Reviewer's report

Title: The efficacy of the COACH prompting system to assist older adults with dementia through activities of daily living: An intervention study

Version: 1 Date: 4 July 2008

Reviewer: Marilynne A Hebert

Reviewer's report:

The following questions were considered:

1. Is the question posed by the authors well defined?

The authors pose a question around the efficacy of the COACH technology in assisting seniors with dementia complete activities of daily living. However, in the Abstract (page 1) the authors note: “The data was [should be were] analyzed to investigate the impact of COACH on the participants’ independence and caregiver burden as well as COACH’s overall performance.”

In an efficacy study, the investigators want to determine whether a technology works as it is expected to in a real world setting. They should keep this study purpose and small sample size in mind when stating their study conclusions. Further studies are needed to determine the bigger effects on independence and caregiver burden. “Independence” in this study is limited to hand washing, not the larger concept of independent living.

2. Are the methods appropriate and well described?

The authors have appropriately distinguished 4 levels of dementia as determined by a standard MMSE assessment. It is suggested from this early stratification that the authors expect differences in how the technology works, given the user’s level of cognitive functioning. However, this is not the case as they pool results from all participants. Suggest the author’s consider a design that includes comparison’s of functioning levels where this is a recruitment strategy and consider the groups separately in the analysis.

It is curious that the computer voice was a male actor, “to emulate a professional caregiver” (pg 10). I would think there were predominantly female caregivers in the facility and wondered whether introducing a male voice may have other consequences not related to the technology, e.g. seen as a voice of authority.

I wondered if participants had routinely washed their hands prior to their Alzheimer’s whether this would make the task of remembering how to do it easier. There is also a ceiling effect for 3 of the participants, who already did well with or without the technology. Perhaps this could be a screening tool on the next round of testing, i.e. on the first round of testing if scores are 4.5/5 or better, participants are excluded.
If participants don’t follow all the steps, they are recorded as being in error. However as one woman noted, she did not want to use soap, perhaps for other reasons than because she wasn’t following the rules, e.g. I am allergic to perfume and all the liquid soap in public facilities smells too strong for me to use! Given the caregiver is close, she may ask the participant why he/she doesn’t want to sue the soap.

3. Are the data sound?
Generally the data appear to be sound. However, it is of some concern that the inter-rater reliability is so low given the interventions and activities seem relatively simple and straightforward, e.g. counting the number of times the caregiver interacted with the participant (pg 14) with Kappa = 0.63. Similarly the 5 steps of handwashing are relatively straightforward and yet Kappa = 0.69. The authors suggest this is OK, given good agreement is between 0.61-0.80, however Kappa is on the low end of good and implications of this should be addressed with respect to quality of data.

4. Does the manuscript adhere to the relevant standards for reporting and data deposition?
All the figures need to be appropriately sized, labelled and numbered.

5. Are the discussion and conclusions well balanced and adequately supported by the data?
Given this is an "efficacy" study – asking how well the technology works - the authors draw unsupported conclusions that this technology will work for people with Alzheimer's and reduce caregiver burden. All they can really conclude is that the technology had partial success in this population with a hand washing task.

They identified a number of issues where the technology did not work as expected (the large number of false positives) that one would have expected to have been worked out prior to testing in the field.

I believe the authors need to re-think their rationale for calculating their statistics based on two groups of participants with differing levels of cognitive functioning: moderate and severe. It is clear from the results noted in the tables that the one participant who is at a low level of function has scores that are markedly different from the other 5 who are functioning at a moderately high level. Yet in the Abstract (p. 3) the authors note “level of dementia did not appear to be a robust indicator of hand washing performance or responsiveness to COACH...”

The opening line of the discussion is also misleading: “Overall, all the participants showed an overall improvement in the hand washing task when COACH was used...” (pg 18). The data suggests 3/6 had only slight (if any improvements); 2 participants had a ceiling effect (started off able to complete 4.5/5 steps of the task without assistance); 1 participant had difficulty completing the task with and without COACH.

Therefore, I do not believe the results can be reported as they stand. Given there
were no exclusion criteria based on MMSE the authors cannot exclude one category of data. However given there is only one participant in category 2, and these results are skewed, I suggest this participant be reported separately, i.e. not included in the averages and ranges. The results suggest that for one group the technology had some benefits but not for the other. However, with only one participant this is not strongly supported.

6. Are limitations of the work clearly stated?
Not that I could find.

7. Do the authors clearly acknowledge any work upon which they are building, both published and unpublished?
Yes, the authors note the previous work that has been completed using COACH. It would be helpful for the reader to have a brief summary (perhaps in a table) of the progress made in this work.

8. Do the title and abstract accurately convey what has been found?
No. The title refers to an intervention study and the study question refers to an efficacy study. The title refers to “activities of daily living” where the study is really only one activity – hand washing. They also call the study “clinical trials” in the paper (pg 29) which it definitely is not.

9. Is the writing acceptable?
The writing is generally acceptable (grammar, spelling, etc.), however prior to this being published, the following changes should be considered:

a. Review the manuscript for agreement of “the individual” (singular) and “their caregiver” (plural) which occurs frequently throughout the paper (e.g. pg 6).

b. The paper needs to be edited. Much of the information is presented in tables/figures, and then repeated in the text. Suggest reducing the text explanations by approximately 1/3.

c. Break the text into paragraphs that each contains one main idea. This makes it easier to read and follow the logic, which is more difficult to do when your paragraphs have many topics and are a page long.

d. Sometimes the referencing is correct (e.g. demonstrated in previous trials by Mihailidis et al. [23, 34]” – pg 14) and sometimes not (e.g. “technique known as flocking was developed by [24]” - pg 9).

Major Compulsory Revisions

1. Drawing conclusions based on study design
a. The authors pose a question around the efficacy of the COACH technology in assisting seniors with dementia complete activities of daily living (ADL). In an efficacy study, the investigators want to determine how well a technology works in a real world setting. They should keep this study purpose and small sample size in mind when stating their study conclusions. They are also studying “hand washing” as a representative exercise, not all ADL.
b. Given this is an "efficacy" study the authors are cautioned in drawing conclusions that cannot be supported with this study design and question. They suggest this technology will work for people with Alzheimer’s and reduce caregiver burden. But all they can really conclude is that the technology had partial success in this population with moderate cognitive functioning.

c. This is an efficacy study – not an intervention study (as noted in the title) and not a clinical trial (as noted in the text).

2. Data analysis

a. I do not believe the authors can accurately reflect the efficacy of COACH by combining the data from their two groups of participants with differing levels of cognitive functioning. It is clear from the results that the one participant who is at a lower level of functioning has scores that are not comparable to the other 5 who are functioning at a higher level.

b. Given there were no exclusion criteria based on MMSE, and the authors carried out the study with this group of participants, they cannot exclude one category of data. Given there is only one participant in category 2, and these results are skewed, this participant can be reported separately, i.e. not included in the averages and ranges. The results suggest that for one group the technology had some benefits but not for the other. However, with only one participant this is not strongly supported.

Minor Essential Revisions

The writing is generally acceptable, however prior to consideration for publication, the following suggestions for change should be seriously considered:

a. Review the manuscript for agreement of “the individual” (singular) and “their caregiver” (plural) which occurs frequently through-out the paper (e.g. pg 6)

b. The paper needs to be edited. Much of the information is presented in tables/figures, and then repeated in the text. Suggest reducing the text explanations by approximately 1/3.

c. Break the text into paragraphs that each contains one main idea. This makes it easier to read and follow the logic, which is more difficult to do when your paragraphs have many topics and are a page long.

d. Sometimes the referencing is correct (e.g. demonstrated in previous trials by Mihailidis et al. [23, 34]” – pg 14) and sometimes not (e.g. “technique known as flocking was developed by [24]” - pg 9).

The figures contribute to understanding the study. All the figures need to be appropriately sized, labelled and numbered.

Discretionary Revisions

The authors identified a number of issues where the technology did not work as expected (the large number of false positives). I would have thought testing in the lab would have highlighted these and worked them out prior to testing in the field.
This may be an area to explore in the discussion in why this result was or was not surprising to them.

**Level of interest:** An article whose findings are important to those with closely related research interests

**Quality of written English:** Needs some language corrections before being published

**Statistical review:** No, the manuscript does not need to be seen by a statistician.

**Declaration of competing interests:**

I declare that I have no competing interests.