Patient non-compliance with drug regimens: measurement, clinical correlates, economic impact

J. Urquhart

From the University of Limburg, Maastricht, Netherlands

Poor compliance with rationally prescribed drug regimens attenuates benefits of treatment, making compliance a key link between process and outcome in ambulatory care. Compliance is defined as 'the extent of correspondence between the patient's actual dosing history and the prescribed regimen'. Electronic monitoring methods reveal that >30% of patients omit many prescribed doses, irrespective of disease, prognosis, or symptoms. Some drugs are better able than others to maintain therapeutic action during the more common lapses in dosing. These are called 'forgiving' drugs; their duration of action is more than twice the prescribed interval between doses, allowing action to continue when one or more doses are missed. Forgiveness has limits, so long lapses in dosing will nullify action of any drug, with economic consequences that depend on the clinical consequences of lapsed action, or, with some drugs, rebound effects. These practical points have only come to light with the use of electronic monitoring of compliance, which avoids the biases created by tablet counts and other methods that make it easy for patients to censor evidence for omitted doses. All else being equal, the most forgiving drug in its class will be associated with the best outcomes, for it will be least impacted by prevalent poor and partial compliance. (Eur Heart J 1996; 17 (Suppl A): 8-15)

Key Words: Compliance, pharmacokinetics, economics.

Introduction

Rationally prescribed pharmaceuticals are a principal interventional arm of primary care and a principal element of economical healthcare. Both therapeutic and side effects of prescription drugs depend upon dosage of drug administered, and the timing of doses. Dosage and dose timing are guided by recommended dosing instructions, which are, in turn, based on findings in clinical trials. The labelling for prescription drugs includes dosing recommendations that are supposedly optimal, in the sense that the recommended dosage and timing of doses are most likely to elicit useful therapeutic action, with fewest/least severe side effects. There are, of course, many caveats specific to therapeutic fields and classes of drug, but it is fair to say that, in the main, dosing which deviates from the optimal regimen will result in suboptimal outcomes of treatment. For this reason, it is appropriate to say that patient compliance in ambulatory care is an important: (1) link between medical process and treatment outcome; (2) determinant of the quality of medical care.

Definition of compliance

Recent methodological advances have made it possible to define drug regimen compliance as 'the extent to which the patient's actual history of drug administration corresponds to the prescribed regimen'.

This definition includes both quantity and timing of doses. Thus, this definition puts the measurement of compliance into terms that allow a pharmacometric interpretation. That is to say, 'compliance', as defined, can be used as input to a quantitative pharmacokinetic/pharmacodynamic model to predict, from the actual dosing record, the time course of drug concentration in plasma and the time course of drug actions. Indeed, this has been undertaken by Rubio et al. who were the first group to utilize electronically measured compliance data to make and confirm such projections.

Prevalence of poor compliance

Substandard compliance would be of only minor concern if it were not so prevalent or if it were limited only to medical conditions of a self-limiting or otherwise minor nature. The reality, however, is that a third or more of patients comply poorly with prescribed drug regimens, more or less irrespective of disease or prognosis. For example, there is a broad consensus that poor compliance with post-transplant immune suppression regimens is a leading cause of rejected organ transplant. Furthermore, compliance varies widely among patients being treated for breast cancer or AIDS.
Clinical correlates

The clinical correlates of common dosing errors depend on the drug, the disease and its severity and comorbidity. To illustrate this fundamental point, consider a patient who misses several days' doses of a loop diuretic, such as furosemide. If the patient is a young woman in good health but prone to premenstrual fluid retention, the diuretic is a convenient way to avoid an unpleasant but not threatening problem. In effect, the diuretic in this situation is an optional, comfort medicine. If such a patient omits doses, she will have some degree of transient fluid retention that may be inconvenient and even uncomfortable, but without threat to her well being. In contrast, if an elderly patient with moderate to severe congestive heart failure (CHF) omits doses, the consequences will be more serious in that, after several days of lapsed dosing, the patient may develop acute pulmonary congestion, with signs and symptoms of fluid volume overload, and present for emergency care in acute respiratory distress.

Thus, as these examples illustrate, the medical status of the patient is an important determinant of the consequences of an error in compliance with a particular drug. The same error made with a different drug will have its drug-specific consequences. For example, the patient with heart failure could omit a prescribed laxative instead of the prescribed diuretic with only minor consequences. Unless serious colonic disease was an important comorbidity, several doses could be omitted with only the discomfort of increasing constipation.

Pharmacokinetic and pharmacodynamic correlates

Dose–response relations of many drugs are non-linear, with some regions of steep slope, and other regions with shallow slope. Dosing variations within a steep-slope region will have large impact on drug actions, whereas dosing variations within a shallow-slope region will have small impact[7]. This is one reason why some drugs are considerably more 'forgiving' of the more common errors in dosing. Another basis for forgiveness is that the drug has a considerably longer duration of action than the prescribed interval between doses. When action persists well beyond the scheduled time for the next dose, the patient can be late in mediating, without loss of action. The more forgiving drugs in a pharmacological class are better able than their less-forgiving classmates to maintain beneficial action when compliance varies[8,9]. Forgiveness is a relatively newly recognized attribute that can differentiate agents that are otherwise considered to be therapeutically similar.

It should be noted, however, that the virtues of superior forgiveness will be most clearly seen in partially compliant patients, who stand most to benefit from the attribute. Obviously it is important to establish that superior forgiveness does not impose needless side effects or hazard on patients who comply fully. To understand most clearly the potential positives and negatives of forgiveness, it is necessary to test the agent in patients whose degree of compliance has been previously defined, selecting partial compliers to reveal most clearly the beneficial aspects, and full compliers to reveal most clearly any negative aspects.

Economic correlates

The medical and economic importance of compliance errors tend to go hand in hand but are, of course, specific to the therapeutic field and drug. If the drug is a simple 'comfort medicine', whose effects are minor and whose use is optional, then variable compliance will have little practical import. In contrast, if the drug is crucial for maintaining vital physiological functions, then the medical and economic consequences of variable compliance can be large. For example, the average contemporary cost of each hospitalization for complications of CHF is US$10 400[10].

Indeed, an economically informative example of the costs of non-compliance is diuretic treatment in CHF, where the omission of as little as 3 consecutive days' dosing can result in sufficient fluid retention to precipitate acute fluid overload, with pulmonary congestion or oedema, and dyspnoea[11]. The $10 400/episode cost of hospitalization for failed diuretic treatment contrasts starkly to the approximately 5 cent/day cost of the omitted doses of furosemide. A patient whose dosing lapses result within a 1-year period in two hospitalizations for acute fluid retention incurs costs of $57/day, representing a thousand-fold increase over the cost of the drug.

A further informative example is provided by the low-dose, combination oral contraceptives. Recently-adopted labelling in the U.K. informs patients that a delay in dosing of >12 h risks breakthrough ovulation and conception. Accordingly, the patient is instructed to take the delayed dose as soon as the delay is recognized, but also to use back-up barrier contraception for 7 days thereafter, to guard against inadvertent conception[12]. The oral contraceptive regimen, though once a day, is convenient but rather unforgiving of relatively minor variations in dose timing. Even more unforgiving is the regimen for the progestin-only 'minipill'. The present consensus is that a patient who is more than 3 h late in taking the once-daily tablet should use back-up barrier contraception for either 48 h or 7 days[13]. It is, therefore, not surprising that so many patients have opted for the implantable, 5-year mode of contraception[14].

The economic consequences of unwanted conception vary from an average $300 cost of first trimester abortion[15], to the tens of thousands of dollars needed to raise a child to the point of economic independence. The passions that surround unwanted conception are a reminder that economic aspects are sometimes only a small part of the healthcare story.
‘Suttonomics' and the non-compliance tax

Focusing on high-cost healthcare situations gives rise to a useful technique called ‘Suttonomic analysis’[19], named after Willie Sutton, a notorious bank robber, who, when asked why he robbed banks, replied: ‘that’s where the money is’. Suttonomic analysis requires a comprehensive database from which one can extract and rank order annualized care costs for all recipients of a particular drug or a class of drugs. Focusing on patients with the highest costs, one sees that some had nothing to do with the drug in question, while others were incurred because the drug failed to work as expected. This distinction must be made by analysis of individual case records. Some failures of the drug to work as expected have organic explanations, but others are due to poor compliance. This last distinction can be made in part on the basis of analysis of case records, and in part by prospective monitoring of compliance in patients who are high utilizers of medical resources.

It is convenient to express costs attributable to poor compliance in monetary terms as a daily ‘non-compliance tax’ on the per-day acquisition cost of a drug, exemplified by the >$57/day costs due to two hospital stays, within a 1-year period, for acute fluid retention in a patient with CHF. The contrast can be striking.

In the usual situation, a relatively small proportion of patients accounts for a large proportion of care costs. It is especially important to understand the reasons why treatment has failed in these high-cost care situations. Some of these are attributable to poor compliance with effective agents, resulting in estimated ‘non-compliance taxes’ that are orders of magnitude higher than the cost of the drug. Therein, of course, is a lesson for looking beyond drug prices in efforts to contain costs.

Biases abound in such data, so sound techniques of pharmaco-epidemiology, together with selective use of controlled studies, are needed to dissect the relevant from the irrelevant, to avoid bias and to distinguish causality from association. Properly controlled studies avoid the problems of bias, but may not be ethically possible in many situations. Controlled study of the consequences of omitting anti-hypertensive drug doses has been pioneered by Johnson and Whelton[17].

Time scales

In gaining an understanding of the pharmaco-economic impact of poor compliance, it is obviously much simpler to deal with situations where drug dosing, drug action, benefit, and risk are closely linked in time. With pharmaco-economics still in its infancy, it is logical to analyse simpler cases first.

Importance of patient targeting

In devising cost-effective interventions, targeting of patients is important, so that intervention is directed at those who stand most to benefit, with the least misdirection of resources to those who either do not need intervention, or cannot benefit from it. An example is the study in type II diabetics described by Matsuyama et al.[18] in which patients were randomized between two bases for management: an objective measure of compliance and patients’ self reporting of compliance. The pharmacists who relied upon patients’ self reporting for information about compliance interpreted almost all problems in management as a need for dose adjustment. In contrast, pharmacists who had an objective measure of compliance found that half the problems were due to inadequate dosing. Thus informed, they could address real, not fictitious problems.

From the economic perspective, the non-compliance tax can guide targeting, to direct $2 interventions at $20 problems, not at 20-cent problems. This point is fundamental in quality management, one of whose pioneers was the late W. Edwards Deming, who pointed out throughout his long career that reliable measurement is essential in the quest to improve quality, in that measurement prioritizes efforts, so that the biggest of the tractable problems are attacked first[19].

That principle is captured in the idea of Suttonomic analysis. Measurement also reveals which interventions work best, assisting one’s efforts at improvement. That is why the development of electronic methods for measuring medication compliance, which came only in 1986/7, is important[20].

Historical aspects

Since 1962, many drugs of proven efficacy and acceptable degrees of safety have been developed. Those and other types of medical progress have brought us from few treatment choices in low-cost care to many choices in high-cost care. If that transition was gradual, its recognition by society has been both recent and abrupt, creating powerful forces for change in the organization and financing of healthcare. We are now in a ‘post-modern' era, where proof of efficacy no longer suffices as proof of value, except when it involves the very first agent available for treating a hitherto untreatable condition. In today’s world, the much more usual case is that a new agent is the second, third, fourth, etc. agent for an already treatable condition. Competitive advantage is to be found and is, of course, keenly sought, from data on medical and economic outcomes of product use.

Statistical aspects

In light of the new information on the prevalence and forms of patient non-compliance, it is difficult to contemplate the process of drug development and pre-market clinical trials without becoming alarmed, because so little attention is paid to non-compliance in most drug trials. Some of this has to do with poor
methods, mainly employing the counting of returned tablets, which have grossly underestimated the extent of poor compliance in drug trials. Statisticians have also contributed to this problem by advocating primacy of the so-called 'intention-to-treat' policy for analysing trials results, whereby one ignores doses actually taken and analyses the results on the basis of the treatments to which patients were assigned. There are, of course, some good reasons for wanting to undertake intention-to-treat analysis, because it avoids certain biases that may enter after randomization to selectively influence results in one or another arm of a randomized, controlled trial.

However, there are a number of problems created by sole reliance on intention-to-treat analysis in drug trials. Sheiner and Rubin point out that intention-to-treat analysis gives a biased estimate of the drug's actions when the dosing instructions are followed incorrectly, as occurs in a large minority of patients. Feinstein has described many of the conflicts between intention-to-treat analysis and common sense, plus the large amount of useful information it suppresses. One such important piece of information is the magnitude of drug responses in patients who correctly follow the recommended regimen. The consequences of taking substantially less drug than prescribed are also suppressed by such an approach, as is the general shape of the curve relating drug exposure to drug effect. All this information that is censored by intention-to-treat analysis constitutes, among other things, a challenge to the recommended regimen: whether the dose and/or dose frequency are set too high, too low, or about right. Given that dosing requirements are sometimes grossly overestimated in pre-market trials and initial labelling, such information can be a useful pre-market 'sanity check' on the recommended regimen.

Of all the problems created by sole reliance upon intention-to-treat analysis, perhaps the most striking is the ethical clash it creates with the long-standing policy for drug labelling. Hutt and Lasagna have pointed out that drug labelling based solely on results of intention-to-treat analysis misleads the substantially compliant majority about the consequences of taking the drug correctly and represents a clash between the ethical foundation of full-disclosure labelling and a widely-practised statistical policy.

Intention-to-treat analysis has the virtue of simplicity, while providing immunity against all manner of self-serving analyses that one might employ, post hoc, to find some basis for claiming that the agent under test does something useful. In that sense, intention-to-treat analysis serves those who prefer to avoid false conclusions about effectiveness, at the risk of falsely concluding that effective agents are ineffective.

It is a policy well suited to a pre-scientific era when proffered treatments were most likely to be therapeutically ineffective or even hazardous. For an era in which new medical interventions arise from an increasingly robust body of scientific knowledge, however, one might reasonably prefer to shift the bias in order to minimize the risk of false conclusions of ineffectiveness. It was appropriate for Hippocrates to enunciate the principle primum non nocere (above all, do no harm) but it is no more applicable to today's medicine than another of the original Hippocratic strictures, in which physicians swore not to perform lithotomies: 'I will not cut for stone'. Most of modern medicine and surgery, and interventional cardiology in particular, is a monument to the irrelevance of such dicta in an era of strong science and technology.

Clearly one way to draw a false conclusion that a new treatment is ineffective is to ignore widespread underdosing and its consequences in the test population. Also, since underdosing in certain temporal patterns practised by a small minority of patients can create exceptional hazard, one can also draw the erroneous conclusion that the agent is too hazardous for general use by ignoring the linkage between misuse and hazard.

Efron and Feldman have described one way to minimize the risk of errors of both types in using patient compliance data as an explanatory variable in trials analysis. As Rubin pointed out in his discussion of the Efron–Feldman paper, the wide spectrum of underdosing that prevails in trials presents a natural experiment in variable dosing that is potentially rich in therapeutically practical information, including dose response and alerts as to certain patterns of non-compliance that may be especially hazardous. As I have discussed elsewhere, the concerns about major bias being linked to variable compliance with drug regimens are probably greatly exaggerated by those who are biased towards preservation of the status quo, avoidance of type I error, and simplicity of analysis.

For the foregoing and perhaps other reasons, increasing attention is being paid to the statistical analysis of compliance data and its challenge to the statistical status quo. In addition to recently published papers on the topic in the statistical literature, a series of symposia has been held on the topic at the 1994 Winter Meeting of the American Statistical Society, the 1995 annual meeting of the American Society for Clinical Pharmacology and Therapeutics, and an international symposium in 1995 at the Belgian University of Limburg.

Cautionary note for economists

Economists must take care about these statistical policy issues, because the average dose taken in a drug trial involving ambulatory patients is usually considerably smaller than the assigned dose, and of course the averaged effect of the drug is likely to be diluted, relative to the full-dose effect, by underdosing among a large minority of trial participants. This 'dilution effect', named by the late Professor Ellen Weber, was a factor of 2 in the analysis of both the Lipid Research Clinics — Coronary Primary Prevention Trial (LRC–CPPT) of...
cholestyramine[^37], and the Helsinki Heart Study of gemfibrozil[^28]. The analyses of these trials, and the role that variable compliance played in them, have been reviewed by Urquhart[^39].

### Basis for hope

Fortunately, as noted above, the distortions created by variable compliance in drug trials have become a topic of strong interest in biostatistics. Several new trial designs have been described to help reconcile the sometimes conflicting demands of full-disclosure labelling, avoidance of bias in trials analysis, and common sense[^17,26]. These new methods should improve the quality of information obtained from randomized, controlled trials. Economists must be attentive to the statistical analyses done, and the manner in which patient compliance was measured and used, or ignored, in the analyses.

Another source of hope is the development of financial analyses and controls in healthcare, which can provide a system-wide perspective and antidote to the 'pennywise, pound-foolish' attitudes that have prevailed for so long in respect to prescription drugs and medical devices. However, this developing capability has had the one prominent side effect of inducing vast turmoil in the pharmaceutical marketplaces of the western world, for this new perspective has suddenly and forcefully prompted questions about outcomes of treatment[^40] not usually asked about prescription drugs, thus changing the basis upon which prescription drugs are valued[^41]. In this new environment, the natural experiment created by prevalent poor compliance can be a very useful aid to understanding, though not a panacea. The present high hopes for 'outcomes research' as a shortcut to better understanding will inevitably be dashed. The predictable outcome of outcomes research will be the relearning of old lessons that proof of causality demands rigour and controls that only rarely present themselves spontaneously in observational studies, e.g. routine medical care[^42,43].

### Outcomes

The assessment of outcomes is different from null hypothesis (NH) testing in randomized, controlled trials which remain a standard for pre-market proof of efficacy. A successful test of efficacy by NH rejection supports only the claim that the product is, to put it in literal terms, 'not no better than nothing'. That claim is synonymous with value when the first effective agent enters the market to treat a hitherto unbeatable disease. In a multi-product market, however, being 'not no better than nothing' is a necessary claim, but not one that asserts competitive advantage over other agents that are also 'not no better than nothing'. In other words, rejection of the NH is not a sufficient indicator of value in a multi-product market[^40].

A model for outcomes studies is, as already suggested, the compliance-stratified analyses of the LRC–CPPT[^37] of drug-induced reduction in plasma cholesterol levels and coronary risk. The labelling[^44] informs both prescriber and patient that the compliant majority had twice the all-patient average reduction in coronary risk, which was diluted by smaller effects of lower doses of drug taken by the large minority of poor and partial compliers. This landmark study teaches many lessons[^39] and has been the basis for an important advance in statistical use of compliance as an explanatory variable in trials analysis[^33].

However, LRC–CPPT went on for 7 years and cost over US$130 million, so one naturally looks for simpler, quicker sources of data. Two experiments of nature have attracted attention. One is variable prescribing of like-indicated drugs to patients with seemingly comparable health status. The other is variable patient compliance with rationally prescribed drugs.

#### Variable prescribing

Sclar et al.[^45] and Hilleman et al.[^46] have gathered such data, which reveal interesting differences in aggregate costs, but leave open the question of whether the observed differences are attributable to differences among prescribed drug products or to unquantified differences in patients' health status. The studies occurred in the usual, powerful cross-current of pharmaceutical promotion, aimed at creating differentiated positioning of like-indicated products in crowded markets, making channelling and other sources of bias likely. Channelling refers to the tendency for certain drugs in a class to be selectively prescribed to patients with special problems, e.g. more severe disease, certain types of comorbidity, or worse prognosis[^47-49]. As Feinstein's group has shown, differences in prognosis are difficult to identify and adjust for, but have consequences that can overwhelm the influences of different treatments[^43].

### Variable compliance

Variable compliance with recommended dosage regimens is also a natural experiment, creating a spectrum of drug exposure ranging from essentially zero to slightly more than the recommended dose. In effect, it creates the opportunity to assess the medical and economic correlates of variably proficient use of the drug, including what Feinstein identified many years ago as a virtual placebo group in those who are prescribed the drug but take essentially none of it[^50]. The natural experiment of variable dosing was long beyond reach due to methodological limitations, but it is now accessible with electronic means for monitoring the time histories of dosing in ambulatory patients. There are certainly valid questions about non-drug influences...
linked to dosing behaviour, but several methods have been used to test for these\[17,26]\.

Suttonomics, as described earlier, is a new and seemingly efficient method of analysis, offering a basis for identifying the specific dosing correlates of costly therapeutic failures. From these, one may identify interventional tactics, more nearly optimal regimens, specifications for novel drug delivery systems, or other means for avoiding dosing errors with powerful drugs. Armed with reliable measurements of the non-compliance tax, one can make an appropriate judgment about the cost of preventing or circumventing crucial errors in compliance, in the interests of better medicine, less misery, and cost containment. There is, after all, a hierarchy of interventions from which one can choose in order to ensure satisfactory compliance with rational treatment. Figure 1 illustrates this hierarchy, the costs of which increase disproportionately with intensity. Incarceration is occasionally used as a last resort in tuberculosis treatment, because of public health hazards of incomplete or erratic treatment, and the likelihood of its favouring the emergence of resistant microorganisms\[51\]. Directly observed therapy is used for the same reason\[52\]. Home telephone-based measurement is a new form of telemedicine, in which electronic compliance monitoring data are automatically gathered and sent by modem for daily review and, if need be, phone-based intervention, at a cost of about $2.50/day. Monthly review and intervention, which are appropriate for long-term prophylactic regimens where a few days or weeks of substandard compliance has little adverse effect and a small non-compliance tax, can be undertaken in conjunction with prescription refills. The answer to which of these interventions are cost effective is both disease- and drug-specific.

Provision of information on drugs and their actions is essential for patients to make informed decisions regarding the initiation or cessation of drug treatment. Such knowledge, however, does little to ensure that doses are taken regularly and on time. If knowledge about drugs and diseases could ensure punctual compliance, then all physicians, nurses, pharmacists, and other pharmacognoscenti would be strictly punctual compliers. That is not so, as many readers can confirm from their personal experiences\[41]\.

**Conclusion**

Powerful prescription drugs cannot realize their full therapeutic and economic values unless administered within certain limits of an optimal regimen. Those limits are narrower for some drugs, broader for others, with the more 'forgiving' drugs being better able to maintain therapeutic effect than those with a less forgiving pharmacodynamic profile\[9\]. Unrealistically narrow tolerances for common errors in dosing can undermine value\[41]\, but these tolerances may be widened by
appropriate regimen design and/or use of novel drug delivery systems technology. Such basics are now amenable to definition in relatively simple studies and should be mastered before investment is made in long, costly outcome studies.

References


