Reviewer's report

Title: Electronic monitoring of non-adherence to medication therapy: examining underlying assumptions

Version: 1 Date: 25 July 2007

Reviewer: John Urquhart

Reviewer's report:

General
The objectives of this study, as stated in the conclusion, are laudable, but there are both methodological and conceptual problems with the paper.

After a goodly amount of hemming and hawing, the authors make the assumption that patients' self-reports are reliable about the temporal sequence of their intake of prescribed medicines. If that assumption were correct, why would anyone bother with electronic monitoring? Why not just ask the patients, write down their dosing histories, and analyze those data for their clinical and economic explanatory power? In fact, there is a considerable history of studies showing, by various means, that, while patient reports of missed doses are probably believable, patient reports of good adherence may or may not be. Yet the authors use these reports as a 'gold standard' for judging electronic monitoring.

The field of adherence research was greatly hampered, until about 2 years ago, by the lack of an external reference standard against which to evaluate dosing history data gathered by various methods. It turns out that the independent reference standard had been described by Rubio et al. (Clin Pharmacokinet 22: 238-246, 1992), but not recognized as such, in a study showing close correspondence between measured concentrations of drug in plasma and projected concentrations derived from using opening times of MEMS monitors as input to a patient-individualized pharmacokinetic model of the drug (nifedipine) in question. In 2005, a study (J Clin Pharmacokinet 45: 461-7. 2005) with 35 HIV+ patients observed for 1 year used the same approach plus improved PK modeling during the intervening 13 years, and showed close correspondence between concentrations of lopinivir, measured bimonthly, and concentrations of lopinivir projected for the times of the blood sampling, from MEMS-compiled package opening times. Fig 2 in the 2005 paper shows the small residual errors of all but 3 of the 210 comparisons. Such correspondence is inconsistent with substantive differences between EM-captured and actual dosing histories. Note that this was a bias-free comparison, because those doing the measurements of drug concentrations were completely separated from those working with the patients, and they in turn were isolated from those who did the pharmacokinetic projections. The present paper does not cite this study and its 1992 forerunner, which are epochal for adherence research, because at last there is an external,
bias-free standard for assessing the reliability of any method that purports to compile drug dosing histories. Even if ‘white-coat compliance’ occurs, the issue is not whether the blood sample that is being drawn is representative of the patient’s prior dosing history, the question is what the magnitude of the residual error between that measured concentration of drug and that of a PK projection based on a patient-specific PK model and the patient's electronically-compiled prior dosing history.

There is no indication in the present paper’s description of methods that staff who interviewed the patients were blinded to the nature of the study design and the study hypothesis. This is an important matter for it impacts on how patients responded to the 'structured interview'. There is a long history of subtle ways in which trial staff, who know their boss’s biases, can question patients in such a way as to elicit answers that will please their boss or otherwise increase the likelihood of proving the study hypothesis. In his masterful book, “Clinical Epidemiology: The Architecture of Clinical Research” (Saunders, 1985), the late Alvan Feinstein (who also coined the term 'white-coat compliance', wrote on these problems on pg 299, under the headings 'interviewer bias' and 'recall (anamnestic) bias'. In the present study, these biases, unless they were in fact satisfactorily prevented, represent a fundamental flaw.

From where comes the assumption that a measurement method is invalid if it perturbs the thing being measured? Consider temperature measurement. You have a beaker containing a fluid, the temperature of which you wish to measure. You insert a thermometer, which necessarily starts out at a temperature different from that of the contents of the beaker. Thereupon, either a warmer thermometer warms the fluid in the beaker, or a cooler thermometer cools the fluid in the beaker. Of course, if you knew a priori the temperature of the fluid, you could bring the thermometer to that temperature, but (a) you cannot know that without a measurement, and (b) if you knew the temperature already, you wouldn't need to make the measurement. Obviously, it is a question of the magnitude of the perturbation. Only an idiot would take a thermometer out of liquid nitrogen and use it immediately to measure something that has been at room temperature. In the case of electronic medication event monitoring method, it is invariably accompanied by one or more human interactions between patient and caregiver, the character of which can be expected to vary and create influences of indeterminate magnitude and direction. Naturally those influences play on what happens after the interaction. The external standard, based on projected vs actual concentrations of drug in blood, looks backwards in time, as the values being compared arise from doses already taken, not taken, or erratically timed, prior to the blood sampling.

As far as the trend of adherence change after the start of MEMS monitoring goes, Vrijens has shown that the strong correlation between missed doses and early discontinuation results in patients who miss many doses dropping out of study cohorts, so that adherence rises with time, not because individual patients do better, but because those who miss doses frequently have departed from the cohort. That is one ‘force’ operating on the time-course of adherence after
monitoring starts. Another is the nature of the interactions between trial staff and the patients, which, as noted above, has a certain ability to encourage or discourage punctual dosing. Yet another is the ‘hawthorn’ effect, which would lead one to expect that the introduction of a new element in the patient’s care would lead to a temporary improvement, and, as the authors note, the study may have picked up the declining ‘tail’ of a hawthorn effect already in progress as the study data started to be collected. When all is said and done, the changes in adherence over time, after the start of EM in this study, were miniscule, so that one could scarcely argue that using electronic medication event monitoring decreases adherence, which of course would pose an ethical issue. If it increases adherence a bit, so much the better, for pragmatists, at least.

Major Compulsory Revisions (that the author must respond to before a decision on publication can be reached)

The authors should acknowledge explicitly that, although an absolute external reference standard now exists (which should be referenced), they cannot, against the long history of unreliability of self reports about when or whether doses were taken, proceed on the assumption that self-reports are robust enough to challenge EM data. See, e.g., fig 1 in the paper by Tu et al. in Clin Pharmacol Ther 77: 189-201, 2005, which shows how wide the nonadherence-driven discrepancies can be in terms of internal drug exposure and clinical assumption that derives from what patients tell their caregivers.

The question of observer bias, and what steps were taken to prevent it in this study, is an essential matter to resolve.

The authors should acknowledge that the website www.aardexgroup.com gives a failure rate of 0.5%, which is essentially what is reported here.

The authors should acknowledge that the MEMS monitors discussed in this paper, MEMS-V, are one generation short of the presently marketed MEMS-VI. There is a passing reference to MEMS-IV, but it is only fair to point out that the manufacturer of MEMS monitors has pursued a many-year course of successive improvements, starting from MEMS-I, which is what Joyce Cramer used in the study that resulted in her JAMA paper 18 years ago. MEMS-VII is in the last stages of development, before introduction. As with reviews of automobiles, it is only fair to the reader to keep the models straight, and not burden perceptions of the 2007 model with problems that were limited to the 2004 model.

Minor Essential Revisions (such as missing labels on figures, or the wrong use of a term, which the author can be trusted to correct)

The Methods section does not identify the source of the MEMS-V monitors used in this study.

Table 4, only referred to in passing on page 12, includes mysterious numbers
that are not commented upon, notably the concentrations of the various immunosuppressant drugs. What is one supposed to make of these figures? There is a vague implication from the title of the table that these values represent 'adherence', but since nothing is said about when the blood was drawn that gave rise to these numbers, how many samples per patient, time in relation to last-taken dose, etc etc. they are uninterpretable.

Discretionary Revisions (which the author can choose to ignore)

In the last word on the 6th line up from the bottom of the last paragraph on page 13, I think the authors mean 'date', rather than 'data'.

What next?: Unable to decide on acceptance or rejection until the authors have responded to the major compulsory revisions

Level of interest: An article of limited interest

Quality of written English: Acceptable

Statistical review: Yes, but I do not feel adequately qualified to assess the statistics.

Declaration of competing interests:

In addition to my wos professorship at UCSF, I am Chief Scientist of AARDEX Ltd, the developer, manufacturer, and marketer of the MEMS monitors which were used in the study, and about the functionality of which much of the paper is concerned. I have invested rather considerable sums of money over the past 20 years in the two firms that have brought electronic monitoring to its current state of development, and am known by many as 'the inventor' of MEMS monitors, which is not literally correct. "Catalyst" is a better term. I do not take a salary from AARDEX Ltd, but own a 27% interest in the company, which only recently, after about 15 years, reached the break-even point. As implied in one of my comments, it has been the policy of AARDEX Ltd and its predecessor, APREX Corp, since the inceptions of each to invest in continual improvement of its products, reflected by the fact that MEMS monitors are now in their 6th generation, and about to go into a 7th. The net effect of the paper’s publication, as submitted, would have a considerably negative effect on the scholarly reputation of BioMed Central, and, in balance, a very small positive effect on the commercial appeal of AARDEX's products. While I am named as inventor on 45 US patents, and have several in my name that are pending, none of them relate to the content of this manuscript. I don't understand the last 2 questions: "competing" with what? Don't you mean 'conflicting’? If so, no.