PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Balance and walking after three different models of stroke rehabilitation: early supported discharge in a day-unit or at home, and traditional treatment (control)
AUTHORS	Gjelsvik, Bente; Hofstad, Håkon; Smedal, Tori; Eide, Geir; Næss, Halvor; Skouen, Jan; Frisk, Bente; Daltveit, Silje; Strand, Liv

VERSION 1 - REVIEW

REVIEWER	Nicola Saywell AUT University, Auckland, New Zealand
REVIEW RETURNED	29-Nov-2013

GENERAL COMMENTS	I enjoyed this paper and found the results to be clear and helpful for those involved in discharge planning for people with stroke. I have made very minor editing suggestions but otherwise think that this adds significantly to the area of stroke rehabilitation.

- The reviewer also provided a marked copy with comments. Please contact the publisher for full details.

REVIEWER	Susan Hillier Associate Professor: Neuroscience and Rehabilitation University of South Australia Australia
REVIEW RETURNED	12-Dec-2013

GENERAL COMMENTS	2. Abstract - need to clarify what usual care is to allow understanding of comparison. 4. Methods - it is unclear until page 10 what the control comparator is - this is important to know up front (ad hoc, uncoordinated, uni-disciplinary). Also it is not known until the discussion that the intervention was intended to be EARLY supported discharge as the early aspect is not mentioned in the intervention descriptions. 9. The results need to be reordered to address the research questions in order. There is a primary outcome and this needs to be presented before the secondary outcomes. ditto in discussion which needs to be rewritten in a logical order. 12. There are other study limitations besides those discussed: what was fidelity like (adherence to trial interventions within subjects). in order for us to make sense of the results it would be very helpful to understnad the levels of intensity achieved by the diffferent groups. One factor that distinguishes between int 1 and int 2 is the environment or context of delivery. However it is highly likely that there will have been a different "dosage" achieved. It would ideally be compared to the dosage received in the control group however

	this was unfortunately not collected. So the factor that distinguishes between the int groups and the control is not only "coordinated team care" bus also potentially dosage (intensity). These issues need to be addressed. Statististics - there are multiple outcomes and some consderation needs to be given to the multiple comparisons issue (capitalising on chance). Particularly important because there are obviously other outcomes that are going to be reported in other papers. generally the english is comprehensible but could be tightened for an international journal. minor edits: p4 line 19 stroke not strokes p4 line 45 tiredness should be referred to as fatigue p9 line 10 "too poor knowledge" say insuffcient norwegian Page 10 first paragraph for interventions - this is confusing about the two health care teams (ambulatory versus community) - please use the same terminology for these two service delivery options throughout the paper. PASS - does this test correlate with anything functional or meaningful to the person? For all outcomes measures state which ICF domain they fall in (you state hey all conform to either body structure or activity). Page 13 - line 3 - BI in full first up. Thank you for this important paper. With editting and clarification as noted, this is important work to guide the way we offer services for people post stroke. Overall the RCT design is strong and to be applauded.

REVIEWER	RUTH PICKERING UNIVERSITY OF SOUTHAMPTON UK
REVIEW RETURNED	28-Jan-2014

GENERAL COMMENTS	This paper reports outcomes at 3 months post randomisation in a 3-armed randomised controlled trial of Early Supported Discharge (ESD) following stroke. A major problem I had with the trial was its description as an RCT with primary outcome being the PASS at 3 months after randomisation (Abstract, page 4, lines 41-48). There is a power calculation for PASS on page 13, lines 20-40. However on page 8, lines 24-29, the current trial is described as being conducted in the context of a larger RCT. The trial registration number (NCT00771771), page 5, line 41, is that for the larger trial, as is the published protocol (Ref 10). The 3 arms and numbers of patients recruited in the current study are the same as for the larger trial, but the primary outcome of the larger trial is the modified Rankin Scale (mRS) at 6 months. The results of the primary analysis of the main trial have not yet been published. I felt the current paper reported analysis of a secondary outcome from the main trial, and shouldn't be described as a separate RCT, albeit conducted in the context of a larger RCT. Otherwise each outcome could be published as a separate RCT and there would be no accounting for the multiplicity involved in analysing lots of different outcomes. The sample size for the trial was determined to show a difference in mRS at 5 months as described in the protocol (ref 10).

	Another general problem I had with the trial was the timing of recruitment, intervention and follow-up. Participants are recruited close to stroke onset (within 7 days), and the knowledge that ESD would be available post-discharge was expected to facilitate earlier discharge, and this along with the support delivered, would be instrumental elements in producing benefit. By 6 months most patients should have received most of the rehab they will get, and the 6 month primary endpoint in the main trial seems a sensible time at which to assess whether the eventual recovery is better in the ESD arms. The current paper focuses on outcome at 3 months: only 105 (34%) of the 306 randomised participants were included in the current analysis. I would have thought that some patients particularly in the non-ESD arm of the trial might still be in hospital at 3 months. This doesn't seem to be the case though: in the CONSORT diagram, page 6, no participants are excluded from analysis because they are still in hospital. In the final row of table 1 the mean days in the stroke unit across the 167 participants discharged home was 8.6 (max=43), in appendix table the days in stroke unit were means 8.5, 8.7, and 8.7 (max=17, 18, 22) in the particpants in the analysis in the three arms. So that it seems all participants recruited were discharged by 3 months. It would be helpful to see to see what happened in terms of hospital length of stay and discharge destination of the 306 randomised participants, split by the three intervention groups - but maybe this is one of the outcomes to be reported in the main trial paper. In an analysis of data relating to all stroke patients within a hospital catchment area in the UK I analysed recently, patients stayed on average 44 days in hospital, I couldn't see anything in the eligibility criteria that would restrict the BERGEN ESD trial to patients with such a short length of stay. Maybe stroke care is organised differently in Bergen so that patients going to a stroke unit are different from those I looked at? Another thing that surprised me about the CONSORT flow chart (page 6) was that none of the participants died between recruitment and 3 months. The flow chart starts with the number of patients meeting the eligibility criteria for the ESD Bergen study. It would be useful to start higher up with the number of patients with stroke in the relevent catchment area, so we can see how representative participants in the study are of the whole. Generalisability is a weakness of this study as the 105 participants constitute only 34% of the 306 randomised. Maybe some description of stroke care in Bergen would be helpful in explaining why length of hospital stay was so much shorter than in the UK. Other points: 1 Abstract, page 4 lines 52-54, 95% confidence intervals are presented for median PASS at 3 months in each treatment group. It would be better to present 95% confidence intervals for effect sizes - the differences in medians between each experimental arm and control. Confidence intervals for differences in medians can't be directly obtained in SPSS, but are available in other packages, otherwise they could give confidence intervals for the difference in means (in table 5 a multiple linear regression is presented for PASS which models mean values assuming normality). 2 Page 8, lines 8-15. The research objectives are stated as whether ESD (either at home or in a day-unit) is superior to standard care. This suggests only two contrasts are of interest ESD-home versus standard care, and ESD-day-unit versus standard care. The

	comparison between home and day-unit care isn't stated as an objective here. In the Abstract it is stated that they wish to compare all three treatment options though. The two statements of objectives should be consistent. If they are only interested in the two comparisons to standard care, then the Bonferroni adjustement for multiple comparisons described on page 14, lines 19-22, need only divide by 2. 3 Page 9, lines 21-30, description of randomisation procedure. Allocations were computer generated but then a nurse at the stroke unit gave the allocations to patients. There is no description of any procedures to conceal allocation from those recruiting participants, who could have subverted the process by rigging the order in which participants were recruited to the trial. They need to describe procedures taken to ensure this couldn't happen. 4 Naming of the experimental groups. I would have found Table 4 easier to follow if instead of referring to experimental groups I and II (notation established on page 10), they referred to the groups as HOME and DAY-UNIT care throughout, for example. I'm not sure if they refer to experimental group I and II other than in table 4 and the CONSORT diagram. 5 Page 11, line 49/50, there should be a reference establishing >=3 as true change for the for the TIS-modNV. Are these true changes assessed using the minimum detectable change (MDC) statistic? If so it would be helpful to use this notation. 6 The sample size calculation, page 13. Note that the SD they quote is for a single value not for change, however this might be expected to over-estimate the sample size required for comparing changes. I wasn't able to replicate the calculation - partly because the value of 4 for a mean change equally spaced between groups wasn't clearly a quantity I needed to feed into my sample size program (nQuery), nevertheless I came up with numbers close to their's. They could have included a Bonferroni correction, reducing the alpha to 0.0167, if they intended to do Bonferroni adjustments. My main issue with the sample size calculation is that it would have been better to acknowledge that it had been determined in relation to the main trial objective of showing differences in mRS at 6 months. 7 Page 14, line 57. Given the relatively small sample size, percentages throughout the paper should be presented with no decimal places, ie 48% here This applies to table 1 (and %s in other tables) as well. 8 It is generally a good thing to address whether there are selection biases in the sample available for analysis - for example moving from the 167 patients discharged home, to those re-tested with PASS vs not (Table 1). But I found it quite confusing in relation to this study. There were 306 randomised, the 167 were a selection of those. In the end the participants included in the analysis are such a small % of those randomised it was difficult to think whether the results presented relate to any patient group. The main trial assessing the consequences at 6 months of the different discharge policies, may give useful comparative material in relation to survival, residential status and the mRS. I'm not sure that outcome amongst a very selected group at 3 months contributes useful information in assessing ESD. The differences between groups shown in Table 4 don't clearly demonstrate the ESD groups to be superior - the

	primary outcome shows little difference. The %s of those randomised included in the analyses (26%, 41% & 35%) are quite different indicating that we can't rely on randomisation to produce similar patients in these groups, even though the three groups at the point of discharge home (total 167 of the 306 randomised) may be similar with respect to the variables shown in Table 3, and similarly in the APPENDIX table amongst subjects available for analysis. So the lack of convincing evidence favouring the ESD here may not give a true picture of the benefits or otherwise of ESD. 9 In Table 4 there are 27, 43, and 39 participants in the analyses reported in the table, but in the CONSORT diagram these are the numbers with test and re-test PASS. Did any of the other variables reported in the table have additional missing values so that the results come from a smaller set of subjects? 10 The multiple regression in Table 5. This doesn't contribute much to the argument. The variable predicting PASS at 3 months - age, PASS at baseline, previous cerebral lesion and previous nursing care - are all as would be anticipated, the lack of statistical significance in relation to the other variable may be due to small sample size. I felt this analysis might be better omitted. 11 Page 25, line 25. Start of the Discussion. They start by listing variables where statistically significant differences were found. But if the PASS is the primary outcome they should start by stating that they found no significant difference for that. Also 14 variables were tested for between group differences in Table 4 of which only 3 demonstrated statistical significance. This is an aspect of multiplicity not encompassed by the Bonferroni corrections which only account for the 3 pairwise comparisons potentially drawn. 12 Discussion, page 25, lines 52-53, reporting lack of earlier discharge in the ESD groups compared to the control group. The lack of statistical significance in length of stay is commented on in the Results section, page 19, lines 18-21, but they haven't previously commented on the similarity in mean length of stay shown in the appendix table which could usefully be mentioned in the Results section. This might seem to be a key finding, but examining length of stay wasn't a stated objective for the current study. It would be useful to see lengths of stay by group amongst the whole randomised group, and also amongst those discharged home, rather than just the subset with re-test PASS at 3 months. 13 Because of the very restricted subset of participants included here, and because of the lack of statistical significance shown for most variables in Table 4 including that stated to be primary, I didn't feel the authors could conclude on page 29, lines 5-14 that ESD for patients is somewhat superior to standard care. 14 In the Conclusion section, page 29, they state their findings as relating to patients with mild disability after stroke. They hadn't previously said that the study related to people with mild disability - or if so I hadn't picked it up. One eligibility criteria states a NIHSS score of 2-26, page 9, line 2/3, and at the bottom of page 12 mild stroke is NIHSS scores of 0-7. In Table 3, though, the median NIHSS score in each group is in the mild range, but the maximum scores are in the severe range for two of the groups. This might explain the short lengths of hospital stay I commented on earlier, and also the excellent survival of participants. Perhaps the

	description of the eligibility criteria on page 8 could be expanded to explain why they resulted in people with mild stroke.

VERSION 1 – AUTHOR RESPONSE

REVIEWER NICOLA SAYWELL
I enjoyed this paper and found the results to be clear and helpful for those involved in discharge planning for people with stroke. I have made very minor editing suggestions [attached] but otherwise think that this adds significantly to the area of stroke rehabilitation.
-Response: Thank you. The editing suggestions have been incorporated within the revised manuscript.

REVIEWER SUSAN HILLIER
Thank you for this important paper. With editing and clarification as noted, this is important work to guide the way we offer services for people post stroke. Overall the RCT design is strong and to be applauded
-Response: Thank you

Abstract - need to clarify what usual care is to allow understanding of comparison
-Response: Changed to “traditional, uncoordinated treatment” in the Abstract and in the Methods section (pages 4 and 11).

Methods - it is unclear until page 10 what the control comparator is - this is important to know up front (ad hoc, uncoordinated, uni-disciplinary). Also it is not known until the discussion that the intervention was intended to be EARLY supported discharge as the early aspect is not mentioned in the intervention descriptions
-Response: We agree. We have therefore added “early” to the intervention descriptions both in the Abstract and Methods sections (pages 4 and 10).

The results need to be reordered to address the research questions in order. There is a primary outcome and this needs to be presented before the secondary outcomes. ditto in discussion which needs to be rewritten in a logical order
-Response: Presentation of the results has been reordered in both the Results and Discussion sections.

There are other study limitations besides those discussed:
What was fidelity like (adherence to trial interventions within subjects)
-Response: Only 1 patient in the day-unit and 2 in the home-rehabilitation groups did not receive the assigned experimental interventions. This information is given in the flowchart.

In order for us to make sense of the results it would be very helpful to understand the levels of intensity achieved by the different groups
-Response: We do not have this information on more than 50% of the participants. Based on available information from the included patients we have calculated a mean intervention intensity/dosage which we assume is representative for the groups. This information has been added to the Results section.

One factor that distinguishes between int 1 and int 2 is the environment or context of delivery. However it is highly likely that there will have been a different "dosage" achieved. It would ideally be compared to the dosage received in the control group however this was unfortunately not collected. So the factor that distinguishes between the int groups and the control is not only "coordinated team care" bus also potentially dosage (intensity). These issues need to be addressed
-Response: According to available information, the total intervention from the community health care

team for the planned intervention period of 5 weeks amounted to mean (SD), min – max: 22.0 hrs. (9.5), 4.0 – 50.1 for the day-unit group; and 16.6 hrs. (9.0), 0.5 – 48.5 for the home-rehabilitation group. Mean difference (95%CI) between the groups was 5.4 hrs. (1.6, 9.1), p = 0.006. The duration lasted for mean 4.4 weeks for the day-unit and 4.2 week for the home-rehabilitation patients, p = 0.240. This information has been added to the Results section (page 23) and is discussed in the Discussion section (page 26).

Statististics - there are multiple outcomes and some consideration needs to be given to the multiple comparisons issue (capitalising on chance). Particularly important because there are obviously other outcomes that are going to be reported in other papers.
-Response: We agree that there are many outcomes, therefore we chose one, PASS, to be the primary outcome. The secondary outcomes were chosen to study different aspects of balance and walking capacity, as well as patients' perspectives mostly of the same aspects of function. We have removed 3 variables from Table 4: NIHSS, BI and mRS, as these were meant to give background information only, and are left with 11 (page 22). We have discussed the aspect of multiplicity, and to Bonferroni adjust for 11 variables would be considered too strict as these variables will be more or less correlated, and a formal correct adjustment is difficult.
For pairwise comparisons of the three groups we adjusted the significance level to 0.05/3=0.0167 when the groups were found to be significantly different after an ANOVA analysis (F-test), which is described in the Statistics part of the Methods section (page 15).
This study is the only study that reports the outcome for the physical tests at 3 months.

Generally the English is comprehensible but could be tightened for an international journal
-Response: Thank you. We have attempted to do so.

Minor edits:
p4 line 19 stroke not strokes
-Response: It has been corrected.

p4 line 45 tiredness should be referred to as fatigue
-Response: We chose to use the word "tiredness", as "fatigue" is a complex term that may involve cognitive, physical and psychosocial aspects. We did not assess these aspects; we only asked the patients if they felt tired due to the stroke. Therefore we did not use the term "fatigue".

p9 line 10 "too poor knowledge" say insufficient Norwegian
-Response: This has been corrected (page 8). We have also specified the exclusion criteria: psychological illness has been changed to "Psychiatric disorder", drug abuse to "current alcohol and substance abuse", to comply with the ESD Stroke Bergen trial protocol.

Page 10 first paragraph for interventions - this is confusing about the two health care teams (ambulatory versus community) - please use the same terminology for these two service delivery options throughout the paper
-Response: The ambulatory team was not community based, but a hospital out-reach team. We understand that this has been unclear, and have changed it to "hospital out-reach team" throughout. We used both terms "community health care team" and "health care team" for the municipal part of the services, and have now changed this to "community health care team" throughout.

PASS - does this test correlate with anything functional or meaningful to the person?
-Response: A strong correlation between PASS and Berg Balance Scale (O'Dell et al. 2013), Functional Independence Measure (FIM) and destination at discharge (Di Monaco et al. 2010) have been demonstrated. The subscore PASS-Trunk Control has been found to be a strong predictor of comprehensive ADL function (Hsieh 2002).

For all outcomes measures state which ICF domain they fall in (you state they all conform to either body structure or activity).
-Response: This information is added to the description of the outcome measures in the Methods section (pages 11-12).

Page 13 - line 3 - BI in full first up.
-Response: This has been corrected (page 13).

REVIEWER RUTH PICKERING
This paper reports outcomes at 3 months post randomisation in a 3-armed randomised controlled trial of Early Supported Discharge (ESD) following stroke.
A major problem I had with the trial was its description as an RCT with primary outcome being the PASS at 3 months after randomisation (Abstract, page 4, lines 41-48).
-Response: We agree that it might be misleading to call this an RCT and have changed the description of the design. In the Abstract: “Group comparison within a randomised controlled trial” and in the Methods section: “The study was conducted in the context of a larger single blinded randomised controlled trial”.

There is a power calculation for PASS on page 13, lines 20-40. However on page 8, lines 24-29, the current trial is described as being conducted in the context of a larger RCT. The trial registration number (NCT00771771), page 5, line 41, is that for the larger trial, as is the published protocol (Ref 10). The 3 arms and numbers of patients recruited in the current study are the same as for the larger trial, but the primary outcome of the larger trial is the modified Rankin Scale (mRS) at 6 months. The results of the primary analysis of the main trial have not yet been published. I felt the current paper reported analysis of a secondary outcome from the main trial, and shouldn't be described as a separate RCT, albeit conducted in the context of a larger RCT. Otherwise each outcome could be published as a separate RCT and there would be no accounting for the multiplicity involved in analysing lots of different outcomes.
-Response: In this study we did not analyse results after 6 months as this will be the focus of the ESD Stroke Bergen trial, and we did not use mRS as an outcome variable. This is the only study reporting results on physical tests after 3 months, being close to the end of the intervention period for the patients who were discharged directly home from the stroke unit.

The sample size for the trial was determined to show a difference in mRS at 6 months as described in the protocol (ref 10).
-Response: That is correct for the main RCT trial. Our study was planned together with the RCT trial, and we calculated the power needed to be able to show a difference in change of PASS scores between groups for this study.

Another general problem I had with the trial was the timing of recruitment, intervention and follow-up. Participants are recruited close to stroke onset (within 7 days), and the knowledge that ESD would be available post-discharge was expected to facilitate earlier discharge, and this along with the support delivered, would be instrumental elements in producing benefit. By 6 months most patients should have received most of the rehab they will get, and the 6 month primary endpoint in the main trial seems a sensible time at which to assess whether the eventual recovery is better in the ESD arms. The current paper focuses on outcome at 3 months: only 105 (34%) of the 306 randomised participants were included in the current analysis.
-Response: For this study focusing on physical function using PASS as the primary outcome, all patients had to be tested with PASS to be included. As the intention of the present study was to investigate the effect of early supported discharge, i.e. rehabilitation with the patients living at home, patients being discharged to other institutions or hospital departments were excluded. The

intervention period (duration) was planned to be 5 weeks after discharge from the stroke unit, and we therefore chose 3 months after inclusion as follow-up time, as physical function would be assessed closer to end of treatment of patients sent directly home. Consequently the results could be linked more directly to type of intervention; i.e. day-unit, home-rehabilitation or control.

I would have thought that some patients particularly in the non-ESD arm of the trial might still be in hospital at 3 months. This doesn't seem to be the case though: in the CONSORT diagram, page 6, no participants are excluded from analysis because they are still in hospital. In the final row of table 1 the mean days in the stroke unit across the 167 participants discharged home was 8.6 (max=43), in appendix table the days in stroke unit were means 8.5, 8.7, and 8.7 (max=17, 18, 22) in the participants in the analysis in the three arms. So that it seems all participants recruited were discharged by 3 months. It would be helpful to see to see what happened in terms of hospital length of stay and discharge destination of the 306 randomised participants, split by the three intervention groups - but maybe this is one of the outcomes to be reported in the main trial paper.
-Response: According to our registration all patients were discharged from hospital before 3 months. We have looked at the total length of hospital stay for all patients in the ESD Stroke Bergen (N=306), irrespective of initial discharge destination (other hospital departments, specialist hospital-based rehabilitation, community rehabilitation, nursing home). We found a mean (SD), min – max: 11.4 days (7.02), 1 – 43 (day-unit 11.3 days; home-rehabilitation 11.3 days ; control 11.6 days) with no difference between the groups (P=0.923). This will be reported in the main paper from ESD Stroke Bergen.

In an analysis of data relating to all stroke patients within a hospital catchment area in the UK I analysed recently, patients stayed on average 44 days in hospital, I couldn't see anything in the eligibility criteria that would restrict the BERGEN ESD trial to patients with such a short length of stay. Maybe stroke care is organised differently in Bergen so that patients going to a stroke unit are different from those I looked at?
-Response: In Bergen, the community health services are organized with rehabilitation beds in nursing homes, which means that the patients are discharged overall relatively early. The municipalities have a legal responsibility to take over the responsibility for the patients when they are medically stable and have received their acute medical treatment and no further specialist hospital services are required. The municipality receives heavy fines if the patients have to stay in hospital because of lack of beds or resources in the community health services. This could explain why length of stay is different between Norway and Great Britain.

Another thing that surprised me about the CONSORT flow chart (page 6) was that none of the participants died between recruitment and 3 months.
-Response: This is a mistake: this information should have been reported in the flow chart. However, only one patient of the included 167 died during the 3 month period. This has been corrected.

The flow chart starts with the number of patients meeting the eligibility criteria for the ESD Bergen study. It would be useful to start higher up with the number of patients with stroke in the relevant catchment area, so we can see how representative participants in the study are of the whole.
-Response: The flow chart is so extensive that we have added this information in text in the Results section. In all, 1749 patients were admitted to the stroke unit during the inclusion period, of which 1430 did not meet the inclusion criteria. The main reason for not meeting the inclusion criteria was that the patients were not residents in the city of Bergen or that the stroke was not confirmed. In addition, we have added two extra text boxes in the flow chart to mark the start of the present study and the number of included participants.

Generalizability is a weakness of this study as the 105 participants constitute only 34% of the 306 randomised.

-Response: 167 patients were eligible for inclusion in the present study (tested with PASS and discharged directly home), of which 105 patients (63%) were retested 3 months after inclusion. Patients that have been found to benefit from ESD services are mainly mildly to moderately disabled. We argue therefore that our results are representative for patients with this level of disability.

Maybe some description of stroke care in Bergen would be helpful in explaining why length of hospital stay was so much shorter than in the UK.
-Response: This is outlined above.

Other points:
1 Abstract, page 4 lines 52-54, 95% confidence intervals are presented for median PASS at 3 months in each treatment group. It would be better to present 95% confidence intervals for effect sizes - the differences in medians between each experimental arm and control. Confidence intervals for differences in medians can't be directly obtained in SPSS, but are available in other packages, otherwise they could give confidence intervals for the difference in means (in table 5 a multiple linear regression is presented for PASS which models mean values assuming normality).
In the Abstract (p. 4), we have reordered the results and start by stating that there was no difference in change between the groups for PASS. The actual results given refer to differences in change for trunk control and some of the self-report measures.
-Response: The regression table (Table 5) has been removed from the manuscript, and the results for PASS between control and each experimental group are presented in the text as differences in means with 95% confidence intervals (page 24).
2 Page 8, lines 8-15. The research objectives are stated as whether ESD (either at home or in a day-unit) is superior to standard care. This suggests only two contrasts are of interest ESD-home versus standard care, and ESD-day-unit versus standard care. The comparison between home and day-unit care isn't stated as an objective here. In the Abstract it is stated that they wish to compare all three treatment options though. The two statements of objectives should be consistent.
-Response: The objectives have been reworded. We intended to compare all three groups, and then allow for pairwise comparison between groups if relevant.

If they are only interested in the two comparisons to standard care, then the Bonferroni adjustement for multiple comparisons described on page 14, lines 19-22, need only divide by 2.
-Response: We were interested in comparing all three groups, and therefore performed pairwise comparisons between all three groups when relevant; day-unit vs home-rehabilitation, day-unit vs control, and home-rehabilitation vs control, based on the ANOVA analyses. We therefore divided by 3 to adjust for multiple comparisons.

3 Page 9, lines 21-30, description of randomisation procedure. Allocations were computer generated but then a nurse at the stroke unit gave the allocations to patients. There is no description of any procedures to conceal allocation from those recruiting participants, who could have subverted the process by rigging the order in which participants were recruited to the trial. They need to describe procedures taken to ensure this couldn't happen.
-Response: The person recruiting participants did not have access to the randomisation list, which was kept by the study coordinator. This information has been added to the Methods section (page 9).
4 Naming of the experimental groups. I would have found Table 4 easier to follow if instead of referring to experimental groups I and II (notation established on page 10), they referred to the groups as HOME and DAY-UNIT care throughout, for example. I'm not sure if they refer to experimental group I and II other than in table 4 and the CONSORT diagram.
-Response: Agree. The terms have been changed in the tables as well as in the text, even though they do appear in some places where appropriate (when the two experimental groups are discussed together).
5 Page 11, line 49/50, there should be a reference establishing >=3 as true change for the the TIS-

modNV. Are these true changes assessed using the minimum detectable change (MDC) statistic? If so it would be helpful to use this notation.
-Response: This is correct and has been changed for both PASS and TIS-modNV (pages 11-12).

6 The sample size calculation, page 13. Note that the SD they quote is for a single value not for change, however this might be expected to over-estimate the sample size required for comparing changes. I wasn't able to replicate the calculation - partly because the value of 4 for a mean change equally spaced between groups wasn't clearly a quantity I needed to feed into my sample size program (nQuery), nevertheless I came up with numbers close to their's. They could have included a Bonferroni correction, reducing the alpha to 0.0167, if they intended to do Bonferroni adjustments. My main issue with the sample size calculation is that it would have been better to acknowledge that it had been determined in relation to the main trial objective of showing differences in mRS at 6 months
-Response: We pre-calculated power for our study and found that we needed at least 60 patients in total (20 in each group) to reach a power of at least 90% with a significance level of 5%. However, the main trial, ESD Stroke Bergen, estimated a sample size of 350 patients with acute stroke based on mRS. We therefore decided to use all the collected data in the ESD Stroke Bergen to analyse the 3 month outcome for the physical tests, as this would strengthen the scientific evidence of our results.

7 Page 14, line 57. Given the relatively small sample size, percentages throughout the paper should be presented with no decimal places, i.e. 48% here This applies to table 1 (and %s in other tables) as well
-Response: This has been corrected.

8 It is generally a good thing to address whether there are selection biases in the sample available for analysis - for example moving from the 167 patients discharged home, to those re-tested with PASS vs not (Table 1). But I found it quite confusing in relation to this study. There were 306 randomised, the 167 were a selection of those. In the end the participants included in the analysis are such a small % of those randomised it was difficult to think whether the results presented relate to any patient group.
-Response: 167 patients fulfilled the inclusion criteria for this study, and the patients in the different groups were similar with respect to background variables. Of the 167 included patients in this study, 52% of the day-unit patients, 72% of the home-rehabilitation patients and 64% of the control patients was retested. Therefore, we argue that the results are relevant for patients who are discharged directly home and have mostly mild to moderate disability.

The main trial assessing the consequences at 6 months of the different discharge policies, may give useful comparative material in relation to survival, residential status and the mRS. I'm not sure that outcome amongst a very selected group at 3 months contributes useful information in assessing ESD.
-Response: We chose 3 months as analysis for final outcome of physical function as this would be closer to the end of the intervention period. The outcome could therefore be linked more closely to type of intervention.

The differences between groups shown in Table 4 don't clearly demonstrate the ESD groups to be superior - the primary outcome shows little difference. The %s of those randomised included in the analyses (26%, 41% & 35%) are quite different indicating that we can't rely on randomisation to produce similar patients in these groups, even though the three groups at the point of discharge home (total 167 of the 306 randomised) may be similar with respect to the variables shown in Table 3, and similarly in the APPENDIX table amongst subjects available for analysis. So the lack of convincing evidence favouring the ESD here may not give a true picture of the benefits or otherwise of ESD.
-Response: For patients with mostly mild to moderate disability, the primary outcome did not show any difference between the groups. However, we did find a difference for secondary measures like trunk control and some of the self-reported measures, which all tended to show more favourable

results after ESD than after traditional treatment, which is commonly uni-disiplinary (physiotherapy) and not coordinated between different health professionals.

9 In Table 4 there are 27, 43, and 39 participants in the analyses reported in the table, but in the CONSORT diagram these are the numbers with test and re-test PASS. Did any of the other variables reported in the table have additional missing values so that the results come from a smaller set of subjects?
-Response: This is correct, and should be 27, 43 and 35 as reported in the flow chart and Table 4. Table 4 shows results for change scores, and therefore for patients who were tested and retested. There were no additional missing values.

10 The multiple regression in Table 5. This doesn't contribute much to the argument. The variable predicting PASS at 3 months - age, PASS at baseline, previous cerebral lesion and previous nursing care - are all as would be anticipated, the lack of statistical significance in relation to the other variable may be due to small sample size. I felt this analysis might be better omitted
-Response: Agree. Table 5 has been removed and the group-comparisons for differences in means for PASS have been incorporated in the text (page 24).

11 Page 25, line 25. Start of the Discussion. They start by listing variables where statistically significant differences were found. But if the PASS is the primary outcome they should start by stating that they found no significant difference for that.
-Response: Agree, and this has been changed accordingly (page 23).

11 Also 14 variables were tested for between group differences in Table 4 of which only 3 demonstrated statistical significance.
-Response: We have removed the following variables in Table 4: NIHSS, BI and mRS, as these were meant for background information and not as outcome variables.

This is an aspect of multiplicity not encompassed by the Bonferroni corrections which only account for the 3 pairwise comparisons potentially drawn.
-Response: We agree that there are many outcomes, therefore we chose one, PASS, to be the primary outcome. The secondary outcomes were chosen to study different aspects of balance and walking capacity, as well as patients' perspectives mostly of the same aspects of function. We have discussed this aspect of multiplicity and to Bonferroni adjust for 11 variables would be considered too strict as these variables will be more or less correlated, and a formal correct adjustment is difficult.

12 Discussion, page 25, lines 52-53, reporting lack of earlier discharge in the ESD groups compared to the control group. The lack of statistical significance in length of stay is commented on in the Results section, page 19, lines 18-21, but they haven't previously commented on the similarity in mean length of stay shown in the appendix table which could usefully be mentioned in the Results section. This might seem to be a key finding, but examining length of stay wasn't a stated objective for the current study. It would be useful to see lengths of stay by group amongst the whole randomised group, and also amongst those discharged home, rather than just the subset with re-test PASS at 3 months
-Response: The web only file shows the difference in length of stroke unit stay between the groups for those who were retested only, while in Table 1 length of stroke unit stay is compared between patients who were retested and those who were not.
We have looked at length of hospital stay for all patients in the main trial, ESD Stroke Bergen (n=306), which includes patients who were discharged from the stroke unit to the hospital rehabilitation unit, other hospital departments, community rehabilitation or nursing homes. Total length of hospital stay was on average 11.4 days (day-unit 11.3; home-rehabilitation 11.3; control 11.6) with no difference between the groups.

13 Because of the very restricted subset of participants included here, and because of the lack of statistical significance shown for most variables in Table 4 including that stated to be primary, I didn't feel the authors could conclude on page 29, lines 5-14 that ESD for patients is somewhat superior to standard care.
-Response: We agree that we should primarily adhere to results of what we have defined as the main outcome measure. There was no significant difference between the groups for the primary outcome. However, in these patients with mainly mild to moderate disability we found significant differences for several secondary outcome measures like trunk control and patient perception of ADL and walking that all went in the same direction, and we think these are important outcomes to report on too. As these measures were only secondary, we have tried to take that into consideration when concluding the results both in the Abstract (page 5) and in the revised manuscript (page 27).

14 In the Conclusion section, page 29, they state their findings as relating to patients with mild disability after stroke. They hadn't previously said that the study related to people with mild disability - or if so I hadn't picked it up. One eligibility criteria states a NIHSS score of 2-26, page 9, line 2/3, and at the bottom of page 12 mild stroke is NIHSS scores of 0-7. In Table 3, though, the median NIHSS score in each group is in the mild range, but the maximum scores are in the severe range for two of the groups. This might explain the short lengths of hospital stay I commented on earlier, and also the excellent survival of participants. Perhaps the description of the eligibility criteria on page 8 could be expanded to explain why they resulted in people with mild stroke
-Response: This information has been added to the Results section (page 17). We assume that this relates to the inclusion criterion of patients having to be discharged directly home from the stroke unit.

VERSION 2 – REVIEW

REVIEWER	Susan Hillier University of South Australia Australia
REVIEW RETURNED	22-Mar-2014

GENERAL COMMENTS	This is clearly written report of an RCT into three modes of health service delivery in stroke rehabilitation. The authors acknowledge the pragmatic limitations of the intervention fidelity - they had hoped to limit the intervention effect to the context of delivery but in the final analysis had other difference (intensity differed between the day unit and home care groups as did focus - impairment and task-oriented versus task-oriented. The latter needs to be discussed more as it explains some of the results. Whilst the report is clearly understood some of the english is a litle awkward and could do with a second edit.

REVIEWER	Ruth Pickering University of Southampton UK
REVIEW RETURNED	26-Mar-2014

GENERAL COMMENTS	I found the paper much clearer this time. I've got a couple of points to raise. 1 The direction of differences shown in Table 4 were not immediately clear. In the title it would help if the direction of

	defferencing were included, that is "... change (3 months - baseline) scores ...". Since higher scores on the TIS-modNV indicate better function, for the change of 4 to indicate better improvement in the home-rehab group than the others, it must be 3 months - baseline. The direction of the numerical rating scales was not described at the top of page 12 so I couldn't work out what the changes shown in Table 4 indicated. On page 20, line 19 onwards, they tell us that walking improved more in the day-unit group, and ADL in both rehab groups compared to control: I couldn't work this out from the numbers presented in Table 4. 3 Similarly when reporting the regression results for PASS on page 22, lines 9-14, it isn't clear which group negative differences favour, and here they don't tell us. 4 Discussion, page 23, lines 19-21, the unexpected lack of differences between groups in length of stay in the stroke unit. This was briefly mentioned in the Results section: on page 18, lines 9-12, the P value showing lack of statistical significance was reported; in Table 1 , final row, summary statistics for the group tested and retested (and not) were reported. It would be helpful to see summary statistics for each of the intervention groups in either table 3 or 4 as well. 5 Though there was not quite statistically significant (P=0.097) difference between the 3 groups in loss to follow-up (page 20, line 6/7), the difference in these rates is worrying - 48%, 28%, 36%. I felt this was another weakness of the study and should be included in their paragraph on weaknesses in the study, Page 25, lines 35/36 onwards.

VERSION 2 – AUTHOR RESPONSE

Reviewer 1, Susan Hillier
This is clearly written report of an RCT into three modes of health service delivery in stroke rehabilitation. The authors acknowledge the pragmatic limitations of the intervention fidelity - they had hoped to limit the intervention effect to the context of delivery but in the final analysis had other difference (intensity differed between the day unit and home care groups as did focus - impairment and task-oriented versus task-oriented. The latter needs to be discussed more as it explains some of the results.
Response: Thank you for this comment. The following information has been added to Discussion: p. 24 “It is recommended that such treatment be meaningful and repetitive, and repetitive task-oriented training has been found to result in small amounts of improvement in ADL, which may explain why the home-rehabilitation group experienced most improvement” and p. 25: “In addition, the day-unit patients received on average more treatment from the community health care team than the home-rehabilitation group. More therapy time and more intensity of therapy seem to be beneficial for functional outcome, and could explain why walking seemed to improve more in this group.”

R1: Whilst the report is clearly understood some of the english is a litle awkward and could do with a second edit.
Response: The article has been edited by a language professional before resubmission.

Reviewer 2, Ruth Pickering
Thank you for your comments. The response is given following each point below.
1. The direction of differences shown in Table 4 were not immediately clear. In the title it would help if the direction of defferencing were included, that is "... change (3 months - baseline) scores ...". Since

higher scores on the TIS-modNV indicate better function, for the change of 4 to indicate better improvement in the home-rehab group than the others, it must be 3 months - baseline.
Response: This has been added to the title as suggested.

R2: The direction of the numerical rating scales was not described at the top of page 12 so I couldn't work out what the changes shown in Table 4 indicated.
Response: The following information has been added to Numerical Rating Scales, p. 11: "…are 11-point scales (0 – 10, lowest score best)" in the section on outcome measures. Information on the direction of change is also available in the subtitle of Table 4; for NRS, "negative values indicate improvement in condition."

R2: On page 20, line 19 onwards, they tell us that walking improved more in the day-unit group, and ADL in both rehab groups compared to control: I couldn't work this out from the numbers presented in Table 4.
Response: More detail on the pairwise analyses has been added to the text on p. 20. For trunk control: "day-unit vs. home-rehabilitation, p = 0.031; day-unit vs. control, p = 0.886; home-rehabilitation vs. control, p = 0.040"; for walking: "day-unit vs. home-rehabilitation, p = 0.215; day-unit vs. control, p = 0.004; home-rehabilitation vs. control, p = 0.126"; and for ADL: "day-unit vs. home-rehabilitation, p = 0.774; day-unit vs. control, p = 0.036; home-rehabilitation vs. control, p = 0.006".

2. Similarly when reporting the regression results for PASS on page 22, lines 9-14, it isn't clear which group negative differences favour, and here they don't tell us.
Response: We have reworded this information and made it clearer which groups are the best by changing:
"Multiple regression analyses for PASS at three months demonstrated a significant effect of age, baseline PASS scores, previous cerebral lesion and previous nursing care with an explained variance of 63%. The differences in means (95%CI) for PASS between the control and day-unit groups were -0.194 (-1.823, 1.436), and between control and home-rehabilitation groups -0.599 (-2.048, 0.840). Group allocation did not explain any of the variance in PASS, p = 0.720." to
"Multiple regression analyses for PASS at three months demonstrated a significant negative effect of increased age (b = -0.070), positive effect of higher baseline PASS scores (b = 0.394), positive effect of not having previous cerebral lesion (b = 2.621) and positive effect of not having had previous nursing care (b = 1.220) with an explained variance of 63% in the final model after backward stepwise selection. The control group had higher means than did the other groups, i. e. the unadjusted differences in means (95%CI) between the control and day-unit groups were -0.194 (-1.823, 1.436), and between control and home-rehabilitation groups -0.599 (-2.048, 0.840). Group allocation was not a significant predictor for PASS, neither unadjusted (p = 0.720), nor adjusted for the other variables."

3. Discussion, page 23, lines 19-21, the unexpected lack of differences between groups in length of stay in the stroke unit. This was briefly mentioned in the Results section: on page 18, lines 9-12, the P value showing lack of statistical significance was reported; in Table 1 , final row, summary statistics for the group tested and retested (and not) were reported. It would be helpful to see summary statistics for each of the intervention groups in either table 3 or 4 as well.
Response: I have added information on length of stroke unit stay in the individual groups to the text in the Results section, p. 18: day-unit, mean (SD), min – max: 8.6 (3.3), 3 – 17; home-rehabilitation: 8.7 (3.9), 3 – 18; control: 8.4 (4.5), 3 – 22. This information is also available in the web-only file.

4. Though there was not quite statistically significant (P=0.097) difference between the 3 groups in loss to follow-up (page 20, line 6/7), the difference in these rates is worrying - 48%, 28%, 36%. I felt this was another weakness of the study and should be included in their paragraph on weaknesses in the study, Page 25, lines 35/36 onwards.
Response: The following information has been included in the paragraph on weaknesses of the study,

p. 26; “However, there was a non-significant difference in loss to follow-up between the groups; day-unit 48%, home-rehabilitation 28%, and control 36%. The loss to follow-up is a weakness of the study.”

Other changes made to the manuscript
In the section on outcome measures, p. 11, 5mTW: “...the first 11 patients were not tested” should read “…the first 13 patients were not tested”. The same mistake has been corrected in the subtitles of Table 2.
The difference between the groups for length of stay, p. 18, “p = 0.458” Should read: “p = 0.948”