PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Validation of Days at Home as an Outcome Measure after Surgery: a prospective cohort study in Australia
AUTHORS	Myles, Paul; Shulman, Mark; Heritier, Stephane; Wallace, Sophie; McIlroy, David; McCluskey, Stuart; Sillar, Isabella; Forbes, Andrew

VERSION 1 - REVIEW

REVIEWER	Robert Sneyd Plymouth University Peninsula Schools of Medicine and Dentistry
REVIEW RETURNED	11-Jan-2017

GENERAL COMMENTS	Title: Days at Home as an Outcome Measure after Surgery Manuscript ID: bmjopen-2017-015828comments for the authors The manuscript reports a simple and hopefully reproducible analysis of data from a group of past studies. The authors test the hypothesis that this easy to compute measure would correlate with complications, morbidity and prove valuable to patients. The work is clearly described and appears to have been conducted carefully. 1. You remind us (page 4, line 55) that “there is a growing acceptance that outcome measures used in clinical trials should be determined in partnership with patients…” Yet (page 7, line 7) “we did not involve patients or their carers in the design or conduct of this study…”. Why not? I am baffled! 2. Page seven, line 55, you report (understandably) that “we were unable to reliably collect secondary length of stay for rehabilitation facilities”. You then went on to assume five days. Might this be an area for some kind of sensitivity analysis? 3. Your methodology is intelligible and the analysis seems sensible. 4. Although you have (in the method section) made it reasonably clear how you calculated your chosen measure DAH30, if your manuscript is published it is likely that it will be cited in subsequent papers reporting DAH30. With that in mind I wonder if a more comprehensive Appendix “method of calculation of DAH30” with some notes on marginal considerations and potential pitfalls might be a good idea. This could help to ensure that anybody used it subsequently reported it in exactly the same way thereby achieving consistency for comparison of reports from different centres/groups.

REVIEWER	Rupert Pearse Queen Mary University of London & Barts Health NHS Trust, UK RP holds research grants, and has given lectures and/or performed consultancy work for Nestle Health Sciences, BBraun, Medtronic, and Edwards Lifesciences, and is a member of the Associate editorial board of the British Journal of Anaesthesia.
REVIEW RETURNED	15-Jan-2017

GENERAL COMMENTS	There is an ongoing international discussion around the optimal choice of patient centered outcome measures for clinical trials of interventions designed to improve patient outcomes after major surgery. This is becoming a very active research area with several large groups around the world now designing and leading high quality large multi-centre RCTs in perioperative medicine. In this context, a methodological paper describing a new clinical outcome measure is topical and potentially important. In this manuscript the authors describe the validation of a new metric, Days at Home within 30 days of surgery (DAH30) as a clinical outcome for major trials. By combining the effects of death, complications, delayed recovery and hospital readmission, the authors argue this outcome is an 'ideal patient-centred outcome measure for perioperative clinical trials'. The author group is extremely strong and includes some leading international experts in clinical trials in this field. The manuscript is well prepared, and the statistical analysis seems to me robust and appropriate. The patient sample on which the analyses are based seems generalizable to the international population of in-patient surgery, albeit derived from a single tertiary referral hospital in Australia. I have very few comments to make on these aspects. However, I must confess I was surprised to see such a seasoned group of experts argue so very strongly for the perceived merits of an outcome measure largely determined by the primary hospital stay after surgery. Duration of hospital stay is dominated by hospital process and structures which may promote or delay patient discharge quite independently from any complication that may have occurred. Factors including, but by no means confined to, availability of senior doctors, pharmacy, weekend availability of staff, patient transport, etc. all combine to speed up or slow down a patient's discharge from hospital. The authors seem to ignore these issues completely. I quite agree that complications, death, need for discharge to a rehab facility, and hospital readmission will be the strongest determinants of the metric, and of course that these are important to patients, but there must surely be a balanced discussion which includes the obvious limitations of DAH30 as well? The manuscript mentions reservations (which I share) about the use of composite outcome measures (e.g. postoperative pulmonary complications), and yet DAH30 is similarly affected by a wide range of factors leading to the same kind of weaknesses. If such key weaknesses are ignored, we cannot expect less expert trialists to use the metric appropriately, leading to further poor trials – a situation I know we are all keen to avoid! This is a good paper describing a worthwhile piece of methodological research. It is likely to be well cited. The manuscript needs to be revised extensively to be much more balanced in how it promotes the use of DAH30. The abstract must include mention of the limitations, in particular due to process of care, so that we can reliably expert readers to have a clear and balanced understanding of the metric from the outset. The main text discussion should then explain these limitations in detail, including both the strengths and the weaknesses of the measure rather than the strengths of the measure and the weaknesses of the validation study as is currently the case. The term 'ideal' should be replaced with something like 'useful' or 'pragmatic'. I also note there is no mention of the single-centre nature of the data. This is not mentioned in the abstract, and it is only briefly alluded to in the methods. There is no discussion of

	this as a limitation.

REVIEWER	Alexandre Stephens Public Health Observatory, Sydney Local Health District, Australia
REVIEW RETURNED	05-Mar-2017

GENERAL COMMENTS	The authors present a study on the validity and utility of days spent at home within 30 days of surgery as a patient centred outcome. Patients included in the study were from 7 recently completed clinical trials, and consisted of 2109 eligible patients. The authors proceeded to evaluate the relationship between DAH30 and a number of factors related to patient health, age, complications and duration of surgery in order to assess the clinical validity of DAH30. Quantile regression (median and 3rd quartile) was the main statistical analysis technique applied. This is an interesting study and the discussion nicely describes the potential benefits of the DAH30 outcome measure. However, there are a few issues that require addressing. Here are my specific comments: **METHODS** In the "Risk Factors and Outcomes" section, it's not entirely clear how a patient ends up in rehab. For example, can a patient go to rehab as part of their hospital stay? Or directly from discharge? And how does this influence hospital length of stay? Is it possible for a patient to have a statistical discharge and then enter rehab as part of their one hospital stay and thus rehab is included in the calculation of their hospital length of stay? My main concern is that if rehabilitation occurs as part of the initial hospital stay, then the hospital length of stay might be overestimated if the rehab part cannot be taken into account. I understand that it was not possible to obtain the length of stay for rehabilitation care which is a major limitation of the DAH30 measure in the study. The authors perform sensitivity analyses assuming 5 days of length of stay for patients attending rehabilitation. This maybe an entirely fair value, but could the authors elaborate on how the came to select this particular value? In the "Statistical Analysis" section, could the authors please write/define ASA and LTR in full before abbreviating? Please also check the second to last paragraph of the "Statistical Analysis" section as there appears to be a word or two missing Are there any admissions following surgery that would be for some type of follow-up care, such as adjustments or subsequent treatment, that is expected or necessary? If so, were these taken into account (as in not included in the calculation of DAH30)? And do they vary by type of surgery? Not sure if I missed it, but I couldn't find the STROBE checklist **RESULTS** The study period spans a considerable amount of time (March 2006

	to September 2016). Did the authors account for any temporal effects on the outcome? Could year be assessed as a covariate in the regression models? Would there have been any changes in care over the study period that would support attempting to account for changes in DAH30? In paragraph 4, sentence 2, the authors write that associations remained after adjustment for "**all of these covariates**". However, Table 2 indicates that adjustment only included age, sex, ASA and surgery time. Smoking, diabetes and heart failure were "not done". Please clarify. Paragraph 5 seems to be a bit of a self-fulfilling prophecy as readmission is involved in the calculation of the outcome. Just seems strange to say hospital readmission was also a factor. Not really sure how it couldn't be. In the Tables 1 and 3 (and also in the corresponding supplementary tables), it's not clear whether the presented estimates are adjusted or unadjusted (raw)? If they are not adjusted, might it be worth showing the adjusted estimates? Table S4 appears to be a replica of Table 1 For completeness, would it not be worth showing the Q3 regression results for type of surgery?

REVIEWER	Hideto TAKAHASHI Office of Information Management and Statistics, Radiation Medical Science Center for the Fukushima Health Management Survey, Faculty of Medicine, Fukushima Medical University
REVIEW RETURNED	10-Mar-2017

GENERAL COMMENTS	Major Comments: (1)The authors applied quantile regression for "day at home up to 30 days (DAH30)", challengingly. I respect this intension, but it needs more statistical information to show in the article for readers. 1)The authors recommended third quartile (75th percentile) of DAH30 as a general index. If so, the appropriateness of the 75th percentile (not 50th, 60th, 70th,or 80th percentile) was unclear. They should show the reason, logically, why no other quantiles but the 75th percentile was appropriate as a general index,. It is quite important. 2)The authors should show the goodness of fit of the model (AIC,BIC, deviance, or their corresponding value), and check the appropriateness of the model. (2)Quantile regression is not popular, generally. The authors should introduce the advantage and disadvantage of quantile regression, simply. (3)It does not always mean DAH30 is a valid index that the authors obtained their satisfactory results using the DAH30 with a minor

	analysis. It would relate the “reliability and validity” in general sense. The reliability here closely relate to generalizability of the features of DAH30, and the validity here relate to “construct validity”. The authors should show them. Minor Comments: We can understand the index Q3 was the third quartile because the definition was written in the tables. The authors should describe its definition at least in the main manuscript.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1 (Robert Sneyd)

1. You remind us (page 4, line 55) that “there is a growing acceptance that outcome measures used in clinical trials should be determined in partnership with patients…” Yet (page 7, line 7) “we did not involve patients or their carers in the design or conduct of this study…”. Why not? I am baffled!
Our Response: The components that make up DAH30 (length of stay, re-admission, discharge destination, and early deaths after surgery) have been considered extensively by others and are clearly highly valued by patients and their families, and as we state in our manuscript, we have received direct feedback about the importance of these aspects from patients booked for surgery. This detail was explained on the next page (3rd paragraph).

2. Page seven, line 55, you report (understandably) that “we were unable to reliably collect secondary length of stay for rehabilitation facilities”. You then went on to assume five days. Might this be an area for some kind of sensitivity analysis?
Our Response: We accept this suggestion, and have adding in a second analysis assuming rehabilitation LOS was 14 days. This extra analysis is explained in the Methods section (red text, page 8 and reported in the Results (red text) and in the revised Supplementary file (Tables S7-S9). The results do not change our conclusions.

3. Your methodology is intelligible and the analysis seems sensible.
Our Response: Thank you.

4. Although you have (in the method section) made it reasonably clear how you calculated your chosen measure DAH30, if your manuscript is published it is likely that it will be cited in subsequent papers reporting DAH30. With that in mind I wonder if a more comprehensive Appendix “method of calculation of DAH30” with some notes on marginal considerations and potential pitfalls might be a good idea. This could help to ensure that anybody used it subsequently reported it in exactly the same way thereby achieving consistency for comparison of reports from different centres/groups.
Our Response: Thank you –we appreciate this suggestion and have included this additional information in the revised Supplementary file (red text, page 2).

Reviewer: 2 (Rupert Pearse)

There is an ongoing international discussion around the optimal choice of patient centered outcome measures for clinical trials of interventions designed to improve patient outcomes after major surgery. This is becoming a very active research area with several large groups around the world now designing and leading high quality large multi-centre RCTs in perioperative medicine. In this context, a methodological paper describing a new clinical outcome measure is topical and potentially important. In this manuscript the authors describe the validation of a new metric, Days at Home within 30 days of surgery (DAH30) as a clinical outcome for major trials. By combining the effects of death, complications, delayed recovery and hospital readmission, the authors argue this outcome is an 'ideal

patient-centred outcome measure for perioperative clinical trials'.

The author group is extremely strong and includes some leading international experts in clinical trials in this field. The manuscript is well prepared, and the statistical analysis seems to me robust and appropriate. The patient sample on which the analyses are based seems generalizable to the international population of in-patient surgery, albeit derived from a single tertiary referral hospital in Australia. I have very few comments to make on these aspects. However, I must confess I was surprised to see such a seasoned group of experts argue so very strongly for the perceived merits of an outcome measure largely determined by the primary hospital stay after surgery. Duration of hospital stay is dominated by hospital process and structures which may promote or delay patient discharge quite independently from any complication that may have occurred. Factors including, but by no means confined to, availability of senior doctors, pharmacy, weekend availability of staff, patient transport, etc. all combine to speed up or slow down a patient's discharge from hospital. The authors seem to ignore these issues completely. I quite agree that complications, death, need for discharge to a rehab facility, and hospital readmission will be the strongest determinants of the metric, and of course that these are important to patients, but there must surely be a balanced discussion which includes the obvious limitations of DAH30 as well? The manuscript mentions reservations (which I share) about the use of composite outcome measures (e.g. postoperative pulmonary complications), and yet DAH30 is similarly affected by a wide range of factors leading to the same kind of weaknesses. If such key weaknesses are ignored, we cannot expect less expert trialists to use the metric appropriately, leading to further poor trials – a situation I know we are all keen to avoid!
Our Response: We are offering a new and practical patient-centred outcome measure. It is not expected to exist in isolation, but be supplemented by more traditional outcomes (eg. complication rates, survival). Although there are frequent concerns raised about using hospital length of stay after surgery as an outcome measure, largely because of the reason outline by this reviewer, it is mostly a source of variation (background noise) in clinical trials and not biased. Very few hospitals have the luxury of extending a patient's stay in hospital for non-clinical reasons. Hospital stay is a reasonable surrogate for quality and speed of recovery after surgery, and it has marked resource/cost implications. Most patients want to go home as soon as possible – it is a desired outcome in and of itself (see page 5, 3rd paragraph). This is one of the reasons that the US-NSQIP organisation record and report such data; it is an outcome variable calculated by the American College of Surgeons' Surgical Risk Calculator. Importantly, unlike many other composite measures, DAH30 is focussed on the actual clinical implications of any complication/adverse event – it will not be affected if the event was minor or transient. Nevertheless, we do accept the points made by this reviewer and have modified our manuscript accordingly (red text, Discussion page 12).

This is a good paper describing a worthwhile piece of methodological research. It is likely to be well cited. The manuscript needs to be revised extensively to be much more balanced in how it promotes the use of DAH30. The abstract must include mention of the limitations, in particular due to process of care, so that we can reliably expert readers to have a clear and balanced understanding of the metric from the outset. The main text discussion should then explain these limitations in detail, including both the strengths and the weaknesses of the measure rather than the strengths of the measure and the weaknesses of the validation study as is currently the case. The term 'ideal' should be replaced with something like 'useful' or 'pragmatic'. I also note there is no mention of the single-centre nature of the data. This is not mentioned in the abstract, and it is only briefly alluded to in the methods. There is no discussion of this as a limitation.
Our Response: We accept this and have improved the balance, giving greater considering to potential weaknesses. We have opted for "pragmatic" in the Abstract. The single-centre design has been clarified, and included as a limitation in the Discussion (we now clearly outline six limitations in the penultimate section of the Discussion).

Reviewer: 3 (Alexandre Stephens)

METHODS

In the “Risk Factors and Outcomes” section, it’s not entirely clear how a patient ends up in rehab. For example, can a patient go to rehab as part of their hospital stay? Or directly from discharge? And how does this influence hospital length of stay? Is it possible for a patient to have a statistical discharge and then enter rehab as part of their one hospital stay and thus rehab is included in the calculation of their hospital length of stay? My main concern is that if rehabilitation occurs as part of the initial hospital stay, then the hospital length of stay might be overestimated if the rehab part cannot be taken into account.

Our Response: Further details are now provided (red text, page 8). We recorded acute hospital (not rehabilitation centre/hospital) data in our study.

I understand that it was not possible to obtain the length of stay for rehabilitation care which is a major limitation of the DAH30 measure in the study. The authors perform sensitivity analyses assuming 5 days of length of stay for patients attending rehabilitation. This maybe an entirely fair value, but could the authors elaborate on how the came to select this particular value?

Our Response: We agree that accurate rehabilitation data would be ideal, and we emphasise this limitation in our Discussion. Five days was common for our hip/knee arthroplasty patients. As per Reviewer 1, we have added in a second sensitivity analysis assuming 14 days – the results and interpretation are largely unaffected. See new Tables S7-S9 in the Supplement.

In the “Statistical Analysis” section, could the authors please write/define ASA and LTR in full before abbreviating?

Our Response: Yes, now clarified.

Please also check the second to last paragraph of the “Statistical Analysis” section as there appears to be a word or two missing

Our Response: Corrected.

Are there any admissions following surgery that would be for some type of follow-up care, such as adjustments or subsequent treatment, that is expected or necessary? If so, were these taken into account (as in not included in the calculation of DAH30)? And do they vary by type of surgery?

Our Response: We did not collect such data, but in any case this is quite rare within 30 days of surgery. We would counsel to include such days in the calculation of DAH30 (we have added this information to the Supplementary file, as suggested by Reviewer 1).

Not sure if I missed it, but I couldn’t find the STROBE checklist

Our Response: It has been resent.

RESULTS

The study period spans a considerable amount of time (March 2006 to September 2016). Did the authors account for any temporal effects on the outcome? Could year be assessed as a covariate in the regression models? Would there have been any changes in care over the study period that would support attempting to account for changes in DAH30?

Our Response: The studies overlapped and it is very unlikely that this could affect the results or interpretation.

In paragraph 4, sentence 2, the authors write that associations remained after adjustment for “all of these covariates”. However, Table 2 indicates that adjustment only included age, sex, ASA and surgery time. Smoking, diabetes and heart failure were “not done”. Please clarify.

Our Response: Yes, after adjustment for age, sex, ASA and surgery time. A global test indicated that the other predictors (smoking, diabetes, heart failure) did not bring anything new to the model – see also response to reviewer 4 (point 2)

Paragraph 5 seems to be a bit of a self-fulfilling prophecy as readmission is involved in the calculation of the outcome. Just seems strange to say hospital readmission was also a factor. Not really sure how it couldn't be.
Our Response: Yes, we accept this but we included this information to demonstrate the relative effect (6 days) for those patients re-admitted to hospital. That is, the impact of re-admission. This quantifiable information is typically missing from perioperative studies, with it being dichotomised to yes or no.

In the Tables 1 and 3 (and also in the corresponding supplementary tables), it's not clear whether the presented estimates are adjusted or unadjusted (raw)? If they are not adjusted, might it be worth showing the adjusted estimates?
Our Response: We intentionally chose to report raw/unadjusted data for these tables because the estimates are useful to clinicians, and for some categories there were small numbers (making adjustment unreliable). Importantly, this information tells us how many days at home for each type of surgery, or following each type of complication. We then went on to do the adjusted analyses when evaluating the effect of complications, to report the more-specific effect of any single variable – we did this in the narrative of the Results section.

Table S4 appears to be a replica of Table 1
Our Response: We tables are very similar because the median values were unaffected (only 245 of 2,109 patients went to rehab.), but the 95% CIs may differ.

For completeness, would it not be worth showing the Q3 regression results for type of surgery?
Our Response: We are loathe to do this because of the small numbers in some categories – the estimates will be unstable, and these data do not add any value to the paper.

Reviewer: 4 (Hideto TAKAHASHI)

Major Comments:
(1)The authors applied quantile regression for "day at home up to 30 days (DAH30)", challengingly. I respect this intension, but it needs more statistical information to show in the article for readers.
Our Response: See below, points 2 & 3.

1)The authors recommended third quartile (75th percentile) of DAH30 as a general index. If so, the appropriateness of the 75th percentile (not 50th, 60th, 70th,or 80th percentile) was unclear. They should show the reason, logically, why no other quantiles but the 75th percentile was appropriate as a general index,. It is quite important.
Our Response: Yes, we agree - the attraction of quantile regression is that any quantile of the distribution can, in principle, be explained. In a clinical trial, it would be appropriate to prespecify which quantile(s) will be considered. In a post-hoc analysis like this one, we chose the third quartile (75th percentile) as it is a common percentile close to the mode of the distribution (i.e. where most of patient DAHs are). The impact of the percentile choice has relatively little impact on the associations in a reasonable range (50th-75th) – we have added some plots demonstrating this in the revised Supplement whereby the coefficients of each covariate are shown to be reasonably stable over this range. See also our response to Point 3) where we indicate the corresponding changes to the statistical analysis section.

2)The authors should show the goodness of fit of the model (AIC,BIC, deviance, or their

corresponding value), and check the appropriateness of the model.
Our Response: Koenker and Machado (1999) – reference [36] – introduced a quasi-likelihood ratio test to assess goodness of fit in quantile regression. This test is a global test that can be used to compare embedded models. The multivariate model we used (10-year age categories, sex, ASA, surgery time (< 2h, 2.0 - 2.99, 3.0 -3.99, ≥4.0) was just as good as a larger model that included all the covariates + heart failure, diabetes and smoking (P-value=0.36). To our knowledge there is no Akaike criterion, only a Bayesian information citerion that has just been proposed (Dunder et al. Communications & Statistics 2016) but never been used.

The part in italics has been added to the corresponding section in the Statistical Methods section: The adjusted models included age by 10-year categories, sex, ASA, surgery time (< 2h, 2.0 - 2.99, 3.0 - 3.99, ≥4.0). A goodness of fit test[39] comparing this model to the full model including the same predictors plus smoking, heart failure and diabetes was not any better (P=0.36)

Linearity was not an issue as predictors were either binary (gender, heart failure, diabetes) or categorised (age, surgery time, ASA).

(2)Quantile regression is not popular, generally. The authors should introduce the advantage and disadvantage of quantile regression, simply.
Our Response: Quantile regression is well established in econometrics and has been commonly used since its inception in 1978. It has been rarely used in biostatistics but it is nevertheless a tool biostatisticians should know about and use more, especially in situations like this. The motivation and advantages were introduced in the Statistical Analysis Section but the section has been strengthened and the part in italics added:

“This approach, well known in econometrics where it was initially introduced, allows the modelling of any quantile of a continuous endpoint, here DAH30, as a linear combination of the covariates. As DAH30 is left skewed with a spike at zero, it is more relevant to model the median (or alternatively, the 75th percentile) that is closer to the major distribution mode and directly interpretable. The choice of the quantile(s) to be analysed can be pre-specified or a range of values selected for their meaningfulness or exploratory purposes. Here the range 50th-75th percentile was deemed relevant. No assumption on the true distribution of the endpoint is required.”(see page 9, red text)

There are few disadvantages as no assumption is actually needed on the distribution that must be continuous (or at least with few ties) for quantiles to be defined as indicated above. The linearity of the quantile as a function of the parameters must be satisfied like in the standard linear regression but this is of no concern here as we categorised age, surgery time to avoid any issues. We added the following.

“The asymptotic distribution of the parameter estimates can be derived but depends on some unknown density estimate. In general, resampling methods are recommended to obtain confidence intervals”. [37,38] [Reference 37 and 38 are two new references, namely Wu 1986; He and Hu 2002] Their calculation is easy as they are implemented in all major statistical software.

This point may be considered as a disadvantage as the resampling method may differ according to the package used. In many cases, they give very similar results but for the sake of completion we also added the part in italic in the next sentence. Raw and adjusted medians and their 95% confidence intervals (CIs) obtained by bootstrapping as implemented in Stata with 1000 replicates were reported for key predictors.

We also added in the supplementary material where we present the Q3 results that: The effect of the

different covariates were largely consistent across a large range of meaningful percentile values (e.g. 50th – 75th) with a slightly smaller effect for age categories as the percentile gets higher but for simplicity we only present the results for Q3 (75th percentile). This percentile is also close to the main mode of the distribution. (see Supplement page 4, red text)

(3)It does not always mean DAH30 is a valid index that the authors obtained their satisfactory results using the DAH30 with a minor analysis. It would relate the “reliability and validity” in general sense. The reliability here closely relate to generalizability of the features of DAH30, and the validity here relate to “construct validity”. The authors should show them.
Our Response: We accept this, and have modified our description on each occasion (red text), using “construct validity” where appropriate. (red text: Abstract, pages 7, 12, 15).

Minor Comments:
We can understand the index Q3 was the third quartile because the definition was written in the tables. The authors should describe its definition at least in the main manuscript.
Our Response: We agree: now corrected.

VERSION 2 – REVIEW

REVIEWER	Rupert Pearse Queen Mary University of London, UK
REVIEW RETURNED	01-May-2017

GENERAL COMMENTS	The authors have made some changes to the main text and added additional information to the supplementary file. The use of the word 'pragmatic' in place of ideal gives a much more objective impression. My only remaining concern is about the generalisability of the DAH30 in international trials, which is likely to be where this will get used. The new paragraph in the discussion suggests the authors could be more aware of this: 'Although concerns are frequently raised about the usefulness of hospital length of stay as an outcome measure after surgery, largely because of social factors and reluctance to discharge on weekends, it mostly adds variance (background noise) in clinical trials and is not biased.' This is probably true. 'Very few hospitals have the luxury of extending a patient’s stay in hospital for non-clinical reasons.' Perhaps not in some countries but I suspect the authors do not know for sure. In the UK we don't have this luxury but we still have huge problems discharging patients for social reasons. Despite this, Germany has longer in-patient stays for similar surgery. I am not sure this issue can or should be dismissed so readily. 'Hospital stay is a reasonable surrogate for quality and speed of recovery after surgery...'. The authors provide no data to support this statement and I suspect it is simply their opinion.

	'...and it has marked resource/cost implications.' Yes, of course. 'Most patients want to go home as soon as possible – it is a desired outcome in and of itself.' Again, this is the authors opinion and not supported by data. I am sure there are situations when this is not as simple as portrayed. Overall, I remain of the view that this is a good piece of methodological research but oversold!

REVIEWER	Alexandre Stephens Public Health Observatory, Sydney Local Health District, Australia
REVIEW RETURNED	10-May-2017

GENERAL COMMENTS	Revisions have adequately addressed comments. Thanks

REVIEWER	Hideto TAKAHASHI National Institute of Pblic Health, Saitama, JAPAN (I moved in this April)
REVIEW RETURNED	16-May-2017

GENERAL COMMENTS	The authors carefully modified the article according to the comemnts. I think it can be published.

VERSION 2 – AUTHOR RESPONSE

Please revise the manuscript to be much more balanced in how it describes the use of DAH30, as per the suggestions of Reviewer 2.
Our Response: We accept this and have modified the manuscript accordingly – see below.

Reviewer Name: Rupert Pearse
The authors have made some changes to the main text and added additional information to the supplementary file. The use of the word 'pragmatic' in place of ideal gives a much more objective impression.
Our Response: Thank you.

My only remaining concern is about the generalisability of the DAH30 in international trials, which is likely to be where this will get used. The new paragraph in the discussion suggests the authors could be more aware of this:
'Although concerns are frequently raised about the usefulness of hospital length of stay as an outcome measure after surgery, largely because of social factors and reluctance to discharge on weekends, it mostly adds variance (background noise) in clinical trials and is not biased.' This is probably true.
Our Response: Thank you.

'Very few hospitals have the luxury of extending a patient's stay in hospital for non-clinical reasons.' Perhaps not in some countries but I suspect the authors do not know for sure. In the UK we don't have this luxury but we still have huge problems discharging patients for social reasons. Despite this, Germany has longer in-patient stays for similar surgery. I am not sure this issue can or should be dismissed so readily.

Our Response: Okay, we accept this and have removed this sentence from our manuscript.

'Hospital stay is a reasonable surrogate for quality and speed of recovery after surgery...'. The authors provide no data to support this statement and I suspect it is simply their opinion.
Our Response: No, there are hundreds (thousands?) of studies that have shown this (and most ERAS studies are premised on this fact) – we have included a few new references to support this statement.

'...and it has marked resource/cost implications.' Yes, of course.
Our Response: Thank you.

'Most patients want to go home as soon as possible – it is a desired outcome in and of itself.' Again, this is the authors opinion and not supported by data. I am sure there are situations when this is not as simple as portrayed.
Our Response: We, and others, have undertaken numerous patient surveys regarding expectations after surgery. We outlined some of this in our introduction: "Our own work and that of others have shown that early return home after surgery,6,16-18 and medical illnesses such as stroke,19,20 is highly valued by patients". There may be some situations when this is not as simple as portrayed, but it shouldn't detract from the obvious reality for the vast majority. This is face validity.

VERSION 3 – REVIEW

REVIEWER	Rupert Pearse Queen Mary University of London RP holds research grants, and has given lectures and/or performed consultancy work for Nestle Health Sciences, BBraun, Medtronic, Glaxo Smithkline, and Edwards Lifesciences, and is a member of the Associate editorial board of the British Journal of Anaesthesia.
REVIEW RETURNED	28-May-2017

GENERAL COMMENTS	Whilst the authors response appeared to dismiss my comments, they have in fact made a very small number of changes to the manuscript to remove some of the statements at issue. Clearly this is a strong, and well written manuscript describing a useful methodological innovation. However, it would be nice to see a properly balanced discussion of the limitations of the new outcome measure so we can promote the best possible use of this amongst the less expert researchers in our field. I was a little disappointed that the authors did not take this quite as seriously as I had expected them to. It only serves to promote the innovation if we provide an honest guide to the future user.

VERSION 3 – AUTHOR RESPONSE

Reviewer Name: Rupert Pearse
"However, I would like to see…"
Our Response: Our manuscript reports our findings and interpretation, and clearly outlines many potential limitations; balance cannot be achieved until further studies have elucidated the genuine strengths and weaknesses of the DAH30 metric in other settings. We have thus added in a comment about the need for external validation in the concluding section of the Discussion.