Author's response to reviews

**Title:** The effectiveness of very early rehabilitation for acute ischaemic stroke patients in Japan: a nationwide retrospective cohort survey

**Authors:**
- Hiroki Matsui (ptmatsui-tky@umin.ac.jp)
- Hideki Hashimoto (hidehashimoto-circ@umin.ac.jp)
- Hiromasa Horiguchi (hiromasa-tky@umin.ac.jp)
- Hideo Yasunaga (yasunagah-jyo@h.u-tokyo.ac.jp)
- Shinya Matsuda (smatsuda@med.ueh-u.ac.jp)

**Version:** 2 **Date:** 16 June 2010

**Author's response to reviews:** see over
Dear Reviewers and the Editorial Office

We would like to express our sincere appreciation for valuable comments from distinguished reviewers to improve our manuscript. We have addressed our responses to the comments point-by-point, and revised the manuscript accordingly, as follows. We hope our revisions successfully meet reviewer’s comments.

**Reviewer: Dr. Eline Lindeman**

1. *The question posed by the authors is defined, however the title concerns 'effectiveness' of VEI, whereas the introduction states (final sentence) that the 'efficacy' is studied. Which term is right?*

We appreciate your clarifying suggestion. As we understand, efficacy is the effect of treatment identified under idealized conditions with selected patients and providers, while effectiveness is the effect of treatment under usual conditions of practice. Since our study intended to see the effect of VEI in a variety of hospitals and patients, we should have used “effectiveness” rather than “efficacy” in the sentence.

(P6, L3-6)

**Original:** Using a larger observational dataset and proper statistical methods such as IV would, therefore, allow us to determine the efficacy of VEI treatment, which shapes the purpose of this study.
Revised: Using a larger observational dataset and proper statistical methods such as IV would, therefore, allow us to determine the effectiveness of VEI treatment, which shaped the purpose of this study.

2. In my opinion ADL at discharge and length of stay are both relevant outcome parameters (which are not independent), however only ADL (measured with the insensitive mRS) is evaluated. I suppose data on length of stay are available, because they are used for calculation of training intensity.

We agree that the length of stay can be the marker of treatment effectiveness of VEI, since VEI could achieve better ADL at discharge and potentially shorten the length of hospital stay, or LOS. However, we chose not to use LOS as an outcome for following reasons. As OECD Health Data shows, Japanese hospitals exhibit longer LOS compared to hospitals in other OECD countries, which is attributed to undifferentiated hospital functions between acute and chronic care, and consequent “hospitalization due to social reasons.” That is, LOS is often determined not only by medical conditions of the patients but also by patient’s social conditions such as availability of informal care in the household, gender roles, and socioeconomic resources in the regions. Since we did not have in our dataset relevant information to control for patient’s socioeconomic conditions, we gave up including LOS in our analysis. We briefly added a following description in the discussion section.

(P20)
Length of hospital stay (LOS) is also an important outcome measure of treatment effectiveness. LOS is also regarded as an confounder because more severe cases may require a longer LOS. However, as OECD Health Data shows [23 (OECD Health Data (2003); available from URL: http://www.oecd.org/dataoecd/10/20/2789777.pdf /[Accessed 2007 Nov 14].)], Japanese hospitals exhibit significantly longer LOS values compared to hospitals in other OECD countries, which is attributed to undifferentiated hospital functions between acute and chronic care, and consequent ‘hospitalisation due to social reasons’ [24 (Ogawa, N., and R.D. Retherford. 1997. Shifting costs of caring for the elderly back to families in Japan. Population and Development Review, 23: 59-94.)]. LOS is determined not only by patient medical conditions but also by patient social conditions such as availability of informal care in the household, gender roles, and regional socioeconomic resources [25 (Ohwaki K, Hashimoto H, Sato M, Tamiya N, Yano E. Predictors of continuity in home care for the elderly under public long-term care insurance in Japan. Aging Clin Exp Res. 2009;21:323-8.)]. Since information to control for patient and regional socioeconomic conditions was not present in the database, we chose not to use LOS in our analytic model.
3. Table 1 is hard to read, please present the data in a way which is easy to understand for the reader.

We apologize for messy appearance of the table. We also had similar advice from other reviewers. We revised the table.

4. Table 3 shows that much more variables than VEI are related to ADL at discharge. Some remarks should be made in the results (and discussion section).

Following your advice, we added some descriptions on the effects of variables other than VEI in the result and discussion section.

(P14)
Original: Table 3 presents the results of single probit and BVP analyses of VEI on ADL at discharge.
Revised: Table 3 presents the results of single probit and BVP analyses of effects on mRS at discharge. Less disability is associated with younger age, lower mRS at pre-admission, lower functional severity score, higher functional capability score, lower training intensity, and administration of VEI.

(P20)
Original: In our analysis, training intensity during hospitalization was unexpectedly and negatively related to ADL at discharge, which requires some discussion.
Revised: In our analysis, every covariate was associated with mRS at discharge as expected, except for training intensity during hospitalisation, which was negatively related to mRS at discharge, a finding that requires some discussion.

We also received advice to revise the title of Table 3 from other reviewers. We change the title “Effect of VEI on ADL at discharge” to “Results of probit model analysis to predict mRS ≤ 1 at discharge”.

(P39)
Original:- Effect of VEI on ADL at discharge
Revised: -Results of probit model analysis to predict mRS ≤ 1 at discharge
5. **in the results section data are presented that are also (exactly) shown in the tables. Such redundancy should be prevented.**

   Thank you for your advice. We remove redundant numbers from the text.

6. **the discussion can be shortened, especially the arguments on page 21 and start of 22.**

   We edited and shortened the corresponding paragraph.

(P21)

**Original:** Ishida et al. [24] reported that patients who were expected to have better ADL recovery were less likely to receive intensive rehabilitation. Previous studies with observational data may over-estimate the effectiveness of training intensity due to this reverse causality with insufficient risk adjustment. In the current study, training intensity was higher among those patients with a higher severity score (data not shown). Another explanation might be that training intensity was confounded by characteristics of the hospitals. Previous studies included a limited number of institutions with relatively high quality care. Their findings may be limited to those institutions with specific characteristics. In the current study, we included 294 institutions that should better reflect the general situation of acute care for stroke in this country. We conducted a sub-analysis by hospitals, and preliminarily found that the average functional outcomes in those hospitals with relatively low training intensity tended to be lower than in those hospitals with higher intensity treatment, even after adjusting for the functional severity of patients at admission. These hospitals with low intensity training and poorer outcomes might result in our negative estimate of training intensity. Indeed, when we stratified the patients by the availability of specialty consulting, we observed a differential impact of training intensity on the functional outcome of patients at discharge. Since the availability of specialty consulting is differentially determined across institutions, the above sub-analysis results may support this argument.

**Revised:** Ishida et al. [27] reported that patients who were expected to have better ADL recovery were less likely to receive intensive rehabilitation. Another explanation might be that training intensity was confounded by characteristics of the hospitals. We conducted a sub-analysis by hospitals, and preliminarily found that the average functional outcomes in those hospitals with relatively low training intensity tended to be worse than in those hospitals with higher intensity treatment, even after adjusting for the functional severity of patients at admission (data not shown). Besides, when we stratified the patients by the availability of specialty consulting, we observed a
differential impact of training intensity on the functional outcome of patients at discharge.

7. the influence of length of stay should be discussed.

We have presented our rationale for not including LOS as an outcome in this study as a response to your second comment. We also did not include LOS as a covariate to explain mRS at discharge, since LOS is strongly related to patient’s severity on admission and treatment trajectory, and severe cases would need longer LOS. Longer LOS, in turn, would relate with better outcomes because longer LOS allows recovery. We added some discussion on the difficulties and complexities of including LOS in our model.

8. In the abstract the results description is not very sound (how to interpret a marginal effect of 0.153?); the sentence: we failed to detect... should be replaced by a more 'scientific' text.

The marginal effect represents a change in the outcome variable for a unit change of a targeted explanatory variable. We revised the sentence to follow your suggestion.

(P2)

Original: we failed to detect any significant association between VEI and in-hospital mortality,

Revised: There was no significant association between VEI and in-hospital mortality,

Original: We found that VEI for acute stroke patients was significantly and positively associated with improved activities of daily living (ADL) at discharge, even after considering endogenous problems due to treatment selection (marginal effect = 0.153, p < 0.001)

Revised: We found that VEI for acute stroke patients was significantly associated with a lesser degree of disability at discharge. Even after considering endogenous problems due to treatment selection, VEI improved the chance of reducing disability by 15.3%.

9. In the methods section, data source is stated: data were collected in a standardized ELECTRIC format, I suppose another term should be chosen.

Our apologies for wrong use of English words. We should have used “electronic”.

Reviewer: Dr. Julie Bernhardt

General comments
The modified Rankin Scale is considered a measure of global disability rather than a specific measure of ADL. Rather than referring to improvements in ADL, I believe you should refer directly to the mRS in all tables and be very clear about the scale you are referring to in the text. At present this is quite jumbled within the document. Also see my comments on the extent of disability across the scale - you need to modify this as well. Instead of saying “improved ADL” say “less disability”. Please amend throughout. Compulsory revision

Thank you for important advice about modified Rankin Scale. Following your advice we replaced the word “ADL” with modified Rankin Scale, and used “less disability” or “better mRS” instead of “improve ADL” throughout the text.

Although Endaravone is favoured in Japan, it is not in common use elsewhere in the world. However, given the common practice of prescribing Endaravone it is appropriate that the authors included it as a factor within this study. No action

We appreciate your understanding.

1. Abstract
1.1 I think your final conclusion that these data support the effectiveness are perhaps a little overstated. This is not an efficacy study and the jury are still out with respect to VEI, you also had some strange findings related to intensity of training which you have attempted to explain in the discussion. I would be in favour of something less strident and more tempered here eg. “These data suggest that rehabilitation commenced within 3 days of stroke may lead to better outcome with no increase in adverse events compared to delayed rehabilitation” or something similar. Minor essential

Following your suggestion, we would like to revise the conclusion as follows.

(P3)
Original: The results support the effectiveness of VEI on functional outcomes and safety even after correction for treatment selection bias, at least among stroke patients with a moderate severity.
**Revised:** These data suggest that VEI may lead to a better outcome with no increase in adverse events compared to delayed rehabilitation.

2. Introduction

2.1 One of the difficulties in the field of early rehabilitation is a lack of clear definition. It is important that you define what you mean by very early rehabilitation early in the manuscript – both in terms of the timing of onset and the content. Please do this early in the manuscript. Minor essential.

We should have provided a clearer definition already. We basically followed the definition in Bernhardt, et al. [ref 7], which defines the very early intervention as “commencing within 48 hours following stroke onset,” and “delivered with the aim of reducing the time from stroke onset to first mobilisation, and increasing the amount of out of bed physical activity” In our dataset, we did have the information on the day of stroke onset and commencement of the intervention, but unfortunately we did not have the detailed contents of the intervention. Thus, in our study, we practically defined the VEI as “any type of rehabilitation commenced by physical or occupational therapists within 3 days after admission for the acute cases admitted within 1 day after onset.” We added above description in the introduction sentence, and moved the description of our practical definition of the term in the top of method section.

(P4) introduction

**Revised:** Bernhardt et al. proposed to define VEI as intervention commencing within 48 hours after stroke onset [6,7]. They also considered VEI to be any intervention delivered with the aims of reducing the time from stroke onset to first mobilisation and increasing the amount of out of bed physical activity [6,7].

And we newly provide the definition of VEI in this study in method section.

(P8)

**Original:** Our main explanatory variable, VEI, was defined as “yes” when patients received any type of rehabilitation by physical or occupational therapists within 3 days after admission.

**Revised:** Our dataset included the information on the day of stroke onset and the commencement of the intervention but not on details of the intervention. With the available data, we defined VEI in this study as any type of rehabilitation performed by physical or occupational therapists within 3 days after admission for acute cases admitted within 1 day after onset.
2.2 Page 6. “instrumental variable (IV) methods have been used to overcome potential selection bias due to the presence of unobserved confounding factors in the observational data” – provide an example (brief) here, with reference. Discretionary

Thank you for your advice. Following your advice, we added some sentences in the introduction section to provide a brief example of the method.

(P6)

**Original**: instrumental variable (IV) methods have been used to overcome any potential selection bias due to the presence of unobserved confounding factors in the observational data, and to provide consistent estimates of the association between treatment and outcome for several conditions and treatments [11-14].

**Revised**: instrumental variable (IV) methods have been used to overcome potential selection bias due to the presence of unobserved confounding factors in the observational data and to provide consistent estimates of the association between treatment and outcome for several conditions and treatments [12-15]. Simply put, IV is a randomly occurring variable associated with the administration of treatment but is not related to outcomes. By using an appropriate statistical model that includes IV, we could statistically control for selection bias and better evaluate treatment effectiveness. For example, McGuire et al. used IV methods to assess the impact of early surgical intervention on the prognosis of femoral neck fractures [13]. The indication for early surgical intervention may depend on patient characteristics, which in turn affects patient outcomes. Thus, simple comparison between only the treatment and control groups leads to biased results. The authors used the day of admission, which should randomly occur, as an instrument to statistically control this selection to distinguish between the two groups.

3. Methods

3.1 Page 10. “Statuses of VEI administration...” unclear please expand – do you mean functional severity and functional capability. Both of these need further definition, in brackets provide the factors that were incorporated into each of these. Minor essential

We apologize for our misleading use of the word. We simply meant the administration of VEI, and corrected the word.
4. Results

4.1 Page 13. 17.6% had small (not moderate) functional deficit – if you mean they had an mRS =0 or 1. This would not be classified as a moderate deficit. 0 is no deficit, 1 = “no significant disability”. An mRS of 1 can therefore not be classified as moderate. This is used twice on page 13 and needs correction.

Thank you for important advice. We change the words accordingly

(P13)

Original: moderate functional deficit (mRS ≤ 1)
Revised: small functional deficit (mRS ≤ 1)

Original: A moderate ADL deficit (mRS ≤ 1)
Revised: A small functional deficit (mRS ≤ 1)

5. Tables

5.1 Page 35. Table 1 – I found this table difficult to follow. I suggest that instead of adding codes to the results column you indent the categories of relevance and then align the data with the category. Not all categories need to be included- for example if you are including patients in the gender column, just put female (or male) and add data. This table does not work as it is currently set out. Footer should read standard ‘deviation’ not division. Essential

Thank you for helpful advice. We also received same advices from other reviewers. So we changed the table 1 following your advice. We also correct legend words. Again our apologies for careless mis-spelling.

5.2 Table 2. Column labels of 0 and 1 are not helpful. Suggest you change to With VEI, Without VEI or VEI, no VEI to help clarify for readers. Please also refer to my general comments for changes to ‘ADL’ categories. Minor essential

Thank you for helpful advice. We changed the table 2 following your advice.
Reviewer: Dr. Michael Leathley

Reviewer's report:

Major compulsory revisions

1. The title needs to be altered. A cross sectional survey cannot really assess effectiveness. It could explore associations. The title should also consider the sample. Yes it is acute stroke patients but it those with moderate disability on admission. It is also only ischemic stroke patients.

We appreciate your clarifying suggestion. As you say, it is very difficult to assess the effectiveness by the cross sectional survey. We should describe our study as retrospective cohort study because we follow the patient from onset to discharge using existing clinical data. Following your advice, we change the title as follow.

(P1)
Original: The effectiveness of very early rehabilitation for acute stroke patients in Japan: a nationwide cross-sectional survey
Revised: The effectiveness of very early rehabilitation for acute ischaemic stroke patients in Japan: a nationwide retrospective cohort survey

2. The introduction needs to be re-written, particularly with regards to the AVERT2 trial. The study cited is a feasibility study and as such it is disingenuous to write that it lacks robustness and generalizability. There are then further negative comments around other studies - a small sample with lack of generalizability: yes this is likely but it is not just the sample size that leads to lack of generalizability. Then a study with 1,716 is also criticised because of "selection bias" but there is no information about how there was bias. More detail is needed around the observed and unobserved clinical characteristics on which the use of VEI appears to depend. There could be more detail here on how VI compares with other methods - what makes it so good? - why is it a "proper statistical method". The study will not allow assessment of efficacy (see comment re effectiveness).

We apologize for our inappropriate description of previous studies. As you mentioned, and we have acknowledged, the AVERT Phase2 study explored safety and feasibility of very early mobilization (VEM), and the investigation of intervention efficacy was left to an on-going larger Phase 3 study. We did not mean to devalue previous studies, but intended to clarify what to be overcome in study design and analytic model to reach scientifically sound evaluation of the
effectiveness of VEI. Due to the nature of observational data where treatment assignment was not randomly determined, simple adjustment for observed clinical characteristics of patients by multiple regression analysis cannot eliminate the impact of selection bias. (e.g. please refer to “Newhouse JP, et al. Econometrics in outcomes research: the use of instrumental variables. Annu Rev Public Health 19:17-34, 1998.”) The indication of treatment may depend on patient’s characteristics, which in turn affects patient outcomes. Thus simple comparison between treatment and control groups only leads to biased results. If we could have a variable which should randomly occur, we could use it as an instrument to statistically control this selection bias to purify the difference between the two groups.

We revised our description in the introduction as follows to make clearer what we meant. We also added some technical explanation on instrumental variable method.

(P4)

Original;

There is an on-going large randomized controlled trial in Australia to determine the efficacy of VEI among acute stroke cases [6]. The preliminarily analysis of the trial (A Very Early Rehabilitation Trial for stroke (AVERT) phase 2; AVERT2) suggested the effectiveness of VEI on functional status and safety [7]. As this was only a preliminary report with a very small sample size (N = 71), the results suffered from a lack of robustness and generalizability. The recent Cochrane systematic review of VEI concluded that the effectiveness of VEI remains to be established [8].

To date, there have been many observational studies that investigated the association of the early initiation of rehabilitation and the outcome of patients. Hayes et al. indicated that stroke patients who received VEI were likely to have a better recovery of their walking ability with a shorter stay in hospital, although this study again suffered from a lack of generalizability due to its small sample size (N = 30) collected in a single institute [9].

In their study of 1,716 patients admitted to several Italian rehabilitation hospitals, Musicco et al. reported that stroke patients whose rehabilitation started within 7 days of the onset were likely to have a better functional status 6 months after discharge [10]. Although their study captured a broader array of patients, their analysis with observational data faced a fundamental difficulty due to “selection bias” [11].

The use of VEI treatment is dependent on the clinical characteristics of the patients and other factors, both observed and unobserved, which in turn influence subsequent health outcome, cost, and health utility. Thus, it is difficult to disentangle the effects of the treatment per se from those of unobserved confounders when we use observational data. In economic studies, and recently in health services research, instrumental variable (IV) methods have been used to overcome potential
selection bias due to the presence of unobserved confounding factors in the observational data, and to provide consistent estimates of the association between treatment and outcome for several conditions and treatments [12-15]. Using a larger observational dataset and proper statistical methods such as IV would, therefore, allow us to determine the efficacy of VEI treatment, which shapes the purpose of this study.

Revised

There is an ongoing large randomised controlled trial in Australia to determine the efficacy of VEI among acute stroke cases [6]. The phase 2 study of the trial (A Very Early Rehabilitation Trial for stroke (AVERT) phase 2; AVERT2) revealed the safety and feasibility of VEI [7]. The recent Cochrane systematic review of VEI, however, concluded that the efficacy of VEI remains to be established [8].

To date, many observational studies have investigated the association between early initiation of rehabilitation and patient outcomes. Hayes et al. indicated that stroke patients who received VEI were likely to have better recovery of walking ability with a shorter hospital stay, although the generalisability of this finding may be limited due to its small sample size (N = 30) that was collected in a single institute [9].

In their study of 1,716 patients admitted to several Italian rehabilitation hospitals, Musicco et al. reported that stroke patients whose rehabilitation started within 7 days of onset were likely to have better functional status 6 months after discharge [10]. This study captured a broader array of patients. Because of the nature of an observational study, however, their analysis may not fully eliminate selection bias in treatment assignment even with adjustment for patient conditions by ordinal multiple regression analysis [11].

The use of VEI treatment is dependent on the clinical characteristics of the patients and other factors, both observed and unobserved, which in turn influence subsequent health outcome, cost, and health utility. Thus, it is difficult to disentangle the effects of treatment per se from those of unobserved confounders when we use observational data. In economic studies and recently in health services research, instrumental variable (IV) methods have been used to overcome potential selection bias due to the presence of unobserved confounding factors in the observational data and to provide consistent estimates of the association between treatment and outcome for several conditions and treatments [12-15]. Simply put, IV is a randomly occurring variable associated with the administration of treatment but is not related to outcomes. By using an appropriate statistical model that includes IV, we could statistically control for selection bias and better evaluate treatment effectiveness. For example, McGuire et al. used IV methods to assess the impact of early surgical intervention on the prognosis of femoral neck fractures [13]. The indication for early surgical
intervention may depend on patient characteristics, which in turn affects patient outcomes. Thus, simple comparison between only the treatment and control groups leads to biased results. The authors used the day of admission, which should randomly occur, as an instrument to statistically control this selection to distinguish between the two groups. Using a larger observational dataset and proper statistical methods such as IV would, therefore, allow us to determine the effectiveness of VEI treatment, which shaped the purpose of this study.

3. Some further clarity around how many hospitals provided data would be useful - is it 975 or 294 and if the latter, were they biased because of the voluntary agreement to join?

We apologize for ambiguous description. 294 hospitals have been participated in our analysis. As you pointed out, because of voluntary participation, there might be a possibility of sampling bias. We compared hospitals who submitted both claim and clinical data (and were included in the analysis N=294) with those who submitted claim data only (N=681). Patients mortality was relatively higher in the excluded hospitals (3.6% vs 4.8%, P<0.001), and average age was older in the excluded hospital (72.1 vs 72.7, P<0.001). There was no difference in gender composition. The average size of included hospitals was slightly larger than that of excluded hospitals. In general, included hospitals were more likely to be high-volume large general hospitals specialized in acute care of stroke while those excluded were more likely to be smaller private hospitals providing mixed services of acute and chronic care. Thus, the presented results in our analysis may be applicable only to those larger acute hospitals, and whether the VEI exhibits the similar effect on patient’s functional outcome in a broader range of hospitals needs further investigation. We added the discussion above in the text.

(P20)
Revised; (limitation section in Discussion)

Due to voluntary participation, there might be a possibility of sampling bias. We compared hospitals who submitted both claim and clinical data (and were included in the analysis, N = 294) with those who submitted claim data only (N = 681). Patient mortality was relatively higher in the excluded hospitals (3.6% vs. 4.8%, p < 0.001), and average age was older in the excluded hospital (72.1 vs. 72.7, p < 0.001). There was no difference in gender composition. The average size of the included hospitals was slightly larger than that of the excluded hospitals. In general, included hospitals were more likely to be high-volume large general hospitals specializing in acute care of stroke while the excluded hospitals were more likely to be smaller private facilities providing a mix of acute and chronic care services. Thus, the presented results in our analysis may be applicable only to the larger acute hospitals, and whether VEI exhibits a similar effect on patient functional outcome.
in a broader range of hospitals needs further investigation.

4. There is mention of speech therapy on page 8, but not on page 9 - if a patient received speech therapy was this counted as VEI?

We thank you for your clarification. We did not count speech therapy as the VEI, since we thought speech therapy will not affect our target outcome of functional status at discharge. We revised the sentence in page 8.

(P8)
Original: In this study, we included those patients who were hospitalized for an acute stroke event and had utilized any type of rehabilitation services during their hospitalization such as physical, occupational, or speech therapy.
Revised: In this study, we included those patients who were hospitalized for an acute stroke event and had utilized rehabilitation services of physical and occupational therapy during their hospitalization.

5. A comment around how patients are diagnosed would be informative. One can be confident of diagnosis where patients are diagnosed by a stroke physician or neurologist and entered onto a stroke register. If the sample is identified from hospital discharge information - how certain can we be that they are all stroke patients.

We added some description about the validity of diagnosis coding in the claim database we used. The administrative claims database that we used for this study is composed of data compulsorily submitted from those acute care hospitals under the case-mix based reimbursement policy. The enrollment to this newer reimbursement policy is voluntary basis, though most of acute care hospitals in this country have joined the system. In this database, the diagnoses of the diseases that were related to major resource use during the hospitalization are recorded by the physicians in charge with referring to the medical charts. Reimbursement rate is set according to registered diagnosis. The administrative claims database has several potential biases including possibility of over- or under-reporting due to deceived coding, as are the same as any administrative claims database such as US Medicare claims database or the US Nationwide Inpatient Samples. We added these comments in the Method section.

(P7)
Original:
The data used in this study were derived from a nationwide cross-sectional observation study conducted in 2007 for the purpose of developing a case-mix classification system for acute care and a related reimbursement schedule [16].

Revised:

The data used in this study were derived from a nationwide survey conducted in 2007 for developing a case-mix classification system for acute care and a related reimbursement schedule in Japan [16]. The administrative claims database that we used for this study was composed of data compulsorily submitted by the acute care hospitals under the case-mix-based reimbursement policy. Although enrolment in this newer reimbursement policy is on voluntary basis, most of the acute care hospitals in this country have joined the system. In this database, diagnoses of diseases related to major resource use during hospitalisation are recorded by physicians in charge while referring to medical charts. The reimbursement rate is set according to the registered diagnosis. The administrative claims database has several potential biases, including the possibility of over- or under-reporting due to deceived coding, which is the same as any administrative claims database, such as the US Medicare claims database or the US Nationwide Inpatient Samples.

6. There is no mention about length of hospital stay being used in the analysis. Status at discharge could be related to length of stay.

We received same advice from reviewer 1. We found that including LOS in our analytic model is very confusing because shorter LOS can be a marker of treatment effectiveness since VEI could achieve better function at discharge and shorten the length of hospital stay, while LOS is also regarded as a confounder to the outcome. Length of stay is determined not only by patient’s conditions such as co-morbidity and disease severity, but also by the process of care. In Japanese case, more complicated is that average LOS in this country is extremely longer than that in other OECD countries, as we depicted in the response to the comment from reviewer 1. The longer LOS can be attributed to the nature of medical care delivery system in this country, or to socio-demographic factors related to informal care provision in households (which is often called as “hospitalization due to social reasons.” We briefly mentioned issues above in limitation section in the discussion. Please refer to our response to Reviewer1, comment 1.

7. The modified Rankin is not really a useful measure in hospital because it questions a person's ability to self care. It is not possible to know if hospital how well a person looks after themselves because they are being looked after by other people. It is appreciated that this measure cannot be changed but it should be discussed. It is more a measure of participation than ADL.
Thank you for this important advice. We received similar advice from reviewer 1, as follows:

From reviewer 1

“The modified Rankin Scale is considered a measure of global disability rather than a specific measure of ADL. Rather than referring to improvements in ADL, I believe you should refer directly to the mRS in all tables and be very clear about the scale you are referring to in the text. At present this is quite jumbled within the document. Also see my comments on the extent of disability across the scale - you need to modify this as well. Instead of saying “improved ADL” say “less disability”. Please amend throughout.
Compulsory revision”

We revised the sentence following the advice. We replaced the word “ADL” with modified Rankins Scale, and used “less disability” instead of “improve ADL” throughout the text.

Additionally, we add some description about mRS in the discussion.

Revised: (Discussion section)

This study has some limitations. We used mRS as a measure of disability levels. Since mRS was originally a measure of a person’s ability to self-care, it may not be a useful measure in hospitalised cases. Our dataset did not include detailed information on patients’ ADLs, such as Functional Independence Measure (FIM). The mRS scores may be less sensitive to changes in functional levels during hospitalisation. In spite of this, we found a significant impact of VEI, which still supports the effectiveness of VEI to alleviate a patient’s disability after acute stroke attack.

8. It is unclear why and how the 2 independent summary scores are created.

Our dataset included several measures of functional levels and severity. These measures had strong correlations with each other. If we put all these measures in our multivariable regression models, strong correlations would lead to biased estimations because regression estimation assumes independence among explanatory variables. One possible solution is to choose a specific measure and give up others, at the expense of losing explanatory power of the model. Another solution, which we took in our analysis, is the use of principle component analysis to integrate the information of measures into a fewer numbers of independent scores. We hope this clarifies what we did and
why. We added a following sentence to help the interpretation of the variables.

Revised; (method section, Definition of variables)

The larger functional severity score indicates the severer functional disability, and the larger functional capability means the better functional levels. These scores have no unit.

9. Table 1 is unclear and needs to be revised.

Our apologies for messy table configuration. We revised the Table to show the baseline characteristics information in a clearer and more concise manner.

10. The statistics employed are difficult to penetrate. Could the authors simplify the statistics or make them clearer to the average reader?

Following your advice, we changed the description of statistics so that we can bring the interpretation of the numbers in English.

(P3)

Original: We found that VEI for acute stroke patients was significantly and positively associated with improved activities of daily living (ADL) at discharge, even after considering endogenous problems due to treatment selection (marginal effect = 0.153, p < 0.001)

Revised: We found that VEI for acute stroke patients was significantly associated with a lesser degree of disability at discharge. Even after considering endogenous problems due to treatment selection, VEI improved the chance of reducing disability by 15.3% (p<0.001).

(P14)

Original: The marginal effect of VEI on ADL at discharge was larger in the BVP model compared to the single-probit model (0.048, p = 0.007 in the single-probit model, and 0.153, p < 0.001 in the BVP model).

Revised: The estimation of the single probit model shows that VEI improved the chance of reducing the disability at the time of discharge by 4.8% (p=0.007). BVP model showed an even larger marginal effect of VEI, the reduction by 15.3% (p<0.001)

11. Remove the negative comments about AVERT2 in the discussion. They are inappropriate given that it is a feasibility study.
We apologize for inappropriate description about AVERT2. We removed the sentence from the discussion following your advice.

Minor Essential Revisions

1. The data are presumably in electronic rather than electric format.
   
   Our apologies for wrong use of English. We change the word.

2. The symbol for functional severity needs to be in the Table.

   The functional severity score comprises a statistically integrated quantity called factor score using principal component analysis. The score is a summary of several measures including mRS at admission, consciousness levels, degrees of communication disorders, Brunnstrom stages, and swallow disorders, as we already mentioned in the text and table legends. The number has no unit, and the larger score of functional severity simply means severer status.
Reviewer: Dr. Chris Sutton

Reviewer's report:

• Major Compulsory Revisions
  None.

• Minor Essential Revisions
  The author can be trusted to make these. For example, missing labels on figures, the wrong use of a term, spelling mistakes.

  Our apologies for misspelling and other failures in the figures. We carefully proofread again and revised accordingly.

1) Review notation for ICD-10 codes and correct inclusion of both I67$ and I679.

  We apologize for inconsistent description in disease code list. We should have used I675, instead of I67$.

  Original: (G45$, G46$, I63$, I65$, I66$, I67$, I679, I693, or I978)
  Revised: (G45$, G46$, I63$, I65$, I66$, I675, I679, I693, or I978)

2) Remove ambiguity regarding mRS=2 as cut-off (cut-point is between 1 and 2; stating cut-off as 2 does not indicate whether 2 is allocated to the lower or higher group.

  Our apologies for ambiguous description on the cut-off point of mRS. We revised as follows for clarification.

  Original: ADL at discharge was dichotomized using mRS = 2 as a cut-off point.
  Revised: The mRS was dichotomized as 1 when mRS was 0 or 1, and as 0 otherwise.

3) Under ‘Statistical Analysis’, 2nd paragraph amend to: ‘The day of admission should be related to service use variation, but not to outcomes except through service variation’.

  We appreciate your suggestion, which more precisely describes our analysis,
Original: The day of admission should be related to service use variation, but not to outcomes.
Revised: The day of admission should be related to service use variation, but not to outcomes except through service variation.

4) Revise terminology regarding ‘weekend admissions’; Friday is not normally viewed as ‘weekend’ – surely using the term ‘Friday admission’ throughout would be more appropriate?

   Again, our appreciation for your comments to make our description more precise. We explicitly used “Friday admission” instead of “weekend admission” throughout the text.

5) Explain to reader that (marginal) effect size is not easily interpretable from a (bivariate) probit model in either Methods or Discussion.

   We receive similar advice from other reviewers. We changed the sentence following the advice.

(P3)
Original: We found that VEI for acute stroke patients was significantly and positively associated with improved activities of daily living (ADL) at discharge, even after considering endogenous problems due to treatment selection (marginal effect = 0.153, p < 0.001)
Revised: We found that VEI for acute stroke patients was significantly associated with a lesser degree of disability at discharge. Even after considering endogenous problems due to treatment selection, VEI improved the chance of reducing disability by 15.3% (p<0.001).

(P14)
Original: The marginal effect of VEI on ADL at discharge was larger in the BVP model compared to the single probit model (0.048, p=0.007 in the single probit model, and 0.153, p<0.001 in the BVP model)
Revised: The single probit model estimation shows that VEI improved the chance of reducing disability at the time of discharge by 4.8% (p = 0.007). The BVP model showed an even larger reduction, 15.3% (p < 0.001).

6) Results (2nd para): mRS of <=1 is not ‘moderate’ – replace by a more appropriate term (e.g. ‘mild’).

   We receive exactly the same advice from other reviewers. We changed the sentence
following the advice.

(P13)
Original: moderate functional deficit (mRS ≤ 1)
Revised: small functional deficit (mRS ≤ 1)

Original: A moderate ADL deficit (mRS ≤ 1)
Revised: A small functional deficit (mRS ≤ 1)

7) Results (Stratified analysis sub-section ...): Replace term ‘insignificant’ by ‘non-significant’.
   We made a revision accordingly.

8) Discussion (para 3): “... might suffer from overestimation of the causal effect of treatment ...”.
   We made a revision accordingly.

(P17,L14)
Original: However, this previous finding might suffer from overestimation of the treatment due to a selection bias since early intervention was more likely to be administered to patients with mild to moderate severity, rather than to those with severe conditions such as coma.
Revised: However, this previous finding might suffer from overestimation of the causal effect of treatment since early intervention was more likely to be administered to patients with mild to moderate severity, rather than to those with severe conditions such as coma.

9) Discussion (para 4): “... recently conducted randomised controlled trial ...”.
   We made a revision accordingly.

(P17,L18)
Original: An early report from a recently conducted randomized control study
Revised: An early report from a recently conducted randomised controlled trial

10) Discussion (para 6): “... suggesting that VEI was not likely to lead to death.”
    We made a revision accordingly.

(P19,L5)
Original: suggesting that VEI was not likely to lead to an adverse outcome.
Revised: suggesting that VEI was not likely to lead to death.

11) Discussion (para 10): “... the average functional outcomes in those hospitals with relatively low training intensity tended to be worse ...” (lower mRS is better!).

We made a revision accordingly.

(P21, L11)

Original: We conducted a sub-analysis by hospitals, and preliminarily found that the average functional outcomes in those hospitals with relatively low training intensity tended to be lower than in those hospitals with higher intensity treatment, even after adjusting for the functional severity of patients at admission.

Revised: We conducted a sub-analysis by hospitals, and preliminarily found that the average functional outcomes in those hospitals with relatively low training intensity tended to be worse than in those hospitals with higher intensity treatment, even after adjusting for the functional severity of patients at admission.

12) Improve layout of Table 1 and change notation for variable contents (e.g. from 2:CI/1< to 2:CI: >1)

We received a similar advice from other reviewers, and apologize for the messy configuration. We revised the table.

13) Missing ‘s’ from ‘variables’ in caption for Table 1.

We revised the table accordingly.

14) Correct captions got Tables 3 and 4: these are not solely ‘Effects of VEI’ but include effects of other variables.

We receive similar advices from other reviewers. We change the title “Effect of VEI-” to “Results of probit model analysis to predict-”.

(P39)

Original: Effect of VEI on disabiity recovery at discharge

Revised: Results of probit model analysis to predict mRS ≤ 1 at discharge
Original:- Effect of VEI on in-hospital mortality
Revised: -Results of probit model analysis to predict in-hospital mortality

15) Review age scale used in Table 4 – estimate and CIs effectively meaningless.
We revised the table following description.

<table>
<thead>
<tr>
<th>Age (by 10 years)</th>
<th>Estimate</th>
<th>CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.002</td>
<td>(0.000, 0.004)</td>
<td>0.002 (-0.000, 0.003)</td>
</tr>
</tbody>
</table>

16) Study flowchart: Replace term ‘Missing variable’ by ‘Missing data’.

Following your advice, we revise the figure 1.

• Discretionary Revisions

1) Discussion (para 5): Rephrase “Thus, treatment of ... effect of VEI.” to clarify meaning (by improving language).

We apologize for unclear English in the sentence. Since we found the sentence is redundant, we just remove it, and modified a following sentence.

Original: Thus, treatment against selection bias by IV was relevant to purify the effect of VEI. The positive effect of VEI on functional outcome at discharge remained significant even after consideration for endogenous problems due to treatment selection.

Revised: Thus, treatment against selection bias by IV was relevant to purify the effect of VEI. The positive effect of VEI on disability outcome at discharge remained significant even after consideration for endogenous problems due to treatment selection using the IV method.