Patient adherence to prescribed antimicrobial drug dosing regimens

Bernard Vrijens1,2* and John Urquhart1,3,4

1AARDEX Ltd, Zug, Switzerland; 2Department of Biostatistics and Medical Informatics, University of Liège, Liège, Belgium, 3Pharmaco-epidemiology Group, Maastricht University, Maastricht, The Netherlands; 4Department of Biopharmaceutical Sciences, University of California San Francisco Medical Center, San Francisco, CA, USA

The aim of this article is to review current knowledge about the clinical impact of patients' variable adherence to prescribed anti-infective drug dosing regimens, with the aim of renewing interest and exploration of this important but largely neglected area of therapeutics. Central to the estimation of a patient's adherence to a prescribed drug regimen is a reliably compiled drug dosing history. Electronic monitoring methods have emerged as the virtual 'gold standard' for compiling drug dosing histories in ambulatory patients. Reliably compiled drug dosing histories are consistently downwardly skewed, with varying degrees of under-dosing. In particular, the consideration of time intervals between protease inhibitor doses has revealed that ambulatory patients' variable execution of prescribed dosing regimens is a leading source of variance in viral response. Such analyses reveal the need for a new discipline, called phamionics, which is the study of how ambulatory patients use prescription drugs. Properly analysed, reliable data on the time-course of patients' actual intake of prescription drugs can eliminate a major source of unallocated variance in drug responses, including the non-response that occurs and is easily misinterpreted when a patient's complete non-execution of a prescribed drug regimen is unrecognized clinically. As such, reliable compilation of ambulatory patients' drug dosing histories has the promise of being a key step in reducing unallocated variance in drug response and in improving the informational yield of clinical trials. It is also the basis for sound, measurement-guided steps taken to improve a patient's execution of a prescribed dosing regimen.

Keywords: compliance, directly observed therapy, drug delivery systems, viral load, antiretroviral drugs, pharmacokinetics, pharmacodynamics, pharmacometrics, phamionics, tuberculosis, human immunodeficiency virus (HIV), assay sensitivity

1. Introduction

Suboptimal use of rationally prescribed anti-infective drugs in ambulatory care appears to be a worldwide phenomenon. It is strongly skewed toward underdosing, created by delayed and omitted doses, sometimes resulting in multiple, sequential omisions of prescribed doses, creating a variety of drug-specific problems. In the case of antimicrobial chemotherapeutic agents, dose omissions, if long enough, allow the concentrations of drugs in plasma to fall to levels too low to inhibit microorganism replication, but still high enough to exert selection pressure. Thus, the combination of inadequate drug action and selection pressure for drug resistance can occur and recur. The consequence is a triad of problems: (i) poor outcomes of individual patients' treatments; (ii) emergence of drug-resistant microorganisms; (iii) depletion of the armamentarium of anti-infective (AI) drugs of the class(es) in question.

These considerations motivated the adoption of directly observed therapy (DOT) in the treatment of tuberculosis (TB), which has achieved a widely recognized superiority over conventional self-administration by ambulatory patients who have infrequent contact with caregivers. Whereas DOT is necessarily labour intensive and thus costly, its benefits in both better treatment outcomes and minimization of emergent multidrug resistance clearly justify the added cost and efforts of DOT. The DOT regimen has been shortened (DOTs/DOTs-plus), but the principle remains of virtually assured exposure to the drugs in question, through direct observation by a caregiver of the patient's administration of adequate doses. Naturally, one of the attributes of the DOT approach is that the proficiency with which direct observation is managed becomes an integral part of the evaluation of treatment effectiveness.

As a means for assuring continuity of a patient's exposure to the prescribed drug, DOT is a 'brute-force' approach, mid-way...
between conventional ambulatory care and in-hospital care. In the latter case, medicines are administered by health professionals, whose error-rates, although not zero, are generally an order of magnitude or more lower than the error-rates in conventional ambulatory care, in which the patient—with or without the help of a relative, neighbour, friend or some other person—is responsible for timely re-medication.

DOT of TB has, however, a pharmacometric feature that is sometimes overlooked: the drugs used thus far in DOT appear to possess a surprising degree of flexibility, allowing change from their original two-to-three times daily administration of relatively small doses to four or fewer times per week administration of much larger doses. If the practical upper limit on the frequency of caregiver visits needed for effective DOT is more or less four times a week, it is unrealistic—given current knowledge—to assume that intervals of 48–72 h between doses could be achieved safely and effectively with other drugs presently used in ambulatory anti-infective treatment, particularly in HIV treatment. This matter is not peculiar to anti-infective drugs but cuts across all pharmacotherapy: for seven-eighths of the drugs in the present pharmacopoeia have plasma half-lives of < 12 h,2 the pharmacokinetic interpretation of which would fix their optimal dosing frequency at once, twice or three times daily, i.e. intervals between doses, respectively, of 24, 12 or 8 h.

It is possible to escape these limits if either of two extraordinary conditions are met. The first of these is a delivery systems-based reformulation capable of extending the duration of drug release in the gastrointestinal tract and absorption into the bloodstream, so that the interval between doses reaches the DOT range.3 The second occurs with drugs that have unusual pharmacodynamics, such that drug action continues at therapeutic levels for 48–72 or more hours after a last-taken dose, i.e. long enough for DOT. It is clear in other areas of pharmacology that some drugs, e.g. reserpine, omeprazole, most of the β-adrenergic receptor antagonists, act by mechanisms that do not depend on the continuing presence of the drug in plasma.3,4 In the early days of antibiotic research, Eagle described a ‘post-antibiotic effect’ of penicillin, by which he meant that there was evidence for persistence of the cidal effect of penicillin well past the time, after a last-administered dose, when penicillin ceased being detectable in plasma. Clearly the validity of that idea depends on, for example, the sensitivity of available assays for measuring concentrations—in plasma or other body fluids—of the antibiotic and of any active metabolites it may have. There is sufficient ambiguity about these factors—that the idea of an AI agent that continues to work when it is no longer present, i.e. Eagle’s ‘post-antibiotic effect’—has little standing in the infectious disease field.

In other areas of pharmacology, however, there is clear evidence that some drugs have the counterpart property to the ‘post-antibiotic effect’. It is well-illustrated by omeprazole, which has the distinction of being the largest selling drug in the history of the pharmaceutical industry. Omeprazole has a plasma half-life of about an hour, but it acts by irreversibly inactivating the Na+/H+ ATPase enzyme that is responsible for the last step in gastric acid secretion. With this enzyme destroyed, acid production ceases, and does not resume until a sufficient quantity of new enzyme has been synthesized, a process that takes several days. Thus the drug’s acid-blocking action disappears, not because the drug disappears, but because new enzyme appears. The process of enzyme synthesis occurs with its own dynamics, uninfluenced by the drug. The resulting pharmacodynamics of omeprazole can be qualitatively characterized as ‘fast onset/slow offset’, in that the drug’s blockade of acid secretion is essentially fully manifest within an hour or so of dose administration, but then continues for 2–3 days without re-medication.5 Such dynamics, which some characterize as a ‘hit-and-run’ action, indicate that the recommended regimen of once-daily dosing provides a great deal of forgiveness for the omissions of scheduled doses that occur so frequently when patients are solely responsible for their own medication.

Are there counterparts in the world of anti-infective drugs to the ‘hit-and-run’ action that is illustrated by omeprazole? If so, then the prevailing focus in AI drug development on the concentrations of AI drugs in plasma, plasma half-lives, onset times, times to maximum effect and the like, will miss drugs with unusually long post-dose durations of AI action. Direct experimentation is required to measure a drug’s post-dose duration of therapeutically effective action, but such experimentation is not part of the standard pharmacometric workup of a new agent, often leaving this therapeutically important property undefined. The particulars of omeprazole’s actions are, of course, unique to it, but the general point is that the time-course of a drug’s actions after a last-taken dose are not predictable from the time-course of its onset of actions, or its pharmacokinetics, and are only revealed by experimental studies.

Such studies are not academic, as their results can have important consequences for the selection of a recommended dosing regimen that is near enough to optimal to preclude the kinds of post-marketing changes that, for example, resulted in a quarantining of the recommended dose of itraconazole, and a consequent quarantining of its commercial potential.5 The post-marketing halving of the daily dose of zidovudine provides another example. A recent study of the compliance-dependent outcomes of doxycycline treatment of chlamydial infections, which found good outcomes in the face of only partial compliance with the prescribed regimen, suggests that the usually prescribed regimen may call for more drug to be administered than necessary.7

Identifying a dosing regimen for recommended use is one of the more difficult tasks in the development of new pharmaceutical products. The history of post-marketing dose reductions, captured in two recent studies,5,8 indicates the need for improvement in this process, which, when not conducted properly, can have major therapeutic and economic consequences. It is clear that the usual forces playing on the process of dose selection in drug development run most strongly towards overestimation; it is also clear that the usual patterns of ambulatory patients’ dosing—with drugs lacking in abuse-potential—run strongly towards underdosing.9–11 Thus, in the clinical development of ambulatory-use anti-infective agents, careful observation of the clinical correlates of reliably compiled dosing histories stands to be a useful tool in helping to identify regimens most suitable for recommendation in product labelling. In effect, the usual patterns of ambulatory patients’ variable compliance with prescribed drug regimens constitute ‘natural experiments’ in variable underdosing that, when captured by reliable, aptly timed measurements, and used to generate testable hypotheses, can improve both the efficiency and the effectiveness of drug development. This idea is not new, as will be evident from the re-reading of a paper written by one of us in this journal a decade ago,12 but its adoption remains outside the status quo in
clinical research. Meanwhile, however, the incidence of post-marketing dose reductions has escalated steadily since 1990.6

In the Introduction, we expose the topic of non-adherence to prescribed AI drugs and explain how DOT, while difficult to implement in clinical practice, could solve the problem of irregular drug intake. The following section introduces quantitative definitions to describe variable drug exposure over time and section 3 develops two practical alternatives to DOT. Section 4 discusses key findings with electronic monitoring as an alternative to DOT. Section 5 describes the historical barriers to, and neglected opportunities in, the routine collection and analysis of ambulatory patients’ dosing histories. Section 6 discusses the clinical consequences of variable drug exposure through a pharmacokinetic/pharmacodynamic approach and poses the drug-specific question of how much compliance is enough. Finally, section 7 lists the current challenges and future directions, while section 8 considers implications of the foregoing.

2. Some definitions

It is useful to choose parameters to characterize ambulatory patients’ dosing histories that support quantitative analysis of their impact, through recognized pharmacometric processes (pharmacokinetics and pharmacodynamics) on clinical and economic outcomes of ambulatory pharmacotherapy. Many use the term ‘adherence’, in the sense of ‘adherence to the prescribed regimen’, and define it as ‘the patient’s tendency to follow medical advice’. Such a definition has metaphorical value, but does not support pharmacometric analysis. It is therefore useful to consider ‘adherence’ as a kind of blanket, qualitative term, but to recognize that quantitative analysis is only served by defining two substituent terms: (i) compliance and (ii) persistence. The former is defined as ‘the degree of correspondence between the patient’s actual dosing history and the prescribed dosing regimen’. The latter is defined as ‘the time elapsed between the first dose taken and time of treatment discontinuation’.

Drug regimen execution is a dynamic process, fluctuating over time, that cannot be adequately characterized by a single number, e.g. the percentage of prescribed doses taken during an interval of observation, or the percentage of treatment days within which the patient took the prescribed number of doses, or average intervals between doses. It is worse to concoct ad hoc threshold values in order to dichotomize patients between ‘adherent’ or ‘non-adherent’, ‘good’ or ‘poor’ compliers, etc. A leading reason why dichotomization is unsatisfactory is the wide range of underlying drug-, disease-, and formulation-specific pharmacodynamics that determine crucial thresholds between ‘enough’ and ‘not enough’ exposure to drug to achieve satisfactory therapeutic results. Many, for example, ritualistically invoke an often-used, but never pharmacometrically justified criterion that taking 80% or more of prescribed doses qualifies the patient to be judged as adequately compliant. Paterson et al.13 reported that virological failure occurred in half of HIV-infected patients taking 80%–90% of prescribed doses of antiretroviral drugs—clearly refuting the validity of assuming that taking 80% of prescribed doses is ‘enough’ in the circumstances and with the drugs used in that study. By the same token, we were we to apply the ‘80% is OK’ criterion to the once-daily, low-dose, combined oestrogen/progestogen oral contraceptives, the resulting rates of conception would be so high as to prompt the conclusion that these widely used products were ineffective. (The low-dose, combined oral steroidal contraceptives command special attention as we have more use-experience with them than with any other pharmaceutical intended for long-term use.)

3. Two alternatives to conventional patient-administered doses or DOT

Drug delivery systems

Whereas there have been a number of applications of the delivery systems principle, there have been, for various reasons, many failures to develop delivery systems that can achieve even a satisfactory once-daily dosing regimen, let alone one of lower dosing frequency, e.g. the 3–4 times weekly dosing regimen needed for DOT. One limitation arises when the drug in question has a low molar potency, requiring doses of drug of several grams or more per day, and of course correspondingly larger amounts for 2–3 day intervals between doses. Formulation to achieve controlled-release requires use of various excipients that add to the bulk of the finished product, which has to be evaluated against the empirical fact that solid oral dosage forms which exceed a volume of ca. 0.75 mL meet with increasing resistance from patients who find them difficult or impossible to swallow. Typically, a controlled-release dosage form of 0.75 mL overall volume will contain one-third or less its volume as active drug, i.e. 250 or fewer milligrams, assuming that the drug has a specific gravity of 1.0, which, with a few exceptions, is a valid approximation. If, for example, the required daily dose of drug was 1 g, it would necessitate four such dosage forms, and, if formulated for DOT, with a 3 day interval between doses, it would necessitate the administration of a dozen or more such dosage forms—an approach likely to create considerable resistance to wide adoption.

These and other constraints on controlled-release mechanisms are discussed by Urquhart.5 In order for the delivery systems approach to be widely applicable, there is clearly a need to find drugs with molar potencies high enough for them to be effective at daily doses of <50 mg, and preferably <10 mg.

Size of the dosage form is not the only problem. The absorption of drugs released from orally administered, solid dosage forms can, under ideal conditions, continue for as long as 18–20 h, but not longer because the normal patterns of gastrointestinal motility and transit of tablet-sized objects carry the dosage form past the point in the lower colon of reliable absorption. Some drugs, moreover, have an ‘absorption window’ that allows absorption only in, for example, the upper small intestine, in which case once the dosage form has passed beyond the absorption window, subsequently released drug cannot be absorbed. Recent work on gastro-retentive dosage forms,2 gives promise of products that can escape this limitation by remaining within the stomach for exceptionally long periods of time, during which released, solubilized drug can continue to exit the stomach and be absorbed within the window.

These volume considerations apply also to long-acting (e.g. 1–3 months) ‘depot’ injectables. Their history dates back to the monthly depot form of injectable penicillin in the late 1940s. Here again, the bulk of the agent being administered is an important matter, for pain at the injection site is more or less directly related to the amount of drug plus excipients that must be injected. Obviously, the choice of intervals between
successive injections is a compromise between the desire for as long an interval as possible, for the sake of cost and convenience, and the desire to minimize pain at the injection site. For long-acting implants (e.g. 1–5 years) drugs with effective parenteral doses are needed in the range of single- or double-digit micrograms per day.

To study the impact of long-acting drug delivery on effectiveness it is useful to revisit the management aspects of the classic trial of depot penicillin and two oral regimens in the prevention of streptococcal infections and recurrent acute rheumatic fever.14 This demonstrated strikingly higher effectiveness in the patients randomized to receive the monthly depot injection, compared with those who were randomized to receive either of the two oral regimens. Even when the latter two groups were stratified by interview-estimated deviations from strict compliance with the prescribed daily dosing regimens, the nominally fully compliant patients (who numbered about half of each of the two groups randomized to receive oral drug) still fell well short of the effectiveness of the monthly depot injection, which was professionally administered, and scarcely ever missed. The two most important messages from this seminal study are: (i) assured continuity of drug exposure can provide a degree of effectiveness not achievable with regimens of patient-administered daily dosing; (ii) in order to understand the shortfalls in effectiveness of oral regimens, reliable, objective data are needed on the time-intervals between successive doses over the full course of treatment. Only with such data can one see when there were intervals between doses that were too long to maintain effective antimicrobial drug action, and the temporal associations between long interdose intervals and microbial status with respect to waxing or waning of the microbial population and disease severity, and emergence of drug-resistant microorganisms.

As discussed below, the striking effectiveness of the depot form of penicillin in preventing recurrent streptococcal infections and acute rheumatic fever has subsequently been echoed in the contraceptive field, where quarterly depot injections and a 5 year subcutaneous implant form of progesterational steroids respectively provide the next-highest and highest levels of contraceptive effectiveness of any of the steroidal contraceptives.15 Remarkably, the same steroid used in the 5 year implant, formulated as a once-daily oral dosage form, has the poorest contraceptive effectiveness of any of the steroidal contraceptives.16 These findings underscore the message that removing the vagaries of ambulatory patients’ execution of prescribed dosage regimens removes a major impediment to realization of full therapeutic effectiveness.

With anti-infective agents, the prospects for long-acting implants to assure continuity of antimicrobial chemotherapy hang only partly on further advances in drug delivery technologies, because there is a clear need, which cannot be overcome by delivery system technologies, for ultra-high molar potency antimicrobial drugs that also have the stability characteristics to carry them, essentially intact, through many weeks, months or years of residence in the warm aqueous environment of subcutaneous tissue. A main incentive for finding agents of ultra-high molar potency antimicrobial drugs is the prospect of unparalleled effectiveness of implant or depot products. It is a grand challenge to medicinal chemists to seek and find such agents, without which foreseeable delivery systems-based products are mostly limited dosing frequencies in the range between once weekly and once daily. Until medicinal chemistry can produce anti-infective drugs active at the micrograms/day dosing level, it will be necessary to find other ways to come as close as possible to assured exposure to the available AI drugs.

**Electronically monitored drug therapy**

On the therapeutics side, the logical complement to caregiver-based DOT is electronic monitoring of therapy, which has a quarter-century of use-experience in many therapeutic applications at many geographical sites. The experience is captured in published work that today covers over 650 publications, including about 250 peer-reviewed papers.17 The principle involved is called ‘electronic medication event monitoring’ (eMEM), realized by incorporating microcircuitry into pharmaceutical packages of various designs, such that the manoeuvres needed to remove a dose of drug are detected, time-stamped, analysed, stored and communicated to the appropriate caregiver(s), to initiate apt corrective action.9,18

The first eMEM package19 was built and tested in 1977; it could capture medication events hourly for only 6 weeks, and cost ca. US$1200 (in current dollars)—$1.20 per recordable medication event. Since then, better design, size- and cost-reduced microcircuitry, and increased manufacturing volume have driven the per-package cost to ca. $80, whereas the functional lifetime has risen to 3 years. The cost per recordable medication event is down to 3.5¢. These changes—typical of electronic products—have occurred even though uses of eMEM are still largely limited to clinical trials. Thus, costs have yet to reflect the economies of largest-volume manufacturing. Recent trials utilizing eMEM have, however, enrolled >2000 patients,20,21 indicating both a growing and broadening use of eMEM methods in clinical investigation, the results of which naturally create the basis in evidence for their use in specific therapeutic areas in routine medical care.

Other methods have been used to ascertain the patient’s exposure to drug, but, as described in Table 1, electronic monitoring is the only method that gives reliable, objective data on the intervals between successively taken doses over long periods of time.9–11,18,22–25

Electronic monitoring of patients’ dosing histories has repeatedly revealed that intakes of patients are frequently irregular, spanning a wide spectrum of deviations from the prescribed regimen, certainly not supporting a dichotomous classification of patients as ‘good’ and ‘poor’ adherers. Under-dosing, relative to the prescribed regimen, is a much more frequent phenomenon than over-dosing. Moreover, under-dosing tends to become both more prevalent and more severe over time in drug regimens meant to be long-term or lifelong.26–28

<table>
<thead>
<tr>
<th>Table 1. Classification of the different methods used to assess patient adherence to prescribed therapy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Objective</td>
</tr>
<tr>
<td>Potentially biased</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
4. Key findings with electronic monitoring

A major surprise has been the finding that life-threatening diseases do not, ipso facto, enforce strict execution of prescribed regimen(s). This fact became evident in the fields of organ transplantation,29–33 cancer chemotherapy34 and more recently in the HIV–AIDS field, where the observational study of Vanhove et al.35 in 1996 first demonstrated with eMEM the occurrence of drug holidays, and their temporal relation to increases in viral replication and emergence of drug-resistant HIV.

The study by Vanhove et al.35 provided data on the temporal sequences of dosing of protease inhibitors, viral loads and changes in drug resistance in six HIV-infected patients. As such, it was the first study to demonstrate the occurrence of multi-day lapses in dosing—drug holidays—in HIV-infected patients, and the temporal sequence of holiday, followed by increased viral replication, followed by increased viral load and genetic changes leading to the emergence of drug-resistant HIV. This seminal report had an interesting pre-publication history, as it was rejected by several leading journals, mainly on the grounds that it was ‘only observational’, not a randomized, controlled study, and involving only six patients, although each showed the same temporal sequence of events in the wake of a drug holiday. In the end, the study was published as a letter in JAMA in December 1996. This misadventure recalls a comment by the late Louis Lasagna—that the pendulum in clinical study designs has swung so far toward randomized, controlled studies as to have come right off its hook. He made this comment in the late 1990s when discussing Austin Bradford Hill’s teachings about the inference of causality.36 Hill’s lucid insights, along with the more recent paper by the late Lewis Sheiner37 on learning and confirming in clinical drug development, constitute the framework for integrating observational and controlled studies for efficient drug development. A central element and largely still untapped resource in this integration is the prevailing range of erratic compliance with protocol-specified drug regimens in both observational and controlled clinical studies involving ambulatory patients.38–45 As such, it is the leading course of variance in ambulatory patients’ responses to prescribed drug.

Hill and Sheiner both provide antidotes to the misconceptions, shared unfortunately by many, including the reviewers who rejected publication of the findings of Vanhove et al.,35 which are now widely recognized as a leading problem in the management of HIV-infected patients, and, indeed, in other major infectious diseases as well, including tuberculosis and malaria. As there is no evident ethical way to impose drug holidays in a randomized, controlled trial in these serious infections, how else might we learn about the consequences of intermittently interrupted exposure to otherwise effective AI drugs than by the kind of observations—on patients’ spontaneously occurring lapses in dosing—that Vanhove et al.35 reported? Their work reflects the heuristic value of the ‘natural experiment’ in variable dosing, revealing the temporal sequence of clinical events before, during, and after holidays with the antiretroviral drugs involved. But the ‘natural experiment’ is only discernible when drug holidays are reliably identified, which is only possible with electronic compilation of ambulatory patients’ drug dosing histories. Such information, obviously, is basic for rational approaches to patient management and to the design of delivery systems and techniques of patient management that are capable of minimizing both the incidence and consequences of drug holidays that occur when the techniques of patient management are inadequate.

5. Historical barriers and neglected opportunities

Variability in dosing due to patients’ faulty execution of rationally prescribed drug regimens, or to early discontinuation (short persistence) with prescribed drug regimens, are widely acknowledged, major sources of variation in drug response in both clinical trials and medical practice,40,41,46–49 and a leading threat to the crucial parameter of assay sensitivity in clinical trials.50,51 However, average treatment outcome, exemplified by the intent-to-treat (ITT) policy, is typically the main focus of analysis even outside the context of confirmatory clinical trials. As the ITT policy censors post-randomization events from consideration in analysis, only a few trials have included a serious effort to quantify the impact of patients’ actual dosing patterns on treatment effectiveness.52,53

Lack of reliable measurement of drug exposure has been one of the main obstacles to such analyses. Growing uses of electronic monitoring methods have begun to remove that obstacle, leaving a variety of concerns—of which a leading one is the concern that, since the discredited count of returned dosage forms regularly gives an overly optimistic estimate of patient adherence, ‘the FDA will think we did a poorly managed trial if they see that many patients failed to adhere with the prescribed regimen’. Besides being a key part of gridlock against methodological advancement, this attitude fails to distinguish between deviations from protocol on the part of the trial staff, on the one hand, and variable execution of protocol-specified drug regimens by patients, on the other. Trial staffs are charged with the professional responsibility of executing the protocol correctly—a responsibility that demands strict fulfillment. Trial subjects, in contrast, are under no such obligation, for they are the subjects of an experiment, and what they do vis-à-vis the execution of protocol-specified drug regimen(s) is an experimental finding. If there is an unfulfilled responsibility in the present conduct of drug trials, it is the perversely struthian attitude among trialists on the realities of ambulatory patients’ actual execution of protocol-specified drug regimens and their perseverance with the long-ago discredited35 returned ‘pill count’ method of guessing estimating patients’ execution of protocol-specified drug dosing regimens.

There is an urgent need for studies that examine—using both direct measurements and modelling and simulation methods—the temporal patterns of drug response that are engendered by patients’ actual dosing patterns. Not only do the magnitudes of drug responses vary widely in a usually downwardly skewed manner, as expected from prior information on the test agent’s dose–response relations, but some holiday-triggered responses differ qualitatively from the dose-dependent drug responses of the majority of patients. In the infectious disease arena, the leading qualitatively different response to antimicrobial chemotherapy, from the dose-dependent drug responses of the majority of treated patients, is the emergence of drug resistance and ‘escape’ from the antimicrobial actions of the drug in question. The holiday scenario, which is largely confined to a holiday-prone one-third of treated ambulatory patients, may be the principal trigger for the proliferation of drug-resistant microorganisms. Routine application of confirmatory (ITT) analysis estimates the risk of this phenomenon as if it were equally likely
in all treated patients, so the risk of emergent resistance is overestimated in patients not at risk, and underestimated in the holiday-prone, who are putatively the sole ones at risk.

The wide range of temporal patterns of drug exposure that occur among ambulatory patients is obscured by continued use of the unreliable methods that afford patients the easy ability to censor evidence of omitted doses, e.g. returned tablet counts, interviews, questionnaires, diaries or methods that make unrealistic demands on patients' recall of past events. One can add to this list the seemingly objective method of measuring drug concentration in the patient's plasma, which is unfortunately biased upward by 'white-coat compliance'—the short-term improvement in compliance that precedes a scheduled visit by 24–72 h. As the vast majority of drugs have plasma half-lives of 12 h or less,\(^2\) it only takes—as pharmacokinetic analysis shows—2–3 doses within the day or two prior to the drawing of blood to put the drug's concentration into the therapeutic range, while underdosing that may have occurred previously is unrevealed, because a measured concentration of drug in plasma can reveal only the pharmacokinetic effects of dosing within a period of time prior to blood-sampling equal to 3–4 times the plasma half-life of the drug in question. For most drugs, this means that so-called 'therapeutic drug monitoring (TDM)' shows the consequences of doses taken only within the 36–48 h that precede the sampling of blood, which is usually a scheduled event, and thus a trigger of white-coat compliance. One could, in theory, do unscheduled sampling of blood, but it is both expensive and intrusive, and, in the end, reflective only of the patient's drug intake during a 1–2 day period of time prior to the sample.

As for diaries, a recent study, using an electronically monitored diary, which recorded the time of each entry into the diary, showed that only 11% of entries occurred within an interval since the event being recorded that was considered plausibly short.\(^55\) Returned tablet counts are easily modified to create the illusion of good adherence by simply discarding or hoarding the remaining tablets, and then returning an empty package, or one that, for credibility, contains one or a few untaken dosage forms. Interviews and histories are vexed by faulty recall and a certain pervasive desire to present oneself as having followed the doctor's orders. The unreliability of pre-electronic means for assessing drug exposure in ambulatory patients has an extensive bibliography, the most searching of which are refs 22–25, in which the reliability of electronic monitoring has been carefully assessed and judged the ‘backbone’ of the assessment of drug exposure in ambulatory care, to use the term of Liu et al.\(^22\) The unique value of electronic monitoring has been reinforced by recent demonstrations that the timing of doses and the resulting intervals between doses have the greatest clinical explanatory power of all expressions of drug exposure in ambulatory patients.\(^42,43\)

The aim in quantifying drug exposure should be to establish a basis of reliable information for sound, ongoing management of the patient's self-administration of prescribed drugs. A single finding of a very low or zero concentration of an AI drug in plasma—the fruits of a one-off trap sprung on the patient—is hardly a basis for sound management of the patient's medications. An ongoing chronological plot of the 24 h clock times of daily dosing times, in contrast, is the logical basis for medication management, for it shows when errors occur, the consistency of daily dosing times, weekend versus weekday dosing patterns, and other aspects that define what needs to be done to drive the patient's dosing history into conformity with the prescribed dosing regimen. Many examples of chronological plots of daily dosing times can be found in Metry & Meyer.\(^55\)

### 6. Clinical consequences of variable drug exposure

**Impact on measured concentrations (PK)**

By combining electronically recorded dosing histories with individual pharmacokinetic parameters of a single patient, one can project individual drug concentrations over long periods of time. This process is illustrated in Figure 1, which shows the model-based projection of a single patient's drug concentrations over 365 days. Drug holidays, which occur in many patients, are seen in Figure 1 and lead to drug concentrations falling below the EC\(_{50}\). From this type of picture one logically questions the relevance of monthly therapeutic drug monitoring when dosing histories previous to the sampling are unknown. In a recent study, Vrijens et al.\(^58\) have shown that eMEM-compiled dosing histories can explain up to 55% of the residual, within-patient variability arising when the electronically compiled dosing data were ignored and the samples were assumed to have been drawn at steady state.

These results suggest that switching from patient-reported to electronically-compiled dosing histories may greatly improve the quality of information derived from both population PK studies and therapeutic drug monitoring. Objective measurements of dosing times and careful attention to accurate recording of sampling times can purge population PK studies of much of their residual variability, thus improving their quality and utility.

**Impact on viral suppression (PD)**

The explanatory power of measured antiretroviral drug exposure on virological response in HIV infections is of great interest. Recently several studies showed the impact of aggregate intake of protease inhibitors.\(^13,59,60\) Those preliminary results suggested further research to characterize better how variation in drug exposure patterns, created by the patient's variable adherence to prescribed dosing regimens of antiretroviral drugs, determine success or failure of treatment. To establish a robust dynamic

---

**Figure 1.** eMEM-PK model-based projection of the time-course of ARV drug concentration in plasma during a 1 year period of electronically compiled dosing history in a patient infected with HIV and treated with a once-daily ARV drug.
relation between adherence to prescribed therapy and virological response, measured repeatedly over time, two major improvements need to be achieved.

First it is important not only to study the virological success rate (achievement of viral load below the lower limit of detection, currently < 50 copies/mL), but also to investigate the rate at which patients move downwards or upwards through defined categories of viral load, as they are, respectively, properly or improperly treated. The main challenge lies in modelling dependencies over time and using information contained in the data efficiently to establish a dynamic relation between drug exposure and viral load. Separate characterization of the onset, when drug therapy starts, and offset, when therapy stops, will be of prime importance. This strategy will be especially useful to assess the forgiveness of the drug for lapses in dosing.

Secondly, one will need good summaries of dosing histories—the basic expression of drug exposure—as input to the model. Errors in the timing of doses are incompletely characterized by the sometimes-used parameter ‘percentage of treatment days during which the correct number of doses were taken’. Substantial errors in dose timing occur in some patients who take all, or almost all, of the prescribed number of doses. Vrijens et al. recently showed that dose-timing information increases the explanatory power of patient adherence data and its association with antiretroviral (ARV) treatment outcomes. In particular, these analyses showed the particularly high impact of multi-day intervals between doses, i.e. drug holidays. Avoidance of these long inter-dose intervals should be a priority in efforts to maintain viral suppression. Explanatory power of dose-timing data varies from one drug and treatment situation to another, with some ARVs drugs being much less forgiving than others. Comparisons among available drugs can include their relative degrees of forgiveness for serially omitted doses. Implementation of this sensible approach begins with comparative data on competing agents’ duration of therapeutically effective antiretroviral drug actions.

Impact on drug resistance (PD)

In building a model for viral load, one should take into account the risk over time of developing viral resistance. However, the relationship between variation in drug exposure and the development of viral resistance is incompletely understood. Whereas it is widely acknowledged that ‘poor adherence’ facilitates the emergence of drug resistance, the precise mechanisms and conditions under which resistance develops have not yet been elucidated. The study of this relationship requires both reliable dose-timing data and a complete characterization of the emergence of resistance to each drug in the regimen of study. Furthermore, while variation in internal drug exposure occurs over hours, the time scale to develop viral resistance is months or years.

Bangsberg & Deeks have reported that the risk of emergent resistance decreases as adherence decreases, starting from high but still incomplete levels of adherence, and decreasing to 40% or fewer of prescribed doses taken, where the risk of emergent resistance falls to a very low level. The reason for this superficially paradoxical result may be as follows. When dosing stops and drug levels fall, they pass through, on their way to zero, a critical zone, which is low enough to allow replication to occur, but high enough to create selection pressure. A patient who takes relatively few prescribed doses has typically discontinued treatment or has long periods of time without any intake and therefore spends most of the available treatment time with zero drug levels, and thus with only an occasional passage through the ‘critical concentration zone’ (CCZ) for selection of drug-resistant mutants, the duration of the passage being determined by the drug’s plasma half-life, with short half-life drugs having rapid passages through the CCZ, and long half-life drugs having slow passages through the CCZ. A patient who takes a relatively high percentage of prescribed doses spends most of the available treatment time with drug present in plasma, and has the opportunity for short lapses in dosing to allow drug levels to run down through the CCZ and then, with the return of dosing, to go back to replication-inhibiting levels for a time, until a future lapse in dosing creates the next dip down through the CCZ. Patients with 70%–90% of drug intake are typically the ones who show frequent drug interruptions of durations equivalent to 2–4 plasma half-lives. It is not surprising that those patients are the ones with the highest probability of developing drug resistance according to Harrigan et al. As adherence increases (>90% of drug taken), drug exposure verges on near continuity, with few or no dips down through the CCZ. The actual story will, of course, depend on the temporal patterning of dose interruptions, but one can imagine a series of frequently repeated lapses just long enough to allow repeated passages through the CCZ and a maximal risk of emergent drug-resistant microorganisms. In contrast, a longer, single lapse, equivalent in magnitude to the sum of the frequently repeated short lapses would create only one passage through the CCZ. Thus, one would predict that not only does the risk of emergent resistance increase with an increasing proportion of prescribed doses being taken, but that the risk is also dependent on the patterning of dose interruptions, with frequently recurring short(- enough) lapses in dosing creating maximal risk of emergent drug resistance. That aspect of the matter has not yet been studied, but is, in effect, a natural experiment that some patients are unwittingly running and which can be captured by electronically compiling their ARV drug dosing histories and the genetic make-up of HIV found in their blood.

To illustrate the present concept, Figures 1–3 present a patient who developed viral resistance after 250 days and lost virological control soon afterwards. The data presented in Figure 1 exemplify the potential explanatory power of having continuous data on drug exposure, together with concomitant data on viral load and fitness that are as near to continuous as feasible. For this patient, the cyclic, peak and trough pattern of drug concentration in plasma, typical of the pharmacokinetic steady state (PKss), stops once time passes the scheduled time of the first missed dose (Figure 2). The concentration of drug in plasma commences to fall, reaching a nadir that is virtually zero by the end of the 4 day holiday, just as dosing resumes. In contrast to the earlier, single trip to the CCZ, shown in Figure 2, Figure 3 shows a segment of the year-long data, with six successive holidays between days 210 and 235, and a single missed dose at day 244. Comparing the data and projections of Figures 2 and 3, the cumulative risk of emergent resistance is proportional to the time spent in the CCZ. Future research should combine eMEM dosing histories with pharmacokinetic and viral information to characterize the CCZ for each drug-regimen combination.
Patients are exposed to more drug than they need in order to give the vast majority of 1 and 2 day lapses in dosing, without generating disparate views. If one has a drug regimen that can for-tis an unresolved question, but one which will undoubtedly in turn, raises the question ‘how much forgiveness is enough?’ It can delay a dose before viral replication starts to resume. That, too, in their forgiveness, and specifically in how long a patient partial compliance with the 600 mg/day regimen. Drugs differ, 1200 mg/day regimen must differ from the clinical correlates of Obviously the clinical correlates of partial compliance with the 600 mg/day regimen must differ from the clinical correlates of partial compliance with the 600 mg/day regimen. Drugs differ, too, in their forgiveness, and specifically in how long a patient can delay a dose before viral replication starts to resume. That, in turn, raises the question ‘how much forgiveness is enough?’ It is an unresolved question, but one which will undoubtedly generate disparate views. If one has a drug regimen that can forgive the vast majority of 1 and 2 day lapses in dosing, without loss of effectiveness, then, implicitly at least, fully compliant patients are exposed to more drug than they need in order to benefit incompletely compliant patients. Of course, sooner or later the vast majority of patients commit errors of omission in their execution of prescribed drug regimens, at which time they benefit from drugs with greater rather than lesser forgiveness. Each drug and disease situation has to be looked at on its merits. From the labelling perspective, as already noted, it would seem only reasonable to include information on what to do when single or serial doses are missed.

7. Current challenges and future directions

Predictive models for viral load outcomes over time are not only useful to assess how viral load predictions can help to explain the reasons for treatment failure, but can also guide the practitioner in how best to design and focus medication management strategies. As we currently understand matters, anti-HIV medicines have to be taken for indefinitely long periods, if not for life, so this approach to medication management can bring long-term value. Compiling a patient’s dosing history from the start of treatment can allow for early detection of hazardous errors in dosing. A next logical step is to integrate dosing history data with the individual patient’s pharmacokinetics, to individualize therapy, with perhaps lower doses, less severe side effects of treatment, and stronger motivation for long persistence with treatment. There is evidence that deteriorating compliance is a predictor of impending discontinuation of treatment.62

Long-term projections of drug internal exposure can then be done, confirmed from occasional blood samples, which is a potentially huge improvement in cost-effectiveness over isolated, single samples of drug concentration in blood, the interpretation of which require reliable data, which are usually unavailable, on the patient’s dosing history during the 1–2 days prior to the blood sampling. Incorporating those pharmacokinetic projections into medical decision-making and combining them with viral load forecasts is, therefore, a now-feasible procedure, the economics of which will depend on the trade-off between costs of electronic compilation of dosing histories and savings from the huge gain in reliability of information on drug concentrations in plasma over time. In clinical practice, when the patient is at the beginning of the therapy, one could start by projecting his/her internal exposure to drug based on population pharmacokinetic parameters and observed dosing histories. Then, over the course of the therapy, serial individual samples of blood can be collected and analysed to compare with, and adjust as necessary, the simulated concentration-time values and viral forecast. Such enriched patient-specific information can help both the caregiver and the patient to monitor the patient’s exposure history, account for the present status of disease, and project the probabilities of future outcomes of treatment under different degrees of attainment of full, or wavering compliance. It is a rational basis for motivating the patient to comply better and persist longer with the rationally prescribed drug dosing regimens.

This task is sufficiently important, and has so many different aspects, that the need exists for an explicit discipline that concerns itself with ‘what the patient does with the drug’, falling in line as a third subdiscipline of biopharmaceutics. The other two are well-known: ‘pharmacokinetics’, what the patient’s body does to the drug; ‘pharmacodynamics’, what the drug does to the patient’s body. The third subdiscipline is called pharmionics.

Pharmionics is the discipline concerned with the ways in which prescription drugs ‘go’ into use—in the broadest sense of

---

**Figure 2.** Projected events during the 4 day lapse in daily administration of an ARV drug occurring during days 52–56 in the eMEM-PK model-based projection of drug concentrations shown in Figure 1. A hypothetical critical concentration zone (CCZ) is added to the plot.

**Figure 3.** Projected events during the series of drug holidays from daily administration of an ARV drug occurring during days 210–244 in the eMEM-PK model-based projection of drug concentrations shown in Figure 1. A hypothetical critical concentration zone (CCZ) is added.

**How much compliance is enough?**

Thus, clarification of temporal patterns of drug exposure that are most likely to generate resistance to ARV drugs is urgently needed. The problem begins with the fact that all recommended regimens are not necessarily optimal. In the AIDS field, we have the example of AZT, which entered the market at 1200 mg/day, and then fell by half, to its present level of 600 mg/day. Obviously the clinical correlates of partial compliance with the 1200 mg/day regimen must differ from the clinical correlates of partial compliance with the 600 mg/day regimen. Drugs differ, too, in their forgiveness, and specifically in how long a patient can delay a dose before viral replication starts to resume. That, in turn, raises the question ‘how much forgiveness is enough?’ It is an unresolved question, but one which will undoubtedly generate disparate views. If one has a drug regimen that can forgive the vast majority of 1 and 2 day lapses in dosing, without loss of effectiveness, then, implicitly at least, fully compliant patients are exposed to more drug than they need in order to benefit incompletely compliant patients. Of course, sooner or later the vast majority of patients commit errors of omission in their execution of prescribed drug regimens, at which time they benefit from drugs with greater rather than lesser forgiveness. Each drug and disease situation has to be looked at on its merits. From the labelling perspective, as already noted, it would seem only reasonable to include information on what to do when single or serial doses are missed.
the word ‘go’. This new field\textsuperscript{18} subsumes matters that meant little when prescription drugs were weak or ineffective, and were usually used singly rather than in complex combinations. Pharmionics has gained in importance as drugs have gained in strength and potential for harm if misused. Pharmionics is akin to avionics, which became essential as flight gained in power and speed, exceeding the unaided pilot’s ability to control flight reliably. The common Greek root in both terms is ionics, from the verb to go. Today, we need to quantify reliably how the drug ‘goes’ in its intended use, as with avionics in flight. One might say that each of the instruments in the cockpit of modern aircraft is a memorial to earlier pilots who once needed, but lacked, the information provided by the instrument in question. Analogously, in the pursuit of pharmionics, we seek to build a body of reliable measurements and analyses of their clinical correlates that can make the ‘going’ of modern, powerful drugs safe and effective for all, wherever they are used.

8. Implications

Intention-to-treat analysis of ambulatory drug trials has been cosseted by trialists’ long perseveration with the returned tablet-count method of estimating drug exposure in clinical trials. Yet repeated studies since 1989 have shown that returned tablet counts ‘grossly overestimate’ (to use the term in the most definitive study\textsuperscript{54} of trial patients) exposure to test drugs. Returned tablet-counts routinely show that >90% of patients are satisfactorily adherent, when the counts are interpreted against an assumed criterion that taking 80% or more of prescribed doses suffices for full effectiveness, and against the assumption that an unreturned dose was an administered dose. The latter point is challenged by a variety of findings, starting with the chemical marker study of Pullar \textit{et al.},\textsuperscript{54} followed by multiple comparisons with electronic monitoring that consistently show higher compliance figures coming from returned pill-counts than from electronic monitoring. Moreover, data on returned counts, when substantially larger numbers of doses are dispensed than are needed for full compliance during the interval between scheduled visits, show that more than a third of patients either discard or hoard untaken doses, returning an empty or nearly empty container at the next scheduled visit.\textsuperscript{53,64} Moreover, the ‘80 is enough’ criterion is obviously not only drug-specific but formulation-specific, and patently can be either an under-estimate or an over-estimate, depending on the agent, as discussed earlier. So the prevailing standard for analysis of drug trials is heavily burdened by almost universal use of a thoroughly discredited method for assessing drug exposure, and failure to recognize that intention-to-treat analysis, through its averaging of sometimes qualitatively different responses to the same drug, underestimates risk in patients who are at-risk from certain dosing patterns, and over-estimates risk in the majority of patients who are not at risk from these patterns, because they do not practice them.

A main factor in this story is the relatively high frequency of multi-day lapses in drug dosing, known as drug ‘holidays’, defined as 3 or more consecutive days of interrupted dosing. Drug holidays serve as a trigger for rebound, recurrent first-dose, and other effects of physiological counter-regulatory responses to primary actions of drugs.\textsuperscript{58} Their incidence in medically unselected patients is about 2400 holidays per thousand treated patients per year,\textsuperscript{58} though they are concentrated in about a third of patients, half of whom have 3–4 holidays per year and the other half of whom have one or more holidays per month. Holidays can occur in patients who are otherwise punctual in their taking of prescribed medicines, so aggregate measures of percentage of prescribed doses taken or percentage of days within which the correct number of doses were taken, may be close to 100%, with still the occurrence of occasional holidays. It is numerically possible, and occasionally realized, that a patient has, e.g. six 4 day holidays in a year, so that 24 days’ doses out of 365 are missing, which results in 93% of prescribed doses taken, or higher if the patient happens occasionally to take an extra dose. Any figure in the 90%–100% range is routinely considered ‘good adherence’ on ad hoc numerical grounds, but is not inconsistent with the patient’s having been exposed, depending on the drug in question, to recurrent episodes of explosive microorganism replication in the case of antimicrobial treatment of short cycle-time microorganisms, e.g. HIV.

What we describe here suggests an opportunity to strike a new balance in drug development, following the learning and confirming sequence\textsuperscript{37} which gives promise of improving current levels of productivity. Research over the past decade has shown, with repeated confirmation, that a large minority of ambulatory patients spontaneously vary their dosing intervals, sometimes widely, and often repeating essentially the same temporal patterns of dose omission. The similarity of patterns of underdosing across many qualitatively different fields of ambulatory pharmacotherapy\textsuperscript{10,11,40} would suggest that ‘compliance’, defined as the degree of correspondence between actual dosing history and prescribed regimen, is largely an attribute of the patient and his/her ability to organize and execute punctually the 8-day task of dose-taking. In contrast to past sayings, mostly devoid of evidential support, this new construct removes compliance from being an attribute of drug, disease, prognosis, regimen convenience and the like. ‘Persistence’, on the other hand, defined as the time between the start of dosing and its permanent cessation, appears to be strongly influenced by drug side effects, the degree of realization of expectations about beneficial effects, difficulties in executing the regimen, the economic cost of treatment and probably other factors that lead patients to abandon the treatment regimen. The roles of these factors, however, have been inadequately studied. In terms of ‘learning’, in Sheiner’s sense of the word, variable compliance, as defined above, is rich with opportunities for observing the temporal sequence of change in dosing followed by change in drug actions, strengthened not only by temporal sequence, but also by repetition of the same sequence from time to time in the same patient. It is also, seemingly, largely free of the biases that have been repeatedly ascribed to measures of ‘adherence’ or ‘compliance’, on the basis of data, both sparse and unreliable, and imprecise terminology. In contrast to ‘compliance’ as defined, the permanent cessation of dosing, which is a one-time event and probably often difficult to monitor because of its frequent association with dropping out from the study in question, is obviously heavily burdened with biases related to individual differences in, perceptions of, and realized or disappointed expectations about, drug action. Patients who do not persist in randomized, placebo-controlled trials carry with them strong biases closely connected to disease severity, strength or weakness of drug responses, difficulties in taking the prescribed medicine(s), thus leaving a complementarily biased residual group of patients, who continue to the end of the trial.
What can be learned from short persisters? Clearly, a careful examination of data gathered from them prior to their cessation of dosing may create a strong basis for predicting whether and when they stop. Delmas et al.1,2 found in the IMPACT study that patients who had complied well but had received no response, or a further loss in bone mineral density, almost universally dropped out of treatment. In contrast, a simple intervention, executed by the prescriber to motivate patients to continue with treatment, was effective in those who also had positive bone marker responses.

A key point in all this is that the arrival of reliable methods of measurement plus refinement in the taxonomy of deviations from prescribed ambulatory pharmacotherapy have created a promising learning opportunity that has only occasionally been exploited in the past.3,13,15,33,35,66 It will, naturally, be obligatory to find ways to confirm salient pharmacodynamic findings that emerge from the clinical correlates of variable dosing, although ethical considerations may preclude such purposeful experiments, as we have seen with the study of Vanhove et al.35 When formal confirmation by randomized, placebo-controlled trials is impossible, one must rely on the approach described by Hilt16 and subsequent work on causal inference.67–71

The advent of modelling and simulation, of better understanding of the pharmacodynamic consequences of the common patterns of aberrant dosing, of recent advances in causal inference in biostatistical analysis, and of electronic compilation of ambulatory patients’ drug dosing histories, all make it timely to use the natural experiment in dose-ranging to increase the amount of information that one can retrieve from clinical trials.

We have not considered the question of whether adherence is improvable, and if so, at what cost and with what degrees of success? Ongoing studies, based on the use of electronically compiled dosing histories,72,73 and sometimes combined with pharmacokinetically projected drug concentration time courses (Figure 1) show promising early results, but this takes us beyond the scope of the present paper.

Acknowledgements

We are indebted to the following for many useful discussions on repeated occasions of various aspects of the subject matter of this Review. In so saying, however, we do not imply that all those listed subscribe to all points made in this review. Those we acknowledge and thank are: Roy Anderson, David Bangsberg, Richard Bertz, Terry Blaschke, Joyce Cramer, Bradley Efron, Charles Flexner, Els Goetghebeur, Robert Gross, Mitchell Levine, Steven Mayer, France Mentré, David Paterson, Carl Peck, Richard Rode, the late Alvan R. Feinstein, the late Louis Lasagna and the late Lewis B. Sheiner.

References


