Author's response to reviews

Title: Unsedated Peroral Wireless pH Capsule Placement vs. Standard pH Testing: A Randomized Study and Cost Analysis

Authors:

Christopher N. Andrews (motility@ucalgary.ca)
Daniel C. Sadowski (dan.sadowski@ualberta.ca)
Adriana Lazarescu (adriana.lazarescu@ualberta.ca)
Chad Williams (ciwillia@me.com)
Emil Neshev (emilnesh@hotmail.com)
Martin Storr (mstorr@ucalgary.ca)
Flora Au (fau@ucalgary.ca)
Steven Heitman (Steven.Heitman@albertahealthservices.ca)

Version: 3 Date: 24 April 2012

Author's response to reviews: see over
To the Editor and Reviewers:

Thank you most sincerely for your kind review of our work. We present a point-by-point response to the issues raised. Please do not hesitate to contact us if you require further information.

Sincerely,

Christopher Andrews, MD, on behalf of the authors.

Reviewer 1

Major Compulsory Revisions

1. All of the inputs that into the costing model need to be included in Table 4 for better transparency. This includes resource use, unit costs and probabilities used in the calculations. It is a simple model and therefore should be easy to replicate.

We agree with this comment and have included all of the model inputs into Table 4 as requested.

2. The authors state that the cost analysis is from a public payer perspective and thus exclude costs resulting from lost productivity (assumptions, paragraph 2) but then calculate the indirect costs (which are productivity costs). It should state “societal” perspective as they have included costs borne to the patient.

We took the perspective of the publicly funded health care system in the Reference Case as recommended by The Canadian Guidelines for the Economic Evaluation of Health Technologies (ref: *Guidelines for the Economic Evaluation of Health Technologies: Canada [3rd Edition]*. Ottawa: Canadian Agency for Drugs and Technologies in Health; 2006.). Patient and family/caregiver time costs spent for travel and receiving treatment should be considered in this perspective in contrast to a public payer perspective where these costs are not included. Time spent traveling and receiving care has an associated opportunity cost, which we valued using the average hourly wage rate for Albertans. Our group has used this approach previously (please see: Heitman et al. J Am Coll Radiol 2010; Heitman et al. Clin Gastroenterol Hepatol 2008; Heitman et al. PLoS Med 2010). In contrast, we did not consider lost productivity costs arising from reduced working capacity during illness or costs to employers to hire/train replacement workers, etc. as these would be only relevant to a societal perspective.
Furthermore, it is unlikely that such productivity costs would be particularly relevant given the very short (one day or less) absences that accompanied the esophageal pH testing in our model.

Nevertheless, we acknowledge that not all jurisdictions consider time costs in the Reference Case and as such, we have modeled a scenario where these costs are excluded (Please see the last sentence of the results section and the 4th paragraph of the discussion on page 16. This has also been included in Table 5). This addition helps to clarify the viewpoint that Bravo “could” be something health care decision makers might fund when the opportunity time costs of patients are considered in addition to the fact that patients prefer WC over SC, even though it is more expensive.

3. In the assumptions section (paragraph 2) is says that “93% of the WC group would have worked during the testing period (see results)”, however this statistic is not in any of the results tables. Furthermore, it is not clear in the cost analysis what percentage was used in the EGD+WC group. As per comment 1, this should all be in a table of model input parameters.

This information is given in the results section in the text but not the tables to avoid duplication. As per the results section, for those in the EGD + WC strategy we assumed that patients would not return to work until the next day given the sedation (thus 100% would not have worked).

4. The primary aim was to determine whether the placement of WC is as effective and better tolerated than SC. The paper therefore assumes that in measuring pH levels, WC and SC are equal in effectiveness. The report discusses patient discomfort in the conclusion and abstract but does not discuss failure rates which a key form of effectiveness for this product. Furthermore, the study was only powered for VAS discomfort scores and not powered to detect difference is technical success. Also the VAS for chest discomfort in the WC was statistically significantly higher than the SC group but is also not discussed.

We did assume the both SC and WC are equal in detecting abnormal pH exposure in the esophagus, in keeping with most of the published literature on the subject. Each modality can provide collateral information about reflux events (eg SC can have 2 probes on the same catheter, one 10cm higher and thus give information of proximal reflux events; WC can be measured for periods up to 96 hours or longer in some reports), but neither modality has been proven to be better or worse than the other on diagnostic outcomes. We certainly agree with the reviewer that failure rates are a key effectiveness indicator in this situation. We have made the failure rate explicit in the abstract. The failure rate is discussed in detail in the 2nd para of the discussion, but we have reframed this paragraph to better emphasize this point.

Regarding sample size, our primary outcome in the clinical trial component of this study was overall tolerability of the peroral approach since this aspect has not been studied in a randomized fashion. SC was not significantly better than WC in this regard. Technical success rates have already been described in large series, and since in most of those series it was similar to that of SC, sample size requirements for a study would have been unreasonably large. Finally, although our failure rate was on the higher end of the reported spectrum, we compensated for this in the cost analysis by subjecting it to sensitivity analysis when brought down to reported rates from other studies. This is further discussed in the Discussion.
We have mentioned the chest discomfort with WC in the results section. We have also commented on this in the discussion.

Minor Essential Revisions

1. More information on the wage rate used would be helpful to understand the indirect costs applied in the model.

We have included this information as requested. Please see page 10.

2. The model assumes that 100% would not work in the SC group, however the mean VAS score of ‘ability to do usual activities’ was 75. Please explain the discrepancy.

The ability to do usual activities was focused on activities of daily living aside from eating and drinking, like moving, exercising, going to the washroom, getting dressed etc. In this respect patients were not significantly inhibited in either group, although SC patients did have more interruption. The ability to work question incorporates not only the physical aspects of working (ie lifting or manual activities in those who do them) but also the social element and embarrassment of being at work with a catheter in the nose, which clearly has a significant impact on those who may interact with customers, clients, or colleagues on a face-to-face basis. Thus almost no patients in the SC group would have chosen to work during the test, vs all patients in the WC group, where the external apparatus is much more discreet.

3. Only 43 patients were asked about whether they would work, but does 43 patients have sufficient power to detect a difference between the groups? Furthermore, the question on whether they would work is hypothetical and is questionable whether this would reflect actual experience.

Given the wide disparity between the groups’ responses and the strongly significant p value of their comparison, a Type II error is very unlikely. We do agree with the reviewer that the question of whether they would work is hypothetical and was asked after the fact. However, we were limited in this regard due to the study design, since patients were not randomized until the day of the test. Thus most people planned to have the day off in advance.

4. It is not appropriate to only drop the failure rate of ESM+WC in the sensitivity analysis, given that the failure rates between the WC and SC groups were not statistically significant.

This is a valid point, however we used both the literature and our study results to inform the cost analysis. We felt that raising the failure rate higher in the sensitivity analysis was not required since the 12% failure rate in our study was higher than almost all studies reporting this metric, and substantially higher than larger recent series. We suspect that part of this is due to relative lack of experience with the technology (ie “learning curve”), and thus may be an overestimation of what would be expected in real clinical practice with experienced operators.

Discretionary Revisions
1. In the introduction (paragraph 2), it may be more appropriate to discuss the duration of the testing period before reporting on how it affects daily living, as well as how long these effects last.

This information has been added to the paragraph.

2. In the introduction (paragraph 4), the authors mention “Endoscopy also increases patients risk...”. It is unclear what this risk is. This same point is in the discussion but the reader has no indication what risks are associated with endoscopy.

Endoscopy has a low but measurable risk of perforation, bleeding, or adverse reactions to the sedation used (such as hypoxia, hypotension, or aspiration of gastric contents). This information has been added to the introduction.

3. In the introduction (paragraph 5) it states that “data from clinical study informed the cost-analysis”, however data was also used from the literature for the EGD+WC group.

This sentence was clarified.

4. As a discussion point, the reasons for not using a validated quality of life instrument, such as the SF-36.

At the time of this study there was no validated utility tool that was specific for this intervention, in both the type of discomfort experienced, as well as the brief duration of that discomfort (ie 24-48h) which then has no adverse residual effects. We thus opted to use VAS scales.

Reviewer 2

Major compulsory revisions

1. Discussion, first paragraph. The authors conclude that their study “confirms that WC insertion is better tolerated than SC both overall and during placement”. However, the study design does not allow the authors to make this conclusion as this would require that all subjects underwent both tests in a cross-over study design. The conclusion needs to be changed.

A cross-over study design would theoretically have been the optimal method to compare both WC and SC strategies in the same group of patients, but would have been logistically more difficult and very difficult to recruit into. Instead we prospectively randomized patients to either modality. There was no significant difference between the groups in their baseline or demographic variables. Although there may have been some differences between the groups in other unmeasured variables, if this occurred it would have been by chance and not due to systemic bias. Therefore we feel justified in making the conclusion that WC insertion is better tolerated than SC in this study, based on the randomized controlled trial paradigm.
2. Statistics: Data obtained from VAS scales are usually not normally distributed. The authors chose to assess differences in discomfort levels using means and the Student t-test. If the data was normally distributed this needs to be stated and if not, non-parametric statistical analysis should be used.

VAS data were similarly distributed in each group and reasonably normally distributed within groups. We therefore compared them using t-test. This information is added to methods.

3. Most, if not all patients who undergo esophageal pH monitoring also undergo an upper GI endoscopy. How many patients in the ESM+SC and ESM+WC groups had an endoscopy shortly before or after the pH-study. As it is the total cost for investigating symptoms suggestive of GERD that is of interest, the cost of this additional endoscopy needs to be included in the analysis.

This point is well taken. However, we would venture that every patient gets endoscopy before pH testing, including those in the EGD+WC group. The effects would thus cancel each other, and thus we did not include this in the analysis. From the literature plus interactions with physicians who use WC regularly, it appears that even most patients who have EGD+WC have already been fully investigated with previous EGD or other imaging, and that the EGD is solely used to place the WC. Since the EGD+WC group is already dominated by non-endoscopic pH methods in our cost analysis, it would remain so if even a portion of EGD+WC patients had 2 endoscopies.

4. The analysis of the costs of the different strategies is based on more or less well-documented assumptions that are regarded as facts in the statistical analysis. This is a limitation of the study that needs to be mentioned in the discussion.

Limitation added to discussion.

5. Studies on adverse symptoms of pH monitoring has previously been performed by other groups. The observations of these studies should be mentioned and compared with the result of the present study.

Discussion modified to reinforce this point.

6. Discussion, second paragraph. The two last sentences needs to be deleted as efforts to validate the accuracy of the conversion factor was not made in the present study.

A disclaimer to this effect was added. However, we have left our comment about the conversion factor as written as since we are the first group outside the original investigators to describe our experience with it, and we feel it is a useful confirmation for clinical practitioners who may wish to try it.
Minor Essential Revisions

1. In the first paragraph of the Introduction it is stated that ambulatory esophageal pH monitoring is the gold standard for establishing the diagnosis of GERD. However, the diagnostic accuracy of esophageal pH monitoring is relatively low and therefore no gold standard.

We agree that the diagnostic accuracy of ambulatory pH monitoring is low compared to other diagnostic tests, however it is much better than, for example, barium esophagram or PPI trial in this disorder. It is generally accepted as an established standard. We have changed “gold” to “established”.

Editorial Requirements:

- kindly change "Introduction" Section into "Background"

Done.

- Structure: Please check the instructions for authors on the journal website to ensure that your manuscript follows the correct structure for this journal and article type.

<http://www.biomedcentral.com/bmcgastroenterol/authors/instructions/researcharticle>

Structure was reviewed.

- Tables: Please ensure that the order in which your tables are cited is the same as the order in which they are provided. Every table must be cited in the text, using Arabic numerals. Please do not use ranges when listing tables. Tables must not be subdivided, or contain tables within tables. Please note that we are unable to display vertical lines or text within tables, no display merged cells: please re-layout your table without these elements. Tables should be formatted using the Table tool in your word processor. Please ensure the table title is above the table and the legend is below the table. For more information, see the instructions for authors on the journal website.

Tables reformatted as requested.

- Conclusions

This should state clearly the main conclusions of the research and give a clear explanation of their importance and relevance. Summary illustrations may be included.

Conclusions heading was added after discussion.