Author's response to reviews

Title: Examining assumptions regarding valid electronic monitoring of medication therapy: development of a validation framework and its application on a European sample of kidney transplant patients

Authors:

Kris Denhaerynck (kris.denhaerynck@unibas.ch)
Petra Schafer-Keller (petra.schaefer@unibas.ch)
James Young (jyoung@uhbs.ch)
Jurg Steiger (jsteiger@uhbs.ch)
Andreas Bock (andreas.bock@ksa.ch)
Sabina De Geest (sabina.degeest@unibas.ch)

Version: 2 Date: 12 October 2007

Author's response to reviews: see over
Dear Mrs Jazayeri,

We thank the reviewers for the helpful comments. This cover letter contains a point-by-point response to the concerns of the reviewers. Please do not hesitate to contact us if anything would be still unclear.

Sincerely,

Kris Denhaerynck

Kris Denhaerynck, PhD, RN
Research Associate
Institute of Nursing Science
Medical Faculty
University of Basel
Bernoullistrasse 28
CH-4056 Basel
Switzerland
tel: +41 61 267-0920 (secr -3040)
fax: +41 61 2670955
Eyal Shemesh

I am not able to offer any constructive criticism: the article is very-very well written. The results are presented in a concise and clear manner, and the discussion describes and explains the results and does not venture beyond the data.

Thank you very much for this feedback. No adaptations are requested with this reviewer.

Ray Ownby

As already mentioned, the four assumptions that the authors argue underlie electronic monitoring are a useful framework for analyzing the MEMS caps’ use in clinical and research studies. The first assumption, that the monitors actually work, may seem obvious but the authors correctly note that the reliability of the MEMS caps has not been routinely assessed in studies, and it is my impression that most users accept the manufacturer’s information on this point. The standard to which the authors appear to hold the device, however, seems to be 100% functioning at all times; when one monitor fails in their study, they argue that the assumption was violated. At one point in the manuscript, they note that “only” 97.5% of the units worked – this gives me the impression that they are being too critical. It would be wonderful if all measurements in research studies were only inaccurate in 2.5% of cases. This standard is thus too strict, particularly in a relatively small sample. While it would be desirable that all electronic devices function with 100% reliability, few do. I’d suggest that the authors reconsider this standard and moderate their conclusions on this point.

It is true that 100% functionality would be a very stringent standard. We have therefore changed the wording of this sentence to “EM failed to register only 2.5% of the generated events”, which implies a less critical perspective.

Their second assumption, that each time the monitored pill bottle is opened a dose is taken, is also an important issue in the use of electronic monitors. The authors correctly note the limitations of monitors’ use and the importance of assessing whether patients using them are using them correctly. The found a substantial need for adjustments of medication adherence data, and appear to argue that this is a defect in the use of electronic monitors. I’d suggest an alternative interpretation is that while electronic monitors are potentially useful, the way in which individual patients use them is important, too.

We have added a sentence to the discussion section and omitted the sentence on the impracticability of MEMS. A more balanced view of the MEMS system is expressed by the following sentence: “The fact that in our study, patients seem to not always adhere fully to the guidelines for MEMS use is not necessarily a drawback of the idea of EM itself. One can conceive of a system that is easier to use and invokes less resistance.” The paragraph continues, “A more practical, easy-to-use system may increase EM use outside of the home, thereby increasing the correspondence between event registration and pill ingestion.”

Their third assumption is that the monitoring device must have no effect on adherence behavior. Here the authors highlight an extremely important issue, but here again I
would suggest that their discussion could be more nuanced. While it is true that under ideal circumstances we would be able to assess any behavioral process without affecting it, in real-world research this is almost never possible. I think the paper might be improved for readers if the authors included a brief discussion of how MEMS caps are different from other methods of assessing adherence, such as self-report, other monitoring methods (e.g., blood levels, pharmacy records) or unannounced pill counts. Each has strengths and weaknesses, but all, it can be argued, may affect patients’ adherence. The result of the authors’ argument, at least for me, is to show that electronic monitoring via MEMS caps isn’t perfect, and I think most thoughtful consumers of research on adherence would agree that’s true.

It is true that every ethical method of monitoring patient behavior has the potential to alter the behavior itself. We have therefore added the following text in the discussion section under assumption 3: "It has to be noted that all ethically permissible adherence assessment methods, electronic or otherwise, require the consent of the subject, and therefore influence the observed behavior. It is well-documented, for example, that blood assays lead to white-coat adherence [1]. Likewise, data from the Swiss HIV Cohort Study, where non-adherence is measured every six months by self-report, show an increased adherence at each measurement time, possibly indicating an intervention effect [28]."

The authors’ fourth assumption concerns the impact of the use of electronic monitoring on sample characteristics. Here, it is not clear whether the use of MEMS caps affects the composition of a sample because of the necessary elimination of users of pill box organizers. The additional information on this point is welcome.

We have attempted to clarify this issue as follows: “Indeed, our study confirmed previous research showing less agreement to participate in EM assessment where patients said they used a pill organizer, suggesting selection bias [4]. This may be because subjects perceive an extra burden in managing an organizer and a MEMS bottle at the same time.”
When the quality tests of the MEMS are discussed on page 3 further details should be presented on the testing performed by the manufacturer. In the variables and measurement section, explanation how testing performed in this study was different than in previous studies and why this approach is superior.

Based on the review by Dr. Urquhart, we now refer on page 3 to the Aardex company website. According to Aardex, this site reports on more recent testing than the internal reports we referred to, and is an easily accessible source.

Page 4, 2nd para, should assumption 2 include patients who “trigger” their bottles while removing the meds from another source which may likely over estimate non-adherence since they may not take the cap off for the required 3 seconds that are needed for an opening (event) to be recorded by the cap?

We have added the sentence: “This also includes patients who “trigger” their bottles while taking medications from another source, but who remove and replace the cap too quickly for an opening to be recorded.”

Page 6, 1st para, describe the guidelines underlying correct EM use (resulting in unreliable data) for the reader when they are first mentioned here. Were these guidelines used to instruct the patients when they received verbal and written instructions on how to use the EM system (page 7 2nd para)?

They refer to the same instructions. To prevent confusion, we have used the same terminology but changed the sentence on page 6 to: "...who do not adhere to the instructions regarding EM use”

In the section on “Variables and Measurement” consider including a clear description and labeling of a priori operational definitions for each of the 3 measures for internal validity and the one measure for external validity. For example, was a non-functioning cap defined as a cap that didn’t record any events, extra events, some events? Why were these operational definitions selected instead those that had been previously used, e.g. by the manufacturer?

We have added some extra clarification regarding assumptions 1 and 2, as these require additional detailed information for any other research groups wishing to replicate this study.

On page 8, we have added an explanation: "When the patients indicated that a drop had occurred, we checked whether the recording system still functioned properly by scanning the uploaded EM data visually for extra recordings or for altered registration patterns (e.g., no further registered openings).

On page 11, we have added the sentence: " The manual check was done by one of the investigators, who unscrewed the cap from the bottle for one week and downloaded the registrations onto the computer. No openings were registered.”

We clarified the second assumption by adding some sentences:
- on page 1: "This also includes patients who “trigger” their bottles while taking medications from another source, but who remove and replace the cap too quickly for an opening to be recorded."
- on page 7: "EM bottles were filled in the hospital pharmacy and then sent to the patients. When a bottle was empty, the patient received a new one to which the EM cap could be attached."
- on page 7: "Examples of such guideline violations include inadvertently opening the EM bottle when no medication was required, stopping EM bottle use for a period, removing pills prematurely, triggering the cap while not removing medication from the bottle opening (e.g. taking medication from another source, cutting a hole in the bottom of the EM bottle and removing medication from there, …)"

Page 7, 3rd para, describe how the 5-point scoring criteria were developed e.g. from the literature, from experts?

We have added the sentence: "This quality assessment tool was developed with consideration for patients’ reports (recorded in field notes and later categorized) of how well they had been able to adhere to the EM use instructions."

Page 8, 2nd para, describe whether the subjects recorded the date of the dropped caps since this seems important to know for data scanning. Were the subjects asked to record any other cap trauma such as temperature or moisture changes?

Because we previously specified that patients’ EM devices were accompanied by a form, we have changed the first sentence of that paragraph to: “We asked the patients to report on their form whether they had dropped their EM caps”. They were not asked to record other cap traumas.

Page 11, first para, was the manual check of the cap completed by the researcher?

We have changed the wording from “A manual check of his cap …” into: “We manually checked his cap, which revealed that it failed to register openings…..”

Page 11, 2nd para, were the 44761 records of raw EM data the events of documented cap openings?

We have changed the sentence to: “1084 adjustments to the 44761 events of the final data base (2.4 adjustments, on average, per person)”

References 5 and 6 should be clarified. When I googled the references, they lead me to the author’s dissertation where references for this information appear to be from 2 other sources.

We now refer to the website of the Aardex group.

The first time that Figure 1 is referenced in the text, it should be described in some detail so that the reader is oriented to the figure since it is quite complex.

This is true: figure 1 was not introduced in the text. We have therefore added two sentences to the background section: “This article discusses processes that
might lead to a violation of these assumptions and describes how the assumptions were empirically tested. Figure 1 outlines the possible effects of these assumption-violating processes on adherence measurement (i.e., whether they lead to overestimation or underestimation of non-adherence).”

Clearly identify Figure 4 as it is labeled as Figure 1 in the lower left corner. The figure is important by is difficult to understand. Consider using the same words in the text that as used in the figure and clearly define them for the reader.

The figure should be labeled correctly now. The legend has changed in "Observed course of non-adherence over time”

Spelling and grammar: Page 2 under “Methods”, last sentence, “dropped” should be “drop”. Page 8 1st para, last sentence should be “were” instead of “was”. Page 13 2nd para, 3rd to last sentence “data” should be “date”. Page 14, 1st para, 2nd sentence should read “The absence of a control group prevents drawing firm conclusions. . . “ Same para, 2nd to last sentence delete “the” after “whole”.

All changes have been made.
Ramani Durvasula

In the abstract and results section, the authors interpret the finding that an initial increase in non-adherence represented an intervention effect. This is actually in contrast to what might be expected as the classic “Hawthorne Effect” which would suggest that patients would show an initial INCREASE in adherence as a result of being monitored. Theirs is almost a “reverse” intervention effect – with the intervention deleteriously impacting the desired behavior. They may consider spelling this out further.

Thank you for this feedback. Our phrasing could indeed be misinterpreted. What we actually meant is that we observed an initially lowered non-adherence and a subsequent increase in non-adherence, which we assume is the waning effect of the intervention. We have therefore changed the sentence in the abstract to: "exploring whether adherence was initially uncharacteristically high and decreased over time" and "adherence decreased over the first 5 weeks of monitoring, indicating that EM had a waning intervention effect". We have also added a sentence to the discussion section: "The prevalence of non-adherence was very low in the beginning of the monitoring period. A subsequent increase of non-adherence probably reflects the waning of the adherence-enhancing effect of introducing EM to patients’ daily lives."

In the background section they overstate the utility of MEMS caps stating that they bring about “unprecedented precision” - while they are a great improvement over traditional methods such as self-report – the weaknesses inherent in MEMS caps have long been cited and the authors are urged to pull back from that statement.

We have changed the sentence to: "The introduction of electronic monitoring (EM) for assessing medication non-adherence has enabled researchers and clinicians to gather detailed data about medication-taking behavior."

The figures and tables the authors use are actually quite useful and do a nice job of organizing the various ways internal and external validity are impacted when using MEMS caps. The authors are also correct in noting that the skewed nature of adherence data requires reliance on non-parametric tests or other non-linear methods, a fact that is often not adequately addressed in the adherence literature.

Thank you for this positive comment.

While this paper is a useful methodological addition to the literature, there are several limitations to this study that the authors should strive to present with greater circumspection.

(1) The period of review was 3 months – a rather long window for retrospective recall. Typically adherence recall deteriorates after 7 days. Issues such as dropping a cap, removing pills etc. could easily have been forgotten, and this deserves mention by the authors.

Patients did not have to recall specific details, as they could note them on a form that accompanied the bottle (described in the text). To clarify this point, we have slightly changed the paragraph in which we explain the event-recording procedure: "Instructions stressed the need to match EM-
bottle openings with actual drug intakes and requested patients to
describe any deviations from this guideline on the form accompanying the
EM bottle. Examples of such guideline violations include inadvertently
opening the EM bottle when no medication was required, stopping EM
bottle use for a period, and removing pills prematurely. Upon completion
of the EM measurements, we integrated the resulting patient notes into
the uploaded EM data." The text continues, "At the end of the 3-month
EM period, we also used a structured interview to assess perceived
adherence to the EM instructions." During this structured interview,
patients were not asked to recall detailed information, rather they had to
give a general impression of their adherence to the system. As shown in
the HIV literature, adherence estimation over several weeks is more
reliable than over the past week [2]. Probably because patients give a
general impression. This probably gives an assessment which is nearer
to the truth than the sample of only 7 days gives.

(2) The collateral report scale is very subjective and poorly operationalized.
Additionally, it does not appear to have been subjected to any form of inter-rater
reliability analysis – thus what is being called good adherence by one health
provider could be rated by fair as another.

We used collateral report as it is used in practice: health care providers
who know a patient give their impression about their adherence. No
attempts were made to enhance assessment accuracy. This assessment
approach was validated against EM (Schäfer-Keller, Am J Trans, in
press). As expected, diagnostic values of collateral report were rather
poor. It shows that a typical collateral report measurement approach is
not a very sensitive method. Although we are aware that collateral report
was one of the weakest methods, we used for reasons of completeness it
in the testing of assumption 4.

(3) The sample as a whole had VERY high levels of adherence (nearly 96%) –
this is unusually high given that most studies of adherence report rates ranging
from 50-85%. At the very least, this speaks to the selective nature of this sample.
These rates of adherence are not representative and the authors should highlight
this fact.

It is true that with most medications, adherence measured with EM is
much lower than the high mean adherence found in this sample.
However, in transplantation, where only small deviations from the dosing
schedule are tolerated, these high numbers are perfectly normal.
Adherence rates recorded here, while exceptionally high for medication
users in general, fall within the normal range for kidney recipients and
other transplant populations [3]. To make this clear to the reader, we have
changed the initial sentence in the design section into: "Data for this
prospective cohort study came from the Supporting Medication
Adherence in Renal Transplantation (SMART) study, which focused on
prevalence and determinants of non-adherence." We refer to reference
21, where a discussion on the prevalences an its comparison to other
(kidney) transplant studies in this sample can be found.
The sample is not well characterized sociodemographically – the authors provide data on age, nationality and gender only. Further data on race/ethnicity, and educational level would have been useful to better understand the sample.

Table 2 gives a more detailed overview of the sample, including the educational level. To make this clearer, we changed the reference (Table 2) into a sentence: "Further details regarding the sample characteristics can be found in Table 2." The ethnicity question was not judged relevant in this European setting, although we have provided the reader with the distribution of nationalities within the sample.

The sample was very selective, and not representative of many of the more heterogeneous samples examined using MEMS caps, particularly HIV+ samples. The need for accurate adherence to highly active antiretroviral therapies used in HIV management made MEMS caps an oft-employed tool in HIV research. However, HIV infected samples in the US are often ethnically diverse, and are often economically disadvantaged. In addition, they often demonstrate much lower rates of adherence. The authors are urged to note the fact that their sample is not representative of the larger population of persons who are post-transplant, and may not reflect the types of samples on whom MEMS caps are employed – as such their findings, while interesting, may not necessarily be replicated in other samples.

Although we are convinced our sample is quite representative for the post-transplant population in Europe, it is clear that it is not representative for all populations in which MEMS is used. We are working on checking these assumptions in other populations as well. Assumption 3 has already been tested in a European HIV population [4], as mentioned in the discussion section. To highlight the need for more research in a broader range of populations, we have added the following sentence to the conclusion of this article, urging other research groups to investigate this issue: "Although the population used in this study is not representative of all populations in which MEMS is used, the approach we present could be used as a model for testing assumptions in other populations."

The study focuses on a relatively select group of patients – patients who are post-kidney transplant and are part of a healthcare system that provides far greater supports than observed in the US. The title of the paper may want to be expanded to clarify exactly what the paper is doing – for example “examining underlying assumptions regarding validity...” (as it is not initially clear what is meant by “underlying assumptions” and possibly acknowledging the type of sample (e.g. in post-transplant patients).

We have changed the title to: "Examining assumptions regarding valid electronic monitoring of medication therapy: development of a validation framework and its application on a European sample of kidney transplant patients."

References

John Urquhart

Before answering to the comments of the reviewer, we would like to stress that our research group does not want to challenge the value of EM as a most valuable method to assess medication adherence. On the contrary, we intend to promote EM as an indispensable measurement method. The importance of EM will continue to increase as the interest in the issues of non-adherence to medication regimens increases at several levels (clinicians, policy level, ...). The framework we propose for validating EM is intended to increase its value by refining EM's accuracy, to contribute to the knowledge of what is entailed by electronic measurement of non-adherence, and to map its strengths and weaknesses. Beyond these aims, we hope to offer other researchers a useful perspective on MEMS use.

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.

We are uncertain as to which part of the manuscript is in question here. We assume, however, that the reviewer questions our practice of supplementing missed registrations with patient notes – a standard practice in the HIV literature. While we are aware that this is not an ideal solution, these patient reports are recorded at the time of each event and do not require patients to recall an event which took place several weeks before. Even if less reliable than other sources, this kind of patient report may result in a more accurate measurement than simply not correcting phantom missing data.

We did have an interview at the end of the measurement period, in which we asked for rather general information (e.g. if they had used the monitor, for instance, during a three week period during which they were on vacation). We believe that, when a patient assures that during a holiday period the monitor was left at home while medication was taken from another source, it would be less useful to discard this patient information and rely blindly on the monitor’s records...
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.

The cited studies support the value of EM as a valid method to measure medication adherence. As with all measurement methods, though, validity testing is imperative and can be achieved via several approaches. In the 2005 study, the extent to which stressing the importance of adherence (p. 462) had an effect on adherence to the EM system remains unknown.

We acknowledge that these studies are an important part of the validation of EM and cited the most recent of them in the discussion section on page 13: "...it may be safely assumed that the non-adherence overestimation would have been even larger if we also omitted the patient-reported discrepancies. Nevertheless, a study supported by Aardex showed that uncorrected EM monitoring enables accurate estimation of patient’s internal drug exposure. How accurately patients adhered to the EM instructions is not known [23]."

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"
We assume that the reviewer refers to assumption 3, in which we examine the effect of EM itself on patient behavior. The interviewers were not blinded to the purpose of the study. This was an observational study, trying to capture phenomena as they occur in realistic situations. This study presents possible assumption violations as they would occur in a typical study. What we call for is that researchers examine their assumptions, whatever their data gathering approach. We showed that an intervention effect may occur and in that studies using a measurement period that is too small, can result in elevated adherence prevalences.

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.

The fact that the measurement itself influences the behavior to be measured does not necessarily in our view degrade MEMS data to the level of invalidity. Few adherence measurement methods are known to have no influence on adherence itself. Self-report does, blood assay does. On page 14, we therefore have added the sentences: “It has to be noted that all ethically permissible adherence assessment methods, electronic or otherwise, require the consent of the subject, and therefore influence the observed behavior. It is well-documented, for example, that blood assays lead to white-coat adherence. Likewise, data from the Swiss HIV Cohort Study, where non-adherence is measured every six months by self-report, show an increased adherence at each measurement time, possibly indicating an intervention effect [28].” Knowing that EM has this influence, and that its effect will wane over a manageable period, is an advantage of EM over other methods, because the intervention period is definable and can be omitted from the analysis if needed.

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.

This is a useful overview of possible reasons behind the observed increase of non-adherence. Although the first scenario is not relevant to our study, as we did not have a significant initial drop-out rate (the first person who stopping to use the EM system did so at day 41), the second and third are possible. As mentioned earlier, whatever the reason, any study of this type may elicit this pattern of behavior. After all, we also show that the initial adherence enhancement has only a minor effect on final period prevalences, which supports the value of EM. The intervention effect can be defined, and thus controlled for. Again, the goal of our paper was to provide a framework to validate EM. Each assessment method needs such a framework also to demonstrate its strengths, not only its weaknesses. Adopting the testing of the suggested assumptions of our framework is therefore an important tool to further substantiate the value of EM.

-------------------------------------------------------------------------------

Major Compulsory Revisions

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.

As we clarified above, it was not our aim to challenge EM data with self-reports, but to use self-reporting to interpret conflicting EM and clinical data (e.g., if we observed an EM-reported drug holiday of several weeks, which we wanted to interpret correctly). We also relied on self-reporting in the form of a kind of diary, which does not suffer from recall bias, and was therefore quite trustworthy.

To clarify the self-report issue, we have slightly changed the paragraph in which we explain the event-recording procedure: "Instructions stressed the need to match EM-bottle openings with actual drug intakes, and requested patients to describe any deviations from this guideline on the form accompanying the EM bottle. Examples of such guideline violations include inadvertently opening the
EM bottle when no medication was required, stopping EM bottle use for a period, and removing pills prematurely. Upon completion of the EM measurements, we integrated the resulting patient notes into the uploaded EM data." The text continues, "At the end of the 3-month EM period, we also used a structured interview to assess perceived adherence to the EM instructions."

We also refer to the 2005 paper by stating that "... a study supported by Aardex showed that uncorrected EM monitoring enables accurate estimation of patient’s drug exposure."

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

We would be pleased to answer this concern. However, it is not fully clear to us in what sense the reviewer is using the term “observer bias,” i.e., it is unclear how exactly our methods could have biased our results (i.e., what kind of possible effects regarding the four listed assumptions). What we can say is that the persons who enrolled the patients and interviewed them were not blinded to the purpose of the study, but this does not diverge from standard practice in many observational studies. Blood sampling and analysis were done independently.

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

We have changed the last paragraph of page 3 to refer to the website of the Aardex group: "Quality tests of the widely used Medication Event Monitoring System (MEMS®-6, Aardex Ltd.) revealed that the system performed well under normal or extreme laboratory conditions (if exposed to heat, cold, shocks, or water). A failure rate of below 0.5% is reported (www.aardexgroup.ch, accessed September 28, 2007). Reports of how MEMS performs in the field show a similar pattern."

We reworded the sentence on page 4, paragraph 1, to read, "EM failed to register only 2.5% of the generated events". This implies a less critical perspective than the original sentence, in which we said that only 97.5% of the monitors worked.

We have also adapted the discussion section on assumption 1 (page 13, paragraph 1): "We identified one EM device that had stopped recording cap openings during the study (0.4%). The failure rate is similar to that reported on the Aardex website (<0.5%), and confirms literature reports that EM devices used in studies can be damaged [5, 6]. Although our study evaluates the MEMS-5 monitors, and not the newer MEMS-6 monitors Aardex refers to, MEMS-6 differs from MEMS-5 mainly in its data upload technology. MEMS-6 is comparable to its predecessor regarding most other features (www.aardex.ch; accessed August 5, 2006). The result of our study can thus be considered representative for the system currently on the market. The existence of a non-registering cap shows that non-adherence overestimation is possible for a small number of patients who have damaged caps that remain undetected."
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.

We agree with the reviewer that each new version of the MEMS has improvements. We commented on this remark in the previous comment (see above). Part of the relevance of our study is the validity framework we offer, which is independent of the system we use.

------------------------------------------------------------------------------------------------------------------

Minor Essential Revisions

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

We have added the source of the monitors to the following sentence: "We used the Medication Event Monitoring System (MEMS-V TrackCap, Aardex, Ltd., Zug, Switzerland) to measure non-adherence to immunosuppressive medications."

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 they are uninterpretable.

To clarify the assay values, we have changed the sentence in the methods section on page 9, 1st paragraph to: "With regard to the blood assay, we considered one measurement moment, namely the patient’s drug trough levels at inclusion in the study (i.e., of cyclosporine, mycophenolate mofetil, tacrolimus, or sirolimus before the morning dose was ingested). We agree that the drawback of the testing of the fourth assumption is the large measurement error (and the resulting lack of power), and therefore urge other research groups to further investigate this assumption (p. 15).

------------------------------------------------------------------------------------------------------------------

Discretionary Revisions

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'.

We have corrected this mistake.