Article details: 2015-0088	
Title	Mechanical thrombectomy in patients with acute ischemic stroke: a cost-utility analysis
Authors	Xuanqian Xie, Anna Lambrinos, Brian Chan, Irfan A. Dhalla, Timo Krings, Leanne K. Casaubon, Cheemun Lum, Nancy Sikich, Aditya Bharatha, Vitor Mendes Pereira, Grant Stotts, Gustavo Saposnik, Christina O'Callaghan, Linda Kelloway, Michael D. Hill
Reviewer 1	Lauren Cipriano
Institution	Massachusetts General Hospital, Institute for Technology Assessment, Boston, MA
General comments (author response in bold)	In this cost utility analysis of intravenous thrombolysis plus mechanical thrombectomy compared to intravenous thrombolysis alone. The authors find that MT+IVT has an incremental cost effectiveness ratio of $12,000 per QALY-gained compared to IVT alone and therefore is good value for money in the Canadian public payer system. Strengths of the model include the model calibration efforts and the transparency in their presentation. Unfortunately some of their analysis does not adhere to best practices in cost effectiveness modeling or in calibration and there are some tables which must contain errors in presentation (Table 3), but I believe all of these are correctable with substantial effort. **Major Issues (issues which must be addressed)** **1.** The model overview section implies that the authors performed a meta analysis of 5 RCTs to estimate the outcomes of each intervention alternative. However, it seems that this meta- analysis is actually a separate paper also under review. The meta analysis should be included in the data sources and assumptions description, but the authors should remove claims that it is part of this paper from the methods (page 6 and 8) and the discussion (or fully include the relevant meta analysis methods, analysis, and discussion). **Both meta-analysis and economic evaluation were part of a health technology assessment by a single group of researchers at Health Quality Ontario. In the revision, we cite a reference for this health technology assessment.** 2. The model overview section does not present a clear description of the patient population: **a.** "we assumed an age range similar to the RCTs (mean age of 65-70 years old)"… but what age is your base case? **Clinical outcomes in the first 3 months of our model were from a meta-analysis of five RCTs (mean age of 65-70 years old). The longer-term (>90 days) outcomes in our model were based on the Oxford Vascular Study. We did not use the age-specific Canadian Life Table, so we did not provide a specific age for our target patients.** **b.** "more than 70% of patients in the RCTs…" so are your strategies pure strategies (comparing MT+IVT to IVT alone) or mixed (comparing 80% MT+IVT/20% IVT alone to 70% IVT alone/30% no treatment)? Using ITT analysis to inform outcomes estimates is acceptable, but this is a very confusing way to present your strategies. Perhaps this detail can be moved later in the methods so it can be more fully (or clearly) explained. **We agree. We deleted the sentence from the Methods section: "More than 70% of patients in the RCTs received IVT in both study arms, and more than 80% of patients received mechanical thrombectomy in the MT+IVT arm."** **c.** Please present a clear description of the base case patient cohort. **In accordance with reviewers' comments, we edited the base case cohort on page 6. "The mean age of patients ranged from 65 to 71 years of age, and there was an equal proportion of men and women. [9-13] Patients had occlusion of either an internal carotid artery or middle cerebral artery, and eligibility for mechanical thrombectomy was confirmed by imaging and established clinical criteria. [13] Patients were functioning independently in the community before the stroke."** **3.** Assumptions (page 7 and 8). The third and fourth assumptions contradict each other. If disability affects risk of mortality and quality of life, then similar annual rates of ICH do not cancel each other out as implied by assumption 4 because the two interventions have different rates of disability at 90 days. As an illustration consider a simple framework where there are two health states "Healthy" and "Sick". Intervention A results in the population being 80% H and 20% S. Intervention B results in the population being 50% H and 50% S. If the sick people die at a faster rate than the healthy people (say, 100% die after exactly 1 year), then the fact that the healthy and sick people have the same rate of ICH is irrelevant. For Intervention A, the 80% of the population (those in state H) live to be exposed to the risk of ICH; whereas in Intervention B only 50% of the population live to be exposed to the risk of ICH. Therefore, similar rates of ICH in the two intervention arms does not indicate that it is safe to ignore ICH since a different absolute number of people will be alive at each point in time to face that risk. **Our previous expression for the symptomatic intracerebral hemorrhage might not have been clear for readers. We edited the fourth assumption in page 8 to read, "The two**

treatments are associated with a similar risk of symptomatic intracerebral hemorrhage within 90 days post-stroke" by adding "within 90 days post-stroke".

4. The authors present a calculation for the QALY gain in the first 90 days in the text (page 8) and present this value in their inputs table with its own distribution forPSA:

   a) The formula is a function of inputs (the health utility at 90 days and the death rate – although it is worthwhile to note that it is not yet stated in the text that the death rate is 0.1786 so this number is initially confusing). These inputs also have distributions in the PSA. So, is the QALY increase in the first 90 days equal to ((0 + 'QALY at 90 days')/2)*0.25*(1- 'probability of death in 90 days) or is it 0.0735 and in the PSA distributed Normal(0.0735, 0.0305) regardless of the values drawn in the PSA for the all- cause mortality rate and the QALY at 90 days? I think that it should be the former, but I believe you may have done the later.

   We used a description in the text to replace the figures in the formula to prevent confusion. The edited formula for QALY gained in the first 3 months was "([(0 + utility increase at 90 days)/2]*0.25)*(1 - probability of death in 90 days)". In the probabilistic sensitivity analysis, the utility increase followed normal distribution, which was independent of the mortality rate and QALY in first 3 months.

   b) The formula itself should be reconsidered. As is, it assumes that all the individuals who survive earn an incremental utility of 0.037 (or linearly increasing from 0 to 0.074) over the 90 days. This might be sufficient for the increase in QALYs for survivors. However, this would imply that all those that die, die immediately and earn no QALYs during the 90 days. If patterns of when people die within the 90 days are known (for example if in- hospital or 28 day mortality is known), then perhaps a better approximation can be made.

   Our primary objective is to estimate the difference in QALY between two treatments. When two treatments have the same mortality rate in the first 3 months, it is reasonable to assume that the QALYs for those who died in first 3 months would be same. In addition, according to five RCTs, most deaths occurred during the first month post-stroke, and their health utility would be very low in this short period. Thus, for those who died in the first 3 months, the QALYs would be negligible in both arms.

5) Productivity costs are not part of the societal perspective and should be excluded. According to the CADTH guidelines, only friction costs should be included to value time lost from paid work. As the cohort under analysis is 65+, these friction costs would only be incurred by the fraction of the population who are employed. Because this is small, it is reasonable to exclude friction costs from this analysis, but the authors may estimate and include them if they wish. For an excellent description of why productivity costs are not included in the societal perspective see Drummond "Methods for the Economic Evaluation of Health care Programmes" 3rd edition pg 78-88. *note: unpaid caregiving should still be included*

   We deleted all analysis from the social perspective. See point 5 of our reply to the Editor for details.

6) The analysis does not conform to CADTH guidelines on analysis horizon (which encourages lifetime horizon as the reference case). Please present lifetime horizon as the base case and other horizons in your sensitivity analysis (figure A4 isgood).

   We understand that ideally the analysis would use a lifetime horizon recommended by guidelines. But the evidence was from RCTs with 90 days' follow-up, and the observational study from UK had 5 years' follow-up. In accordance with expert opinion, we decided to set 5 years' follow up as the base case to reduce potential uncertainty. Fortunately, we conducted the sensitivity analysis to cover up to 15 years' follow-up (close to lifetime).

7) Page 13, line 32 "In the Canadian…" Since total costs are increased, the increase in spending on the MT is only partially recouped through downstream savings. This is also somewhat misleading because of siloed budgets within the health care system, these extra expenditures at the hospital level will create reductions in expenditures within rehabilitation facilities. These are still improvements, but the shifting of budgets is not seamless. The second sentence in this paragraph seems arbitrary as it restates a methods point but without any additionaldiscussion.

   We deleted this paragraph from the Discussion on page 14: "In the Canadian health care context where general tax revenues pay for both acute and long term care, upfront investment in acute stroke thrombectomy services can be recouped by reduced need for long-term care of the neurologically disabled. Indirect costs such as loss of productivity and the cost of unpaid caregiving are partially accounted for in this analysis because of metrics

extracted from the Economic Burden of Ischemic Stroke study.[20]"

8) Calibration. For each time step, the calibration process has 3 inputs for 3 calibration targets which are the exact state distribution of individuals at the end time point. For example, we can write the equation for the transition from 3 months to 6 months as

$$\begin{bmatrix} .317 \\ .420 \\ .263 \end{bmatrix} = \begin{bmatrix} \gamma & 0.037 & 0 \\ 1-\exp(-R_{ab4-6}/12) & 1-0.037-pMort*RR_{bc4-6} & 0 \\ pMort*RR_{ac4-6} & pMort*RR_{bc4-6} & 1 \end{bmatrix}^3 \begin{bmatrix} .2874 \\ .5340 \\ .1786 \end{bmatrix}$$

$$\text{Where } \gamma = 1-\left(1-\exp\left(-\frac{R_{ab4-6}}{12}\right)\right) - pMort*RR_{ac4-6}$$

And where the vector on the left contains the 6 month calibration targets, the vector on the right is the position at the end of three months, and the matrix in the middle is the one-month transition matrix for months 4-6. I use pMort for the baseline probability of death which varies by age. This system of equations has multiple solutions all of which would have a SSE of exactly 0. That set of multiple solutions is mathematically defined. You do not need to search for it using a grid search. Using a pMort of 0.003, all of the following input sets have a summary goodness of fit of 0 (a perfect fit) (subject to some rounding in the numbersprovided).

Set 1: R = 0 ; RRac = 9.31; RRbc = 13.19

Set 2: R = 0.00046; RRac = 9.157; RRbc = 13.283

Set 3: R = 0.02255; RRac = 1.837; RRbc = 17.725

Set 4: R = 0.025; RRac = 0.783; RRbc = 18.365

The solutions to the optimization problem for each stage can be identified by closed form methods and there is no need to perform a grid search to identify solutions. There is a set of solutions which solve the above system exactly. Similarly, there will be a set of solutions which solve the system exactly for all other values for pMort. If you want to consider a linear combination of ages, each with their own pMort, there will also be a set of 3 (messy) polynomials to solve, but to which the solutions are mathematically defined. Optimization software will quickly be able to reveal all of the solutions which satisfy these equations, and the three for 6-12 months, 12-24 months, etc. without a computationally intensive grid search approach. If you want to incorporate uncertainty in the targets, then you can find the set of solutions varying the targets within their range of uncertainty. But, conditional on these targets, there is a simple set of solutions which can be algebraicallydefined.

a. Some of the sets above are inconsistent with the biological system (consider set 4 where the RR on mortality for the disabled population would be less than the general population). Additional constraints may need to be set on the system to generate reasonable input values. Solutions to a similar set of equations (as in, for other time periods) could result in cases where RRac > RRbc; if this is not reasonable clinically, then the constraint should be added to the system. Furthermore, the literature may reveal additional bounds – such as whether the RR of mortality for functional independence is between 1-3 or closer to 9. (After accounting for these additional constraints and incorporating additional knowledge about the system, you may find that for some periods, the system has very little to no uncertaintyremaining).

We reported the process of calibration of natural disease history for the stroke patients in the in Appendix in the previous version. In this revised version, we describe selecting parameters, selecting ranges in search parameters, and justifying the calibrated results. We added the following sentence to the introduction of Appendix 1.

"We aimed to obtain calibrated parameters with the following features:

- **They are the most common measures or statistics (e.g., relative risk and odds ratio) in epidemiology studies**
- **The values of calibrated parameters are consistent with the natural biological system (e.g., relative risk of mortality for post-stroke patients versus general population > 1)**
- **Model outputs and the observed data (i.e., Oxford Vascular Study) must be consistent**
- **The values of calibrated parameters (e.g., relative risk) are consistent with external data (e.g. a study in Australia)**

Parameters should be reasonable for projection of long-term outcomes beyond the observed period"

9. Calibration (as is). **There are various ways to calibrate parameters. We admit that the method we used might not be the best approach. Yet our calibrated parameters have all features mentioned above, so they can adequately serve our present economic evaluation.**

a) Page 13, line 57. The authors state that the calibration approach they use provided "relatively reliable parameter estimates". Calibration does not necessarily ensure reliable parameter estimates – especially in systems with many degrees of freedom and few calibration targets. The reliability of estimates from calibration should not be overstated.

We have removed the phrase "relatively reliable".

b) The authors should follow the best practices for presenting model calibration (Stout et al. 2009 Pharmacoeconomics 27(7)).

Stout et al. 2009 provided important guidance. But we followed the methods introduced by Vanni et al. 2011. Fortunately, both articles had considerable overlap. Reference: Vanni T, Karnon J, Madan J, et al. Calibrating models in economic evaluation: a seven-step approach. Pharmacoeconomics. 2011;29(1):35-49.

c) The authors do not explain why they chose absolute deviation for the measure of fit for mortality and then squared deviation for the measure of fit in the summary GOF statistic.

There are a couple of measures for goodness of fit (GOF). The observed data of mortality were accurate (no missing data and no misclassification), so we set mortality rates as the primary goal and defined a narrow acceptance range (±1%). For sets meeting this criterion, the mortality would be fairly similar, and then we move to the combined measure of GOF, including both disability and mortality at different times.

d) The authors also do not explain why they assumed weights of 1 for each calibration target in the summary goodness of fit score. Weights of 1 can create issues with scale. For example, a 1% error from a target of 31.7% at 6 months is relatively small in terms of squared error, but a 1% error from a target of 56% at 5 years is much greater. Therefore, your summary statistic will penalize more heavily mismatch on 5-year survival than the 6 month or 1 year proportion of patients in functional independence. This contradicts your confidence in these targets as you are likely more confident in your near term targets than you are in targets further into the future. The authors might consider using weights that attempt to adjust for the scale of the targets.

We agree that the weight influences the selection of best fitting set. Properly assigning weight to each calibrated target is very challenging, given targets at different observation times and varying reliability of targets (e.g., results of mortality would be more reliable). For simplicity, we assigned a weight of 1 to all targets.

e) For c & d. Choices about the measure of fit and the weights influence which parameter sets are identified as best fitting. See Taylor et al. 2010 Pharmacoeconomics 28(11) and Enns et al. 2014 Medical Decision Making 35(2).

We agree that the weight influences the selection of best fitting set. Properly assigning weight to each calibrated target is very challenging, given targets at different observation times and varying reliability of targets (e.g., results of mortality would be more reliable). For simplicity, we assigned a weight of 1 to all targets.

f) The authors do not state how they weight the 1000 inputs sets identified from their calibration process in the PSA. Are the inputs equally weighted? Wouldn't it make more sense to weight them based on the quality of overall fit such that better fitting sets are weighted more heavily?

Either using equal weight or assigning weight as a function of overall GOF is used in practice. We used equal weight in this study.

10. Please show the results of calibration input validation in more detail (page 33, line48).

We provided more details of calibration input validation in Appendix 1. "On the basis of the calibrated relative risk of mortality for the general population versus risk for function independence and disability patients in Table A1-7, and the percentage of patients of function independence and disability in the best-fitting model in Table A1-8, we estimated that the relative risk weighted by the function status were approximately 2.07, 2.16 and 2.27 at 1, 2 and 5 years after stroke, respectively. This relative risk was very close to that reported in Australia, ranging from 2 to 2.3 between year 2 and year 5."

11. Figure A1. Please add whiskers on the results from model to indicate the range across the best fitting input sets.

We deleted the Figure. Because both modelled and observed proportions of patients in three health states are reported in Table A1-8 and A1-5, we decided to delete this plot.

12. Table A6. Please be specific which of these analyses relied on observational data which has since been refuted by RCT evidence and which use technologies similar to those included in your analysis.

Treatments for patients in the Oxford Vascular study have not been reported in the articles published. It could be that the objective of these studies was to predict population-based incidence, disability and institutionalization rates. Guidelines by the National Institute for Health and Care Excellence (NICE) published in 2007 recommend that IVT should be used within 4.5 hours of onset of stroke symptoms (unless the patient has an intracranial haemorrhage). However, as early as 1996, the American Heart Association stated that "Intravenous r-TPA (0.9 mg/kg, maximum 90 mg) with 10% of the dose given as a bolus followed by an infusion lasting 60 minutes is recommended treatment within 3 hours of onset of ischemic stroke". We added a few lines explaining this information in Appendix 1: "Treatments for patients in the Oxford Vascular study have not been reported in the articles published. Because intravenous thrombolysis treatment was recommended by the National Institute for Health and Care Excellence in 2007, most patients in the Oxford Vascular Study might not have received IVT therapy. (13)". References:

National Institute for Health and Clinical Excellence. Technology appraisal guidance: Alteplase for treating acute ischaemic stroke (review of technology appraisal 122). London and Manchester; 2012 [cited 2015 Nov 16]. Available from: https://www.nice.org.uk/guidance/ta264/documents/stroke-acute-ischaemic-alteplase-review-of-ta122-final-appriasal-determination-guidance2

Adams HP Jr, Brott TG, Furlan AJ, et al. Guidelines for thrombolytic therapy for acute stroke: a supplement to the guidelines for the management of patients with acute ischemic stroke. A statement for healthcare professionals from a Special Writing Group of the Stroke Council, American Heart Association. Circulation. 1996;94(5):1167-74.

13. Model validation. There are many models which include analysis similar to the IVT arm of this study. A good model validation would include a comparison to these findings – are the life year, QALY, and lifetime costs estimates similar to IVT arms in other CUAanalyses?

We focus on the internal validity of our model. The long term survival of patients in the IVT group in our model was similar to that in Oxford Vascular study, the main source of our data input. See Appendix 3 for details.

The evidence for long-term outcomes in acute ischemic stroke is sparse. The stroke patients' long-term outcomes have substantially improved over time (Rothwell et al 2011), so it is inappropriate to use historical data to validate our model. The costs in our model were also not necessarily in accordance with those in other studies, which were strongly related to the location and the perspective of analysis. Of course, it is not difficult to compare results in our model with those in other models. But, given the model outputs are determined by the inputs, we do not think that we can use the outputs from other models to validate our model. Fortunately, our model inputs (e.g., the calibrated parameters) have been validated externally by the studies in Australia.

Reference: Rothwell PM, Algra A, Amarenco P. Medical treatment in acute and long-term secondary prevention after transient ischaemic attack and ischaemic stroke. Lancet. 2011;377:1681-92.

14. Table 1-1.

    a. Please include ranges for the calibration inputs (what is the range across the 1000 randomly selected sets)?

The calibrated model inputs in Table 1-1 were derived from Table A1-7 and Table A1-2 (Life Tables). We reported the ranges of 1,000 randomly elected parameter sets in Table A1-7 in this revision.

We added "And the ranges were reported in Table A1-7" in text, and updated Table

A1-7 (See below).

Table A1-7: Values of the Best-Fitting and Good-Fitting Parameter Sets

Parameter	Value in best-fitting (range of 1,000 good-fitting) parameter set	Definition
$R_{ab4\text{-}6}$	0.392 (0.34, 0.44) per patient-year	Annual disability rate from functional independence to disability for months 4 to 6 post-stroke
$R_{ab7\text{-}12}$	0.267 (0.23, 0.28) per patient-year	Annual disability rate from functional independence to disability for months 7 to 12 post-stroke
$R_{ab13\text{-}24}$	0.161 (0.16, 0.20) per patient-year	Annual disability rate from functional independence to disability for months 13 to 24 post-stroke, i.e., at 76 years old
OR_{ab_age}	0.830 (0.83, 0.92)	Odds ratio of age for risk of disability
$RR_{ac4\text{-}12}$[a]	2.646 (2.1, 2.9)	Relative risk of mortality versus the age-specific general population for patients with functional independence for months 4 to 12 post-stroke
$RR_{bc4\text{-}12}$[a]	7.57 (7.5, 8.2)	Relative risk of mortality versus the age-specific general population for patients with disability for months 4 to 12 post-stroke
$RR_{ac13\text{-}60}$	1.035 (1.0, 1.1)	Relative risk of mortality versus the age-specific general population for patients with functional independence for months 13 to 60 post-stroke
$RR_{bc13\text{-}60}$	2.899 (2.6, 3.0)	Relative risk of mortality versus the age-specific general population for patients with disability for months 13 to 60 post-stroke

[a]Because values of time-dependent parameters in 4- to 6-month and 7- to 12-month groups were fairly close, we combined them.

b. Costs "from the societal perspective". Remove the direct health care costs and make the unpaid caregiver costs transparent.

We removed information on the cost from the societal perspective. See our reply to comment 5 from the editor.

c. Cost for end of life care. Please explain in the methods text which patients get this cost and when. Is this given in the month of death to everyone who dies (including the 17% who die in the first 90 days)? This is a very high cost for a fatal acute stroke. In Table 3, it appears that the cost of death are only included in sensitivity analysis. This does not seem appropriate. Please make assumptions about the cost of death in the first 90 days, and subsequently, more clear in the text and in the tables.

Cost for end-of-life care was included in one sensitivity analysis. It applied to mortality after 90 days post-stroke. We added this piece of information in Table 1-2, "Cost of end-of-life care for death after 90 days post-stroke ($CAD)". Also, we have provided the calculation of cost for the first 3 months in the footnote to Table 1-1.

The cost for end-of-life care would apply to all patients. An intervention could postpone but not avoid these costs, so the only issue is discounting of cost at different times, which often has marginal impact on final results. Thus, it is appropriate to exclude these costs in the base case.

15. Probability distributions for PSA.

a. All-cause mortality with distribution Beta(64, 294) is listed twice. Are these inputs treated as independent in the PSA? I would suggest independence as assuming that the mortality is exactly equal is a strong assumption; whereas, allowing for mortality to vary, but with equal means,

probably better accounts for true uncertainty. Also note, this relates back to item 4a above because with independence equal mortality would not always be the case and should then be accounted for in the calculation of incremental utility in the first 3 months.

In our simulation, we used a shared mortality rate for both groups. Following one reviewer's suggestion, we tested the independent morality rate of two treatments with equal means. Although the standard deviation of incremental cost and incremental QALY were greater than that using a shared mortality rate, the results (probability of cost-effectiveness for Mechanical Thrombectomy) changed only marginally.

b. Why do the authors assume a gamma distribution for costs? The gamma distribution might be appropriate for the costs of individuals in the system (which are very right skewed), but the central limit theorem is pretty strong and so it is likely that the uncertainty around the mean value (the purpose of PSA) is normally distributed. Were the person-level observations in the BURST study so skewed that CLT doesn't apply?

Both Gamma distribution and normal distribution should be fine for the cost data. We used the Gamma distribution in this case.

c. The distributions used for some parameters are overlapping, but a rank order should still apply. For example, the annual costs of functional independence should alwaysbe less than the annual cost of disability and the utility of functional independence should always be greater than the utility of disability. Do the authors incorporate any limits to prevent values that violate logical rank orders?

The utilities (mean [95% CI] of 0.71 [0.68, 0.74] for functional independence and 0.31 [0.29, 0.34] for disability) and costs (mean ± SE of $1,384 ± 277 for functional independence and $3,080 ± 616 for disability) for functional independence and disability states had almost no overlap with our study (See Table 1-1), so it is not necessary to incorporate the logical rank order issue in the simulation.

16. Table 3 presents negative ICERs which are very confusing as they have two possible (but very disparate meanings). When the incremental cost is negative and in the incremental QALY is positive, the ICER for MT+IVT should be labelled "Dominates" or "Dominant strategy". When the incremental cost is positive and the incremental QALYs are negative, the ICER for MT+IVT should be "Dominated". The values presented in the table for time horizon are not consistent with Figure A4. I would suggest the table be expanded:

Scenarios	Societal perspective			Payer perspective		
	Incr. Cost	Incr. QALY	ICER	Incr. Cost	Incr. QALY	ICER

The negative signs in Table 3 were typos. We corrected them in this revision. Also, we have used the term "dominant" in this Table.

Minor Issues

1. Since the 2015 annual inflation rate is not yet available, please specify the month in 2015 which was used for inflation adjustment.
   We edited the sentence: "Costs are expressed in April 2015 Canadian dollars ($CAD)." In page 10.
2. Model validation should be briefly described in the methods (in the overview is fine) or in the results, but right now it is just tacked onto the first paragraph of the results without any information or context.
   In the revised manuscript, we mention that "The model validation can be found in Appendix 3" in the end of the Methods section.
3. The comparison to other CEA in the literature is awkwardly phrased. This should be re-written.
   We removed the discussion of economic evaluation for older generations of mechanical thrombectomy and Appendix 6.
4. The appendix contains many unnecessary phrases. Eg. Page 28 line 6. "we were able to estimate" "we estimated"… many "should have" and "ideally" occur throughout. Page 30, last paragraph is very awkward. Ideally, the authors would improve the readability of the appendix.

   We edited those paragraphs slightly, and removed the "ideally", "should have" etc.
5. Calibration. The authors repeatedly misuse the word "convergent". A random grid search does not "converge" and it doesn't result in identifying "convergent parameters". It may

	reveal multiple parameter sets satisfying pre-determined fit criteria. **The term "convergent" referred to meeting the acceptable goodness of fit. It was used in some publications of calibration. To avoid confusion, we changed "convergent" or "convergence" to "acceptance" or "good-fitting".** 6. I could not match your best fitting numbers in calibration, but it may be because Table A1-2 does not contain the baseline monthly probability of death for individuals in your initial cohort (age 65-70). Please extend this table to include ages 65-75. **The mean age in the Oxford Vascular study was 75 years old, so we provided the Life Table from 75 to 89 years old in Table A1-2. We gave an example of parameter estimation. The probability of death in one month was 0.003027 for 75-year-olds in the general population (Table A1-2), and the relative risk of mortality for functionally independent patients in months 4 to 12 post-stroke was 2.646 (Table A1-7), so the calculated p value of mortality per month was 0.008 for functionally independent patients for months 4 to 12 post-stroke (Table 1-1). We admitted the age differences between the RCTs and the Oxford Vascular study was one of the main limitations of our study.** 7. Table A1-7 would benefit from a brief text description of each input parameter. **We defined each parameter in Table A1-7.** 8. The statement that you randomly selected 1000 input sets for PSA is stated twice (on page 32 and 33) **We removed the duplication on page 33.** 9. Figure A3-1 and Figure A3-2 can be removed; they are not informative (and they do not constitute model validation). **We focused on the internal validity of the present study. These Figures showed that the model output reflect our inputs.** 10. Table 3. Ages for sensitivity analysis should be clear. Sensitivity analysis for discounting should consider 0%, 3%, and 10% **We reply to the issue of age in point 2a above. Also, our sensitivity analysis presented the ICER by age groups, ≤ 70 years old and > 70 years old. We have included the discounting rates of 0, 3% and 10% in this revision in Table 3.** Table 3. Consider including additional sensitivity analyses. **Yes. We conducted some additional analysis in this revision.**
Reviewer 2	Alastair Buchan
Institution	University of Oxford, Acute Stroke Programme, Nuffield Dept. of Clinical Medicine
General comments (author response in bold)	1. Since only two of the RCTs (SWIFT PRIME and EXTEND-IA) mandated IV tPA in the inclusion criteria, while the 3 other studies (ESCAPE, MR CLEAN and REVASCAT) tested MT against "best medical management" that may or may not include IV tPA, is it possible for the authors to calculate the cost-effectiveness for MT with best medical treatment without IVT? **Of the 3 studies (ESCAPE, MR CLEAN, REVASCAT) that did not mandate IVT in the inclusion criteria, 13%-32% did not receive IVT in the intervention arm and 9%-22% did not receive IVT in the control arm. We were able to examine participants who were IVT eligible and ineligible on the outcome of functional independence (mRS 0-2) in the ESCAPE and REVASCAT studies (data for MR CLEAN were not available). The subgroup analysis is shown below. The subgroup difference was not significant (p = 0.72). Thus, the economic implication for those without IVT should be similar to that for the base case.** **We added a sentence to in the methods section on page 12. "Given no significant differences in functional independence were found among subgroups of status of IVT (P = 0.72) and occlusion site (P = 0.94), the analyses for those subgroups were not conducted."**

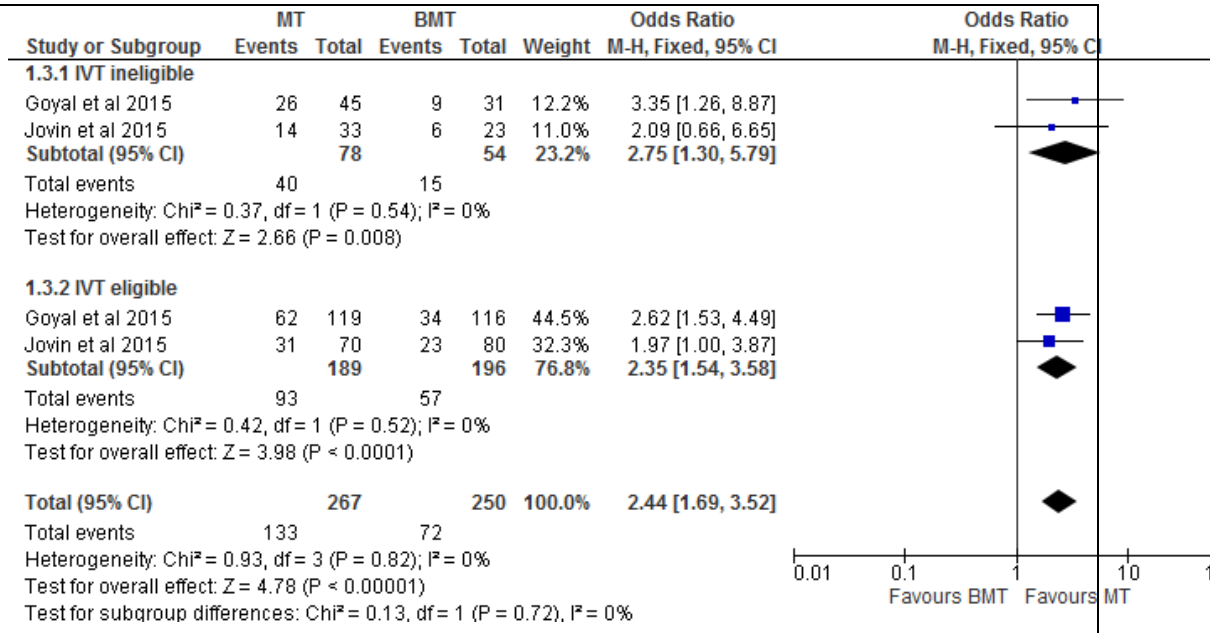

Figure: Mechanical Thrombectomy Versus BMT on the Proportion of Functionally Independent Patients at 90-Day Follow-up by Status of IVT

2. The authors stated "The pooled estimate of the adjusted beta coefficient in the linear regression in 2 RCTs [8,10] showed that MT+IVT increased health utility by 0.074 (95% confidence interval (CI): 0.014,0.133) at 90 days, compared with IVT alone." Is there any reason why these two RCTs (MR CLEAN and REVASCAT) were used?

Only those two RCTs reported the EQ-5D utility.

3. Although the authors have stated that one of the limitations of their model input was the short interval follow-up (90 days) in the 5 RCTs identified and the need to combine results with a cohort study to model longer term outcomes". Nevertheless, these are two different cohorts in the sense that in the Oxford Vascular Study, a large cohort study from the United Kingdom the patients were not treated with MT or IVT. Also, this was a cohort of patients with TIA or minor strokes.

See our reply to comment 3 from the editor above.

Reviewer 3	Scott Sloka
Institution	St John's, Newfoundland
General comments (author response in bold)	1. The abstract is clear and has the appropriate detail. **Thanks!** 2. The introduction is also clear and concise. The population statistics are useful, but it might be enhanced if an estimate of the number of people that might actually benefit from this procedure (ie the percentage of all strokes [62000] that are candidates for MT who receive tPa and then fail this first line treatment and are therefore a candidate). This would give the reader a target size for the usefulness of this new intervention. One might have to be geographically close to a center that can use this MT device, and we are a geographically spread out population. **We added a potential target size for this intervention to the introduction on page 5. "About 87% of strokes are ischemic, and 20% of those are caused by large vessel occlusion in the internal carotid artery and middle cerebral artery. Therefore approximately 8,700 people per year may be eligible for endovascular treatment in Canada."** 3. The meta-analysis includes MT with or without IVT vs IVT and/or best medical therapy, and their decision tree includes MT+IVT vs IVT alone. The authors might comment on whether this is a valid review to make a comparison. Were they able to separate out the 70% and 80% subsets, or is the meta-analysis "mixed". **The meta-analysis results were analyzed from the intention to treat approach for the "mixed" treatment, and we attempted to reflect this issue (i.e., a proportion of patients who do not receive the planned treatment) in our economic model. But our presentation might confuse**

readers. We deleted one sentence from the methods section on page 5-6: "More than 70% of patients in the RCTs received IVT in both study arms, and more than 80% of patients received mechanical thrombectomy in the MT+IVT arm."

4. The use of UK data for clinical outcomes could be reasonable given the dataset size, although it might be useful if the authors comment on whether they feel the UK study is easily translatable to our population or whether there are some caveats to be made in using it for a Canadian analysis.

See our reply to comment 3 from the editor.

5. Is there a good reference to add to "Disability is associated with increased risk of mortality"? It makes intuitive sense, but might be better served with a reference.

We agree. We added the reference for this assumption in page 7.

Reference: Hankey GJ. Long-Term Outcome after Ischaemic Stroke/Transient Ischaemic Attack. Cerebrovascular Diseases. 2003;16(suppl 1):14-19.

6. Health utility is affected by stroke severity, co-morbidity and age. Is the MT intervention outcome also affected by these parameters? The authors did a comprehensive sensitivity analysis which likely covers these variables, but is the meta-analysis inclusive of the entire population or are there exclusion criteria in the trials that will affect the generalization of the results of this study?

Results from randomized clinical trials (RCTs) are largely specific to the patients whose occlusions are confirmed by imaging. These patients were defined as the target population of our economic model, too. We agree that patient inclusion criteria could affect the generalizability of our results. In this revised manuscript, we examined the cost-effectiveness for the more severe stroke patients (based on Interventional Management of Stroke III study). We reported the results by age group in the previous version of this manuscript (See Table 3).

7. In the interpretation with the sensitivity analysis, can the authors summarize one or two scenarios that would make the addition of the intervention unfavourable? I realize they are contained in the table and the CEAC, but these situations could be easily explained away for a general readership.

Mechanical thrombectomy could be less favourable for patients with more severe ischemic strokes. We conducted another scenario analysis using patients with more severe ischemic stroke in the Broderick et al (2015) study, examining patients with a National Institute of Health Stroke Scale score of ≥20. However, in this study more patients (25%) in the endovascular treatment group were functionally independent than in the intravenous thrombolysis group (14%) (adjusted odds ratio, 1.97; 95% confidence interval, 1.09–3.56). Again, these results were seen without any statistically significant increase in mortality between groups (28.8% mRS 6 in endovascular treatment group vs. 34% in intravenous thrombolysis group). We updated the data inputs in Table 1-2 and results in Table 2. We also added the information in the text.

Methods section, on page 12, "We also analysed the scenario of stroke patients with severe neurological deficit (National Institutes of Health Stroke Scale score, ≥20), on the basis of pooled results from the Interventional Management of Stroke III and Multicenter Randomized Clinical Trial of Endovascular Therapy for Acute Ischemic Stroke in the Netherlands trials."

Footnote to Table 1-2 on page 24, "This study [30] did not find a statistically significant difference in mortality between groups (28.8% mRS 6 in the endovascular treatment group vs. 34% in the intravenous thrombolysis group). Thus, we assumed no survival benefit of mechanical thrombectomy in this scenario analysis."

Results section, on page 13, "For patients with severe stroke, assuming no improvement in mortality, the ICER was increased to $81,651 with QALY gained of 0.106 and with the incremental cost of $8,691."

Reference: Broderick JP, Berkhemer OA, Palesch YY, et al. Endovascular Therapy Is Effective and Safe for Patients With Severe Ischemic Stroke: Pooled Analysis of Interventional Management of Stroke III and Multicenter Randomized Clinical Trial of Endovascular Therapy for Acute Ischemic Stroke in the Netherlands Data. Stroke. 2015. [Epub ahead of print]

8. Is there a reference stating that the outcomes at 90 days continue to be valid for a certain duration afterwards?

Unfortunately, no reference states that the outcomes at 90 days continue to be valid for a certain duration afterwards for the new generation mechanical thrombectomy therapy. But,

	Patients' long-term health outcomes (i.e., more than 3 months after a major stroke) are static after endovascular treatment (Palesch et al, 2015), and IVT therapy (Kwiatkowski et al, 1999). Therefore we included the following: **"Patients' long-term health outcomes (i.e., more than 3 months after a major stroke) would be conditional on their health status at 90 days (i.e., functional independence or disability)" as an assumption of the economic model in page 8.** **Reference: Palesch YY, Yeatts SD, Tomsick TA, et al. Twelve-Month Clinical and Quality-of-Life Outcomes in the Interventional Management of Stroke III Trial. Stroke. 2015 May; 46:1321-7. Kwiatkowski TG, Libman RB, Frankel M, et al. Effects of tissue plasminogen activator for acute ischemic stroke at one year. N Engl J Med. 1999 Jun 10;340:1781-7.**