors: bioavailability, volume of distribution and elimination rate (plus a series of absorption and conversion rates, and distribution across separate physiological compartments (e.g. (10)) that we ignore in the interest of simplicity). This results in six main PK/PD parameters per drug, and hence 12-18 for a combination therapy, all of which contribute to therapeutic outcome. Each of these parameters shows substantial variability (10): human PK typically varies over a three-fold range (discussed further in (10)) while parasite IC50 for a drug typically varies 50 to 1,000 fold (e.g. Figure 3 of (13)). Human immunity also plays a substantial role in the outcome of treatment.
role in outcome (14, 15). Another determinant of therapeutic success, often overlooked, is that of multiclonal infections. Malaria infections consisting of several genetically distinct clones are commonly observed (up to around 8 clones per infection in higher transmission areas e.g. (16)). The clones are likely to vary in PD and therapy must clear all the infections including the most resistant. Hence increasing the number of clones (quantified as a patient’s multiplicity of infection, MOI) will increases failure rates(17). Consequently, higher MOI introduces another factor contributing to therapeutic outcome (17). In summary, the substantial variation in PK/PD, human immunity levels, and MOI will obscure the relationship between d7c and therapeutic outcome. We therefore decided to review the evidence base for using d7c as a predictor of therapeutic outcome and used two strategies: PK/PD modelling, and critical appraisal of clinical data previously invoked as support for d7c predictive ability.

The consensus method of quantifying the predictive capability of a diagnostic measure is by receiver operating characteristic (ROC) analysis (12) as described above. We investigated the predictive ability of d7c by analysing simulated data of antimalarial drug treatment outcome generated with the PK/PD model described previously (6, 18). PK/PD modelling has the advantage that we know, and can alter, the factors underlying treatment outcome. This “mechanism-based PK/PD modelling” was recently reviewed in (19), and is widely used in infection biology. These PK/PD simulations have been applied by other investigators to malaria (20-27) and recently extended by us to incorporate factors such as multiple dosing and drug conversion (6, 10, 28).
Our simulations showed that d7c is, as expected, generally a good proxy for drug exposure as measured by the area under the drug/time concentration curve (AUC): the correlation coefficient of d7c with AUC measured up to 100 days post-treatment was 0.98 for lumefantrine (LF), 0.94 for chloroquine (CQ), 0.93 for piperaquine (PPQ), and 0.92 for mefloquine (MQ); details in Table S3.1 of (6). Population attributable risk percentage (PAR%) simulations showed that between 3% (artesunate (AS) plus MQ) and 17% (dihydroartemisinin (DHA) plus PPQ) of failures could be avoided if adequate drug levels were achieved throughout the patient population (details in Tables 6 of (6)). The simulations also showed ‘low’ d7c was associated with a statistically significantly increased odds of failing treatment (details in Tables 5 and 6 of (6)). Furthermore, a simulated clinical trial of AS-MQ suggested that low d7c was a more important risk factor in treatment outcome (measured using the Wald statistic) than patients initial parasitaemia, high malaria transmission intensity and patient age <5 years (details in Tables 5 of (6)).

Despite the clear association of d7c with overall drug exposure and treatment outcome, simulations show d7c would have a very poor predictive capability when evaluated by their auROC curve (see above and Figure 1); generally in the range of 0.55 to 0.65 (Tables 6 and S3.2 of (6)). This was consistent with clinical data (see below). We also noted that even a d7c with a very poor predictive capability, quantified by its auROC, could still have a significant association with outcome as quantified by its ‘p’ value; for example we predicted a p=0.001 associated with a d7c cut-off value (<15th centile) for MQ (Table 5 of (6)). This apparent discrepancy arises from differing roles of a ‘p’ value and a ROC curve. The ROC curves quantifies the extent to which d7c
is a good (or bad) diagnostic predictor of therapeutic outcome. In contrast, the ‘p’ value simply
tests a null hypothesis i.e. that d7c has absolutely no association with therapeutic outcome. The
latter is, hopefully, unlikely so it is entirely consistent that low ‘p’ values can be associated with
d7c whose ROC curves reveal a lack of any useful predictive value. The use of ‘p’ values to
identify ‘target’ or ‘cut-off’ values of d7c is therefore problematic and is discussed further in the
Supplementary Information.

Clinical data were then collated and reviewed to ascertain whether they follow the patterns
predicted by PK/PD modelling. We identified and read all the papers we could find that
reported use of d7c as predictors of therapeutic success (Supplementary Information).
Disappointingly, no authors reported auROC and all relied on the use of ‘p’ values to justify a
d7c cut-off which, as described above, tests the hypothesis of no association rather than
quantifying predictive ability. Some papers reported success/failure rates associated with the a
d7c ‘threshold’ which enabled us to make a crude reconstruction of auROC curve (dashed black
line on Figure 1); the auROC values we extracted from these clinical reports lay in the range 0.6
to 0.7 which is disappointingly low but entirely consistent with the values obtained by PK/PD
simulation described above and previously (6). In summary, the empirical evidence base for
statements such as “drug concentration on day 7 is predictive of the outcome” (e.g. page 73 of
(29)) appears weak at best. In fact, our review of the literature reveals that most clinical data
point towards d7c being an extremely poor predictor of therapeutic outcome.
Potential pitfalls when using day 7 concentration in clinical trials

It is a current WHO recommendation that d7c be measured in antimalarial drug clinical trials to assess drug exposure. These data are being collected and reported so it is constructive to consider what role they can play in clinical trials and, equally important, to discuss the dangers that may arise from their uncritical use.

One potential use of d7c is to detect patients who poorly adhere to their drug regimens; once identified, such patients could be removed from the analysis and drug cure rates calculated separately for completely- and poorly-adherent patients. Figure 2 shows our simulation of the effect of poor adherence on the d7c of PPQ, given as 3 daily doses with DHA, in 10,000 patients. As expected the mean/median d7c declines as adherence decreases but the large amount of natural variation in human PK means that d7c is unlikely to be diagnostic of poor-adherence except in cases where patients take only the first of the three doses; obviously the proportion of such patients should be very small to negligible in most clinical trials.

One tempting way of using d7c data is to simply remove the patients failing treatment who have ‘low’ d7c from a clinical trial; one obvious justification is that they may have been adhering poorly to the treatment regimen. However, removal of only these drug failures would overestimate the true drug efficacy as it might for example exclude fully adherent patients with particularly extreme PK parameters. In some cases, this overestimation may keep treatment efficacy above the 95% threshold of initiating policy change (30) thus removing the necessity to
consider a replacement drug, an interpretation with potentially fatal consequences. This bias
towards over-estimating drug efficacy can be avoided by removing all patients with low d7c but
the ROC curve analysis and PK/PD simulation suggest this will also remove a lot of drug
successes, thus the results would be unbiased but sample size will fall and the confidence
intervals around drug effectiveness would increase. This problem arises because d7c is such a
poor predictor of outcome and is best illustrated using a trivial analogy: suppose we thought,
erroneously, that patients born on a Monday are more likely to fail treatment. Searching
through records of patients failing treatment and removing all failures born on a Monday will
reduce the number of failures and hence artificially increase the apparent drug efficacy; the
correct strategy would be to remove all patients born on a Monday irrespective of their
therapeutic outcome. This will eliminate the bias but reduce sample size analysed in the clinical
trial and hence increase the confidence interval(s) around the estimates of treatment efficacy.

There are considerable dangers in using d7c of a drug to identify a single d7c cut-off that can be
used to distinguish between ‘adequate’ and ‘inadequate’ dosing and hence used as a ‘target’
concentration. Informative threshold concentrations only occur with high values of auROC
where the ROC curve is non-linear and shows a clear cut-off point with high specificity and
sensitivity (e.g. the hypothetical drug concentration cut-off of 45 ng/ml on the blue ROC on
Figure 1). Note that it would be inappropriate to use the concentration with the lowest ‘p’
value as a cut-off because, as discussed above and in the Supplementary Information, the p
value simply tests the assumption of no association and, critically, is affected by both the size of
effect and sample sizes. Cut-offs between 175 and 600 ng/ml for LF have been suggested (for
review see page 74 in (29)) typically with ‘p’ values of <0.001. The near-linear relationship
between d7c and the chance of falling treatment means that virtually any ‘cut-off’ value can be
chosen and supported by statistics (‘p’ value, odds ratio, etc.) to support its selection. It is
important to realise that, unless the d7c is associated with a large auROC, the choice is largely
arbitrary.

Drug dosage, regimen and adherence (which jointly determine d7c) are the only factors we can
actually control in malaria therapy, all the other PK and PD parameters are encoded by human
and malaria genetics, so it is important to assess the impact of changing dosages. Analyses of
clinical trials consistently show that the higher the dose given, and d7c attained, the better the
chance of successful treatment; moreover this relationship appears to be roughly linear (e.g.
Figure 5 of (31), and Figure 4 of (32)). On one level this result is entirely expected and trivial
(33): all antimalarials have had their dosages increased after their initial deployment and
Guinea-Bissau overcame its problem of CQ-resistant malaria by simply doubling the dosage of
CQ given to patients (34, 35). The reason dosages are not routinely increased is because of
concerns over toxicity (and no other country has followed Guinea-Bissau’s policy of doubling CQ
dosage); hence the results emphasise the need for better toxicity data on antimalarials drugs
which is, in our opinion, a under-researched area.

Finally, WHO mandate drug efficacy must be >95% (36) so the expected failure rates will be low
in the routine efficacy monitoring trials and any diagnostic ability of d7c will probably never be
observed. It is pointless doing a ROC analysis when few or no treatment failures occur in a trial
so it is not clear how these d7c can be properly incorporated except by large meta-analyses of numerous trials (e.g. pooled analysis of day 7 PPQ concentrations (32)). Several authors have already noted that d7c will be of little value when cure rates are very high. For example, White et al (9) note that “Relationships between PK variables and cure rates are not evident when cure rates are very high. Such relationships are apparent only when resistance develops or doses are inadequate”. Most clinical trials of antimalarial are now testing highly effective ACTs, often in non-inferiority trials, to allow decisions to be made on factors such as cost, side-effects, ease of regimens, length of post-treatment prophylaxis, and so on. It therefore follows that d7c collected in these trials will be largely superfluous until resistance starts to arise, in which case of course, the ACT would have to be replaced.

In summary, there are a number of tempting but ultimately incorrect ways of using d7c to interpret clinical trial data. It appears that d7c has no consistently clear predictive cut-off on a ROC curve so that subsequent analyses tend to draw the rather trivial conclusion that the more drug given to patients, the higher the subsequent d7c, and the more chance they have of being cured. The question therefore arises as to whether we can identify more informative and/or cost-effective ways of bringing PK/PD measures into current clinical trial methodology.

Can we gather more informative pharmacological data in clinical trials?
Designing antimalarial therapies would be straightforward if all people and all parasites were identical. The enormous variation in parasite sensitivity to drugs, human variation in how a drug is absorbed, distributed and metabolised, and how toxicity may occur makes the rationale design of drug regimens enormously complex (37). It is this complexity that limits the use of d7c to a proxy for drug exposure. The subsequent realisation that drug exposure is not the sole predictor of failure forces us to consider what other factors contribute to failure and how we can collect and integrate this information during routine clinical trials.

The first requirement would be to move away from taking drug measurements at a single time point. Treatment is a dynamic process that requires repeat measurements either by intensive sampling or, more likely, through ‘sparse’ sampling and appropriate population PK analysis (29). The use of sparse sampling and population PK modelling is highly informative as it allows PK parameters such as absorption rates, volumes of distributions, elimination rates and the number of physiological compartments to be determined in different human populations, alongside the intra- and inter-individual variation of these parameters (e.g. (38)). PK/PD modelling has been successfully used in other organisms to optimise dosing strategies (e.g. (39)) and it seems reasonable to adopt the same approach for malaria. However, this requires the measured mean and associated variation in individual pharmacological parameters to determine treatment outcome. The d7c would be inadequate for such modelling as it is a composite measure determined by several distinct PK parameters that needs to be split into its constituent parts i.e. dosage, bioavailability, volume of distribution (Vd) and elimination rate.

Day 7 concentrations could still be measured in each patient (day 7 is a routine surveillance
time point) but be augmented by additional sampling around this day to generate the sparse datasets required for PK analysis. PK parameters alone cannot address the issue of how to deal with natural variation in parasite drug sensitivity which typically ranges 10 to 100-fold even in the absence of ‘major’ mutations controlling drug resistance (e.g. Figure 3 of (13)). This implies that PK data need to be accompanied by some indication of the drug sensitivity of local malaria population(s). This would be best achieved by in vitro measurement of fresh parasite isolates. These strategies have been worked out in some detail (reviewed in (40)) and are now part of the WWARN depository system.

It would then be necessary to integrate these separate sources of data into a comprehensive framework linking the parameters to therapeutic outcome. We would suggest PK/PD modelling as a framework (6, 41) but are not dogmatic, provided that some sort of coherent framework is used to link parameters to therapeutic outcome (19, 39). The application of such a framework has several major advantages.

- Data can be combined from different trials, locations and patient groups to calibrate these models. This greatly increases the scope and value of each dataset including historical trials that may have collected PK or PD data for different reasons. In particular it also allows current ACT trials, where few or no failures may occur, to contribute data useful for drug regimen optimisation.

- Clinical trials are run in highly controlled settings making it problematic to extrapolate their efficacy estimated under near-ideal conditions into effectiveness under real-life
conditions. PK/PD modelling can be used to make these extrapolations. Obvious applications are to investigate how robust the regimens are to poor adherence (42); it would be clearly unethical to give patients incomplete dosages in a clinical trial but we know that non-adherence is routinely observed in the field (recently reviewed in (43)). Other common exclusion criteria in clinical trials are patients who are severely ill or on co-medication both of which may substantially alter their PK. Co-medication is important in many endemic countries where long-term medication for TB and HIV is relatively common; changes in PK cause by co-medication can then be fed back into PK/PD model to see if and how the drug dosages given to these sub-groups of patients should be altered. We do not argue that PK/PD modelling will necessarily give an exact dosing schedule for such people but we do argue it can provide a dosing scheduling starting point for clinical trials in such patients.

- Current trials cannot easily be used to predict future events, in particular the consequences of increasing drug resistance, which many commentators regard as inevitable (e.g. (44)). Once again PK/PD modelling can investigate how robust regimens may be to small increases in levels of parasite drug tolerance/resistance and hence their likely therapeutic lifespan (28). The PK/PD data can also be used to try and implement regimens that minimise the selection pressures for resistance.

Conclusions
The realisation that d7c is apparently poorly predictive of therapeutic outcome raises the obvious question of whether this PK component of clinical trials could be improved. If measured, our first conclusion would be that the predictive capability of d7c be evaluated and reported using ROC analysis. The collection of d7c data is not trivial. It requires correct blood sample collection, processing, storage and transportation to a central reference laboratory, as well as collation of d7c with clinical outcome and possible stratification in the subsequent analysis. As argued above, the outcome of all this effort is likely to be the trivial conclusion that giving patients more drug improves their chance of treatment success. It is therefore arguable whether this approach represents the best use of resources and that better resource allocations might be achieved through a two-tier clinical trials framework using two distinct types of clinical trials to guide policy.

First, routine efficacy monitoring trials designed to check that local first line drug(s) have remained effective. There seems little point in measuring d7c in such studies given effectiveness is likely to be high so that few, if any, failures will occur. In light of the poor predictive ability of d7c, and the difficulties inherent in pool resulting from different studies (discussed in the Supplementary Information), we suggest that WHO drops their recommendation that d7c should be collected in such studies and leave individual investigators to consider whether the costs of measuring d7c can be justified. One such justification would be to compare whether different populations vary in their drug absorption and/or subsequent metabolism. In this case, rather than relating the d7c to treatment outcome, investigators would examine whether d7c differs in ‘at risk’ populations. Examples of ‘at risk’ populations
could be young or malnourished children, or patients with co-medication (e.g. antiretrovirals). However, these ‘at risk’ populations are generally excluded from clinical trials so may be better investigated as part of Phase IV trials. One other justification would be as a crude measure of factors affecting PK: if the same population exhibits differing d7c over time then it suggests, for example, adherence to the drug regimen was altered as it is unlikely that human PK parameters will have changed.

Second, we recommend the periodic use of more focussed and specialised PK/PD trials, including a sophisticated PK component designed to generate the data required for long-term optimisation of regimens and enrolling the full range of patient types. This PK component would necessitate an optimised, drug specific, sampling schedule, which may or may not include measurement on day 7 (see for example (45)). Furthermore, these trials should also be accompanied by culturing of local parasite isolates to estimate PD parameters, thereby allowing investigation of changing circumstances, such as the possible evolution of drug resistance. However, conducting focussed PK/PD trials will be challenging and likely to need external support. Analysing one drug concentration sample point costs typically 20 USD or more. This does not include the cost of staff, consumables, sample storage, sample transport, acquisition or maintenance of analytical equipment, etc. The culturing component is also technically and logistically demanding as the blood is sometimes collected in very remote areas and needs to reach the lab early enough that parasites are still viable (40).
We also require an ethical framework for performing the PK/PD trials. There are limits on how much blood can be drawn from any individual patient, particularly infants. Blood is routinely drawn at prescribed days of follow-up to check for parasitaemia but it is likely that at least some of the sampling (for example for parasite culturing for IC50 analysis, repeat sampling for artemisinin PK) may be of no direct clinical benefit to the patient. This is not ethically impossible but it does require some justification that individual patients will accrue future benefits from effective antimalarial drug provision that outweighs the inconveniences of providing additional blood sample not required for immediate clinical purposes. In summary, we need a consensus protocols for these strategic trails and a consensus that they can be deemed ethical.

Statisticians often recommend that researchers planning a study should first simulate, and then analyse, realistic dataset so that their eventual study design can avoid any likely pitfalls revealed in the simulations. For once we have heeded our own advice. The collection of PK data (d7c) was recently discussed in the WHO publication (“Global report on antimalarial drug efficacy and drug resistance: 2000-2010”) who noted that “The interpretation of the results of blood concentration studies for determining drug resistance is not, however, always straightforward”. We think this is an understatement and that too little preparatory work has gone into identifying best ways of analysing the data. Part of the problem is that, of course, it is impossible to analyse the data until has been generated and methods of analysis typically have to be refined according to the data. In summary, we believe that simply collecting d7c is unlikely to be best use of PK resources, and that more sophisticated PK elements of clinical
trials be designed and rolled-out. The resulting detailed PK data on mean values and
distributions can then usefully contribute to the rationale design of robust and effective
antimalarial drug regimens.

Funding

This work was supported by the Bill and Melinda Gates Foundation [grant number 37999.01]
and the Medical Research Council [grant number G1100522].


Figure 1. Graphical representation of a ROC curve.

Receiver-operator characteristic (ROC) curves are widely used in medical research to quantify the predictive value of a measurement as explained in the main text. The black diagonal line is a ROC curve for a diagnostic with no predictive value, the blue line is an illustrative example of a good predictor, the red line is a ROC curve for day 7 drug concentrations obtained previously by simulation (18), and the black dashed line for point 4 illustrates how we estimate the area under the ROC curve from point estimates of sensitivity and specificity (see Supplementary Information). These point estimates correspond to the following sources: point (1) is White et al. (46), point (2) is Denis et al. (47), point (3) is Price et al. (48), point (4) is Ezzet et al. (31), point (5) is Checci et al. (49), point (6) is McGready et al. (50).

Figure 2. A simulation of how day 7 concentration (d7c) varies according to the level of patient adherence to the regimen.

The Figure shows the mean d7c of piperaquine (PPQ) in mg/L simulated for 10,000 African patients given dihydroartemisinin (DHA)-PPQ once daily for three days using the PK/PD model described in (18, 28). The patients were either dosed with the exact mg/kg dosage (green; 18mg/kg as per (51)), dosed according to their age (red; as per (52)) or dosed according to their weight (blue; as per (53)). To investigate patient adherence it was assumed that patients took the full course i.e. all three doses (left), only the first two doses (middle) or the first dose only (right). Outliers are not displayed. The Figure was produced using a published one-compartment model. A more realistic two-compartment model parameterisation based on
Staehli Hodel et al. (54) gave identical results except that the d7c in the central compartment is approximately 4-fold lower (see Figure S1 in Supplementary Information).

[Figure 2 parameter values were as follows: Elimination rate per day is 0.03; volume of distribution is 150 L/kg]