PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	What intervention is best practice for vestibular schwannomas? A systematic review of controlled studies.
AUTHORS	Wolbers, John; Dallenga, Alof; Mendez Romero, Alejandra; van Linge, Anne

VERSION 1 - REVIEW

REVIEWER	Michael Gleeson, MD, FRCS, FRACS, FDS. Professor of Skull Base Surgery The National Hospital for Neurology & Neurosurgery, Queen Square, Londo
REVIEW RETURNED	06-Aug-2012

GENERAL COMMENTS	The authors have attempted to answer a question that has confounded those managing patients with vestibular schwannomas for the last 25 years. Like those who have preceded them, they have found the available evidence both lacking and flawed. In the end, their conclusion, stated boldly in the title "Radio-surgery is best practice for medium-sized vestibular schwannoma", has been based on two studies with level 2b evidence and five studies with level 3b evidence. As they so correctly point out, the Cochrane Central Register of Controlled Trials has no further studies. The paucity of evidence in the literature has deterred many from providing it, simply because their conclusions would be less than helpful. This submission does not make easy reading for someone who practices in the field, let alone anyone who has limited medical knowledge but might be seeking advice on which to base their choice of treatment. It contains statements that some might find derogatory and inflammatory. For example, "A recent survey ..... showed that 70% (of patients) were informed only about microsurgery and not on the radio-surgery option." This is certainly not the practice in the UK or in other countries where each patients' best interests are considered by multi-disciplinary teams. The main outcome measures defined in this study as predictors of best outcome, and hence best clinical practice were : mortality, surgical or anaesthesia-related complications, facial nerve function, preservation of hearing, quality of life, speed of return to employment and health-related costs. All these appeared to favour radio-surgery. It is assumed that the reported studies used the same selection criteria that would be considered by a multi-disciplinary team, before advising their patient on a recommended treatment plan. But this is not the case, only one study clearly defined the indications for intervention. In the other 6 studies, "just having a vestibular schwannoma seemed sufficient to initiate an intervention." Let it not be forgotten that less than 33% of patients who present with vestibular schwannomas ever require active management in

	their lifetime, other than interval clinical and radiological review. The efficacy of radiosurgery can only be properly assessed after about 5 years. Tumour expansion is common, indeed expected, after radio-surgery and is seen to develop for about 1-2 years after treatment. Tumour size may then stabilise, decrease or, in about 15% increase, denoting in the latter that therapy has failed and the tumour has in fact been continuing to grow all the time. The selection of six studies in which radio-surgery follow-up was less than 5 years, and one study in which more than 20% of patients were lost to follow-up detracts from the conclusions of this review. No-one doubts that surgery is without complications or has inherent risks to life and limb. Surgeons take that responsibility very seriously and it is a part of the consent process. But, the long term complications of failed radio-surgery cannot be ignored either. The surgery for patients with tumours that continue to grow following radio-surgery is not simple, rarely results in a complete resection and carries significant risk for the function of adjacent cranial nerves. This potential outcome that affects up to 15% of patients has not been addressed or considered. There are symptoms that define the most appropriate treatment intervention for specific groups of patients. For instance, vertigo or imbalance, symptoms that may affect patients with small tumours and very definitely those with tumours approximating 3cm within the cerebello-pontine angle. None of these patients are helped by radio-surgery and surgical removal or near total resection can help the majority. It should not be forgotten that it is precisely those symptoms that prevent many from returning to work and have a significant impact on their families and those who care for them. Quality of life studies abound in outcome studies. Most rely on simple generic scales, for example the SF36 and Glasgow Benefit Inventory, as in the studies selected for this review . There isn't a disease-specific tool for vestibular schwannoma which would give some degree of robustness to any outcome statement. Those who have examined this topic find that anxiety and depression are pivotal to perceived quality of life. Put simply, patients do not like having tumours inside their heads and it is for this reason that many with tumours eminently suitable for radio-surgery elect for surgical management. Patient choice, accurately informed, is something encouraged by most health systems. Furthermore, the costs of radio-surgery are usually understated as the ongoing expense of interval scans and failed therapy is never included. Those working in this field recognise the difficulties of controlling trials for patients with vestibular schwannomas. The radio-surgery arms are usually dominated by patients with relatively small tumours while the surgical arms contain an abundance of patients with much larger tumours. Linear measurements of tumour size fail to express the difference in volume adequately between a small tumour of 1cm and one measuring 3cm, a real difference in terms of volume of more than 25 times. The range and median size of tumours treated in the selected studies has not be quoted and is not available. Why is size so important? Simply because symptoms referable to the trigeminal nerve feature prominently in patients with tumours in excess of 2.5cm. Facial anaesthesia is unlikely to respond to radio-surgery while many with facial paraesthesia or anaesthesia are

	made better by surgery. Furthermore, and perhaps more important, failed radio-surgery in a medium-sized tumour of 3cm isn't admitted or recognised until the tumour is at least 3.5cm. By this time, the tumour is no longer a medium-sized tumour but a large one about to become a giant tumour. Morbidity and mortality parallel tumour size and it is the patient that risks this. There is no doubt that radio-surgery can help a large number of patients with vestibular schwannomas. Likewise, surgery also has its place. Combined treatments are likely to become more popular in years to come and will probably improve outcomes for our patients. Systematic reviews relying on small numbers of low evidence series are not going to help, particularly when the title is superficial at best and clearly misleading at worst. The British Skull Base Society has wrestled with this problem for decades. It now has a national registry that documents every new patient identified by the contributing, multi-disciplinary, centres in the UK. Almost 1000 patients were accrued in its first year, a figure which is already re-writing the textbooks on the incidence of this relatively common tumour. Much more reliable information will become available in the years to come and this will determine what might be better regarded as "best practice." In the meantime, best treatment will continue to be delivered by ethical multi-disciplinary skull base teams who consider the best interests of their patients and can provide fully informed consent.

REVIEWER	Janine Dretzke Senior Systematic Reviewer Unit of Public Health, Epidemiology & Biostatistics University of Birmingham Birmingham, UK I have no conflicts of interest.
REVIEW RETURNED	06-Sep-2012

THE STUDY	A table should be provided in an appendix detailing study and patient characteristics (including age/gender, disease and treatment history, co-interventions, outcomes etc.) Search strategy: Why were search terms relating to the specific interventions not used? Where any MeSH terms used? How many hits were there in Cochrane Central? Were validated study design filters used? Did you consider contacting experts in the field? A sample search strategy (e.g. for MEDLINE) should be provided as an appendix. Further, a PRISMA flow diagram should be completed and provided. Existing evidence. Were any other (systematic) reviews identified? Is this the first in this area? You mention a meta-analysis in the discussion-what types of studies did this include? Study selection: What were the inclusion and exclusion criteria? These should be specified a priori, particularly with regard to the

	most important outcome measures and adverse events. Why was a study reporting only costs included, and was there a systematic search for studies reporting costs? How were disagreements resolved? What were the main reasons for exclusion? Quality assessment: It might be more correct to state that you devised your own quality criteria, but were guided by the criteria listed in the SIGN checklist for cohort studies. You dichotomise studies into trustworthy (4 studies) or not (3 studies). You need to acknowledge that this is to some extent an arbitrary cut-off point of quality (how may 'yes's' do there need to be for a study to be considered trustworthy?) It would be better to decide which the most important quality criteria are and then to make a judgement on the studies. Data synthesis. You need to state why it was not appropriate to synthesise results across studies (if this is the case) and why analysis was qualitative rather than quantitative. In some instances the language needs to be clearer in order for the meaning to be unambiguous: Examples: Abstract, study selection: "Seven prospective and retrospective studies" Page 3: "..causing possible advantage for the surgery arm." Microsurgery or radiosurgery? It's also unclear to me which way this bias might be operating and why. Why would older patients being allocated to the radiosurgery arm make this seem more effective? P3: "four studies remained that showed a trustworthy association between intervention and outcome" Do you mean that there were four studies that were more likely to give unbiased results? The language overall needs to be checked. Examples: "quality of live" (p5) "..the mortality rate is much smaller and occurs many years later in a patients' life." (p6) "if an intervention wisely should not to be postponed." "there are no proven parameters predicting a tumour to grow" (p2)
RESULTS & CONCLUSIONS	Table 1. For outcomes, "radiosurgery is better" is not that informative-though you may not want to list all the outcomes, perhaps you could state to how many outcomes this applies (e.g. 4/5) and whether the difference is statistically significant. I'm not sure that this table is actually needed as most of the relevant information is also contained in Table 2 (or the text). Table 2. What are the different outcome measures used for quality of life and are they comparable across studies? There need to be more references in the text (results and discussion), so that it is clear which study or studies are being referred to, e.g. p3, "the one study focused on intervention costs"- which one? It states on p4 that "after radiosurgery there was no mortality and no surgical or anaesthetic complications..." but in Table 2, there do

	appear to be complications associated with radiosurgery. Some of the arguments in the discussion seem a bit confused, e.g. it states that "the validity of high-quality observational studies is demonstrated by remarkable similar results in randomised and observational studies when comparing treatments. Such studies may provide trustworthy information on the risks of the intervention..." Are you trying to say that the validity of observational studies is best confimed by RCTs giving similar results? But you say earlier that there are unlikely to be RCTs in this area. This needs to be re-phrased for clarity (p5): "..but the disadvantage was at the side of the best outcome." There is some repitition, e.g. (p5) "Only Van Roijen et al did not report on intervention..." and further down "Only Van Roijen et al did not report on clinical.." This needs some further clarification:"In two studies there were co-driven (?) interventions, evoking a relevant weakness to the confidence of the outcome" Which studies? Are these better or poorer quality studies? How is the result likely to be affected? This also needs clarification: "Although only one study clearly defined the starting point of an intervention..." Which study, which intervention? How is the indication likely to act as a confounder? How does this sentence relate to the previous sentences? "Therefore, the overall assessment of study quality gave..." (p5) There should be some detail in the discussion about similarity of patients in the studies to patients to whom the intervention might be applied. The discussion needs to be rewritten in a more stuctured way, with a focus on how specific biases identified are likely to influence the results.
REPORTING & ETHICS	The PRISMA checklist should be used when reporting a systematic review.

VERSION 1 – AUTHOR RESPONSE

Response to the comments of prof. Gleason

Indeed the question on best practice is pending for about two decades since radiosurgery became established next to tumour excision. It needs a new look as exemplified by the derogatory statement - as perceived by the reviewer - on 70% of patients not been informed on radiosurgery by their doctors. This statement is not from these authors, however, but a citation from the referenced German study. When learning this, we also experienced the same shock as the reviewer and it urged us to write this article. Therefore we start our discussion paragraph with that statement. All the same, it seems a proper study and we are not aware of any studies from the UK or elsewhere addressing this subject. The reviewer provides no foundation – other than being shocked - for it not being true. Above and beyond one would suppose in Germany also multi-disciplinary teams and individual doctors act in each patients' best interest. The article by Muller et al triggered us to do our research, but we have only a trivial dilemma with skipping the opening lines of our discussion paragraph.

It is a misunderstanding, that if no class 1 evidence from randomised, double-blinded studies with

narrow confidence intervals is available, the evidence for best practise is lacking and flawed. For example, there was never a controlled study to proof that using an operating microscope is of benefit to the patient in various surgeries, such as in vestibular schwannomas. Yet nobody is asking for a randomised study, because it feels sensible that using dedicated light and magnification improve circumstances and the outcome of the surgery. Likewise, we expect that no level 1 proof will ever become available for vestibular schwannoma interventions. One of the class 2b trials (Myrseth, 2009) started off as a randomised trial, but patients – receiving standardised written study information - quite sensible felt microsurgery and radiosurgery not to be equal in terms of strain and risk. They refused participation and made their own choice. Next best evidence, however, does exist as we concluded in this manuscript based on two class 2b and two class 3b studies. These four studies showed little risk to bias in Appendix 1 (new Appendix 2) of the manuscript. More over all four consistently pointed to radiosurgery as best intervention. This is not contradicted by the reviewer.

We searched to answer a specific, restricted question on the management of patients with solitary vestibular schwannomas. The restrictions are the six eligibility criteria.
Having come to our conclusion, we also stated that it feels sensible to choose for the non-invasive, less risky, less expensive, in short the most patient-friendly intervention in an out-patient setting. And the more so, since the clinical outcome is the same or better. In fact the same common sense is advocated by the reviewer stating that no-one doubts that microsurgery has inherent risk to life and limbs. This study objectifies this feeling.

Rightly, there is a difference between selection criteria and outcome measures. Both have been separately addressed and presented in the Appendix 1 (new Appendix 2) and in Table 2. The common denominator for selection was comparative intervention studies on newly-diagnosed, sporadic vestibular schwannoma. Comparative studies not complying with these 5 inclusion criteria were excluded. In this revision not reporting on clinical outcome became an exclusion criterion also. For various quality reasons in the end four studies remained that showed trustworthy association between the two interventions and outcome. Our conclusion on best practice is based on these four studies, which emerged to concern only tumours less than 3 cm cisternal diameter.

Indeed, many patients will not need an intervention. Those patients are not part of our study. We researched the best practice, if an intervention is considered necessary by the involved doctors and patients. And yes there is no research and no consensus on when to intervene. This study, however, does not interfere with the indication to start an intervention.
Tumour progression was mentioned once as starting point in the comparative studies of this research. Not mentioning a specific reason to start treatment does not disqualify the comparison of (two) interventions proviso the same disease at the same stage. We valued on this in new Table 1 and the revised discussion paragraph.

The four comparative studies have a mean follow-up of 3.5 -5.9 yr. Not included in detail in the first version of the manuscript, we checked the results of the comparative studies with those of large, long-term case series reporting the beneficial results from centres of excellence; some already have a minimum follow-up over 5 years. We added this information in new Appendix 3. They compared well. These long-term results are beneficial of its own, but can't be compared with surgery, while to the best of our knowledge there are no long-term surveys with qualifying losses to follow-up after surgery for vestibular schwannoma.

An incidental increased tumour volume in radiosurgery series is not considered a failure. Commonly it is attributed to osmotic swelling due to tumour necrosis. The failure rate of radiosurgery (that is a second intervention needed) varied from 0-5% and compares well with the 0-18% of microsurgery. (Table 2) Next to these numbers from our review, the radiosurgery failures are in the same order 1-3% (also new Appendix 4) and 2-4% for microsurgery in the long-term case series comprehensively

reviewed by Arthurs et al. as recently as 2011.
May be the esteemed reviewer tangled his number of 15% potential reoperations after radiosurgery with the about 15% recurrent tumours after previous surgery that is reported on in various radiosurgery case series (range 6-21%: Regis 2007, Friedman 2006). Or the reviewer might have tangling up with the number of further surgeries in microsurgery patients for post-operative haemorrhage, repair of CSF-leakage, tarsoraphy and gold plate insert after facial nerve palsy, being 17% (Myrseth 2005) and even 37% (Pollock 2005), 30% (Pollock 2006) and 25% (Myrseth 2009). No feast for the patient and his surgeon.

We are not aware of controlled studies showing evidence for symptom relief due to surgery (or any intervention) in vestibular schwannomas of any size. Nevertheless, in the case of large tumours it is conceivable that (near total) resection may provide some (partial) symptoms relief. However, these kind of symptomatic tumours showed not to be part of our review. The patients in the studies had minor symptoms and certainly not incapacitating. For the quality of life assessment validated questionnaires have been used allowing for conclusions within a study. Independent from the used questionnaires all showed advantage for radiosurgery above microsurgery (Table 2). Also dedicated case series showed decreased quality-of-life after microsurgery.(See references 23-25) The nice thing of these controlled comparative studies is that it provides information on both interventions on matters as clinical outcome, perceived quality of life and also work resume.

Overall the tumours were small or medium-sized. The moderately large tumours (over 20 mm) formed the smaller third part and were not typically over represented in the microsurgery arm. In all four trustworthy studies the tumour size matched well between the arms as judged in (old) Appendix 1 and also detailed in new Table 1. No tumours over 3 cm were in the four high-quality studies and our conclusion does not deal with this category as prominently stated in the previous title of the manuscript.

Since a very small number of all patients will start off with a tumour over 3 cm and since 95-98% will not need a second treatment after radiosurgery, a negligible number will remain for salvage surgery. Moreover a second surgery after failure of first surgery might even put a patient more at risk to life and limbs than after having had radiosurgery firstly. One doesn't know from studies, but it accords to our experience that repeated surgery is cunning and risky to the patient. Likewise not known from studies, we would, however, not be surprised that repeated radiosurgery may effectively control tumour re-growth and comes with fewer risks than salvage surgery. Yet, these questions are not part of our exclusive research.

Registration of all vestibular schwannomas is good. The Danish do it already for many years and provided us with rather interesting observations – also concerning the incidence of the tumour (see manuscripts introduction) – but the registry will not answer the question on best practice. And yes, surgery has its place, be it, - according to the current evidence - a smaller one than in the past.

Response to the comments of dr. Dretzke

We added a new Table 1 with patient characteristics from the reviewed studies. The old Table 1 was removed. The column with EBM level was removed from Table 2 and included in new Table 1. Appendix 3 (old Appendix1) provides study characteristics. All patients have the same disease in the same stage, which is smaller the 30mm, same base-line symptoms and without previous intervention. The one exception is the study of Karpinos et al. We added this exception plus reference in the text. We added the mortality numbers mentioned in the text also in Table 2.

On search strategy. We started the search with the specific intervention terms and retrieved 156 studies. We noticed that one study (Regis, 2002) that we were aware of was not included. Since the

Regis 2002 study was retrieved in this extended search with more general terms we decided to spend more time browsing through irrelevant material. We sought for high sensitivity and took the relatively low precision for granted.
The authors are experienced medical specialists in the field of ENT, radiotherapy and neurosurgery. They visit at a regular base their own disciplinary congresses, dedicated meetings on skull base tumours and vestibular schwannomas and keep up-dated also through their professional literature. We think that the change that studies comparing two interventions for solitary vestibular schwannoma are missed is negligible.

We added the extra flow diagram on study selection. (new Appendix 2)

An example of the Medline search strategy is added as new Appendix 1

On existing evidence. Only systematic reviews on case series have been published. This applies also to the (wrongly called) meta-analysis of Arthurs et al. mentioned in the discussion paragraph. In this review was no quantitative synthesis involved. We corrected the text accordingly. To the best of our knowledge this is the first systematic review on studies comparing the various interventions in an intended controlled way.

On study selection. In the text we gave the eligibility criteria more emphasis and we included clinical outcome measures. Indeed we are primarily interested in clinical outcome and not in costs. Since not reporting on clinical outcome, we excluded in this revision the study by van Roijen et al.
In the end only one “trustworthy” study provided data on costs. We felt that the way we reported in this manuscript, it made transparent the availability and quality of the various clinical and economic outcome measures.
We elucidated how disagreement was resolved (by consensus).
The main reason for exclusion was not comparing interventions in controlled ways, which are all Oxford CEBM level 4 studies. (new appendix 2)

On quality assessment. Here also, we agree with the reviewer. The applicability to the intervention studies on vestibular schwannoma was judged. It resulted in the formats of Appendix 2 and Table 2, which were used for the extraction of the various measures. We adapted the text accordingly.

On the arbitrary cut-off point of trustworthy studies. Looking again at the summed number we realised a mistake in old Appendix 1. In the Myrseth 2005 study all patients were offered treatment right at diagnosis of their vestibular schwannoma and only after a failed intervention a second intervention was offered in 6% after surgery and 5%.after radiosurgery. We corrected the current Appendix 2. In the revision all trustworthy studies have no relevant biases. The (non-)relevance of encountered biases was elucidated item by item in the discussion.

We checked the conclusion of the authors of the retrieved comparative studies and scored the various biases in order to rank the trustworthiness of the conclusion. We decided not to perform quantitative synthesis, because the clinical outcome measures seem reliable within one study, but difficult to compare between studies. More over the beneficial outcomes all pointed to one intervention, radiosurgery.

Yes, we mean that four studies were more likely to give unbiased results. We gratefully added this to the text.
We hope to have improved clarity of content and language at the mentioned sites.

Table 1 was skipped.

For quality of life measures various tests were used that are mentioned in a new column in Table 2. Indeed only the same tests are comparable across studies. (This elucidates for example why quantitative synthesis in our opinion is not feasible with this review's material.)

On more references in the text. Old Appendix 1 (= new Appendix 2) provides the references/first author and year. In this revision we added these references also in the text.

By surgical and anaesthesiological complications we mean meningitis, pulmonary and urinary infection, CSF leak etc. Other than the name may suggest radiosurgery is a single session. high dose, precision irradiation and not open surgery. So the regular complication to life and limbs of invasive surgery involving general anaesthesia and use of scalpel do not apply to radiosurgery. In fact, the crucial – common sense - advantage of radiosurgery is not to be in need for anaesthesia and scalpel and yet to be similarly effective.

On confusing arguments in the discussion and re-phrasing on p5. We meant to say that high-quality observational studies might be as meaningful as RCT in circumstances that two interventions are compared, because the study population might be less selected by sharp inclusion criteria. High-quality observational studies may comply better with the overall disease population. We adapted the text and hope to have cleared the confusion.

In this revision van Roijen et al. (and its double phrasing) was excluded due to not reporting on clinical outcome.

On co-driven interventions. Only Karpinos et al. had co-driven intervention in 14-26% of their cases, that is earlier surgery for the same solitary schwannoma. (We mentioned above already the change made to old Appendix 1 concerning the Myrseth 2005 study.) A previous intervention is a relevant bias. Every second intervention comes with more risks and less beneficial outcome for the patient. In this case in up to 24% patients of the surgery arm the study-intervention was a second intervention, making it a poorer quality study. And moreover the intention of our review is to inform on newly diagnosed patients without previous treatment in order to secure reporting on the same stage of the disease. The results from first and second intervention were not separated in the article. We adapted the text accordingly.

We elucidated on the starting point issue in the discussion paragraph.
We rearranged and rewrote the discussion by following the current appendix 3 and using the offered suggestions.
We verified the 27 items of the PRISMA checklist and, if possible, addressed them in this revision.

VERSION 2 – REVIEW

REVIEWER	Professor Michael Gleeson Department of Neurotology The National Hospital for Neurology & Neurosurgery Queen Square London, W1N 3BG.
REVIEW RETURNED	03-Dec-2012

THE STUDY	This paper will require significant input from your copy editor.

REVIEWER	Janine Dretzke Senior Systematic Reviewer Public Health, Epidemiology & Biostatistics

	University of Eirmingham B15 2TT UK I have no conflicts of interest.
REVIEW RETURNED	06-Dec-2012

GENERAL COMMENTS	**2nd round comments peer review BMJ open vestibular schwannomas 6/12/12** Some comments made previously have been addressed, but not all comments regarding systematic review methodology have been addressed satisfactorily. The writing is still not very clear/accurate and the team might benefit from some assistance with this, and with systematic review methodology, if planning to resubmit again. Only a proportion of language/writing issues have been addressed below. <u>1 Existing reviews</u>. You state there are numerous existing reviews on this topic, but cite only one (Arthurs). There is an overlap for at least some studies between this and your review, so the Arthurs review is clearly not limited to only case-series. There needs to be more of an explanation of why your review is different/better compared with this existing review. <u>2 Quality assessment</u>. You need to make it clearer that you have devised your own quality criteria and used the SIGN 50 criteria as a guide only. For example, you have not looked at whether the main potential confounders are identified and taken into account in the analysis. The term "*confidence effect is due to intervention*" (Appendix 3) does not make sense. You do not state in the methods which are the more important quality criteria and why (and why some 'no's' were in bold and some weren't). Again, it needs to be stated that the cut-off for good and poor quality is somewhat arbitrary. If you want to include the Oxford EBM criteria, then you need to define what 2b and 3b mean. I'm not sure how useful this is, as you then use more detailed criteria in the cohort checklist, which also incorporates study design elements. Could you reference in the text which the four better quality studies are (without the reader having to refer to an appendix).

	Other comments 3. Page 1, article focus. Would suggest rephrasing along the lines of: Systematic review of evidence from controlled studies on the effectiveness of interventions for the treatment of solitary vestibular schwannomas 4. Page 2, line 21 *publication rate*-should this be publication status? 5. P5, line 8. This sentence does still not make sense in the context of the evidence you found, i.e. no randomised trials: *The value of high-quality observational studies is validated by the remarkable similar results, which were witnessed when comparing specific treatments through both randomized and observational trials.* 6. Page 4, line 3. *There was minimal or no losses to follow-up.* This should be loss to follow-up. 7. Page 4, line 6. The sentence: "*That is, were more likely to give unbiased results*" seems to be stuck on as an afterthought. Please rephrase. 8. Page 3, line 53*. Baseline patients' characteristics were quite similar in the study groups.* Is this both between and within studies? 9. Page 4, line 52. *Appreciating a patients' individual preference, ideally counselling is based on the outcome of high-quality clinical trials.* Does this sentence mean: Whilst taking into account patients' individual preferences, ideally the choice of treatment should be based on high quality evidence from well conducted clinical trials. Please rephrase. 10. Page 4, line 11. *After radiosurgery, there was no mortality and no surgical or anaesthetic complications, better facial function, better hearing preservation and better quality of life.* As stated in my previous comments, there do appear to be complications after radiosurgery as outlined in table 2. 11. *In general higher age, co-morbidity and larger tumours are*

	drawbacks for a good outcome. In those studies showing significant imbalance of these variables the potential disadvantage, however, was at the side of radiosurgery, being already the best outcome in these (all) studies.{Pollock, 1995 #6854}{Karpinos, 2002 #7605}{Myrseth, 2005 #7511} Therefore, we considered these imbalances as not relevant. This is an important statement but still not well worded. Are you trying to say that any potential confounders are likely to work in favour of microsurgery appearing more beneficial (or radiosurgery appearing less beneficial). 12. *Only one study clearly defined the starting point of an intervention.{Myrseth, 2009 #7707}, Nevertheless confounding by indication between the various studies appears unlikely, since major adverse events, like invalidating neurological deficits, do not occur in the natural history of vestibular schwannomas smaller than 30 mm. It is very implausible that any of the major adverse events occur in the absence of an intervention. Therefore the risk that such outcome occurs due to chance is not realistic and we assigned no relevance to defining the indication to intervene.* This is unclear, particularly the last sentence. Are you talking about the importance (or otherwise) of the intervention starting at similar points in the patients' disease progression? 13. Abstract, second sentence in background: *If intervention (microsurgery, radiosurgery or fractionated radiotherapy) becomes necessary, the preference appears to be subjective, while it might be based on research-based evidence.* Do you mean: if intervention becomes necessary, the choice of intervention appears to be driven by patient or clinician preference rather than being evidence based. (or similar) 14. Discussion, 3rd sentence: *We searched for evidence and found that radiosurgery is best practice in medium-sized tumours.* This implies that you found evidence that radiosurgery was already being used as best practice, rather than you found evidence of greater clinical effectiveness of radiosurgery compared with microsurgery (and that it should therefore become best practice).(see also key messages)

	15. Key messages. Cohort study already implies that the study is observational, you do not need both. 16. Article summary (and throughout). Stick to one term for the studies you found-either cohorts or controlled observational studies or comparative studies etc. 17. Methods. Page 2, line 23. *The retrieved articles were screened by title and by abstract if necessary.* I would assume that all studies would have been screened by title and abstract (where available) and that some hardcopies would have been screened where necessary. 18. The flow diagram indicated that 4 studies were included, while in the article 6 studies are included. Two are subsequently downgraded due to quality, but this is not clear from the flow diagram. Why are those two specific criteria (for exclusion) chosen in the flowchart when there are more quality failings? The inclusion and exclusion criteria should be pre-specified. 19. Regarding the primary and secondary outcome measures listed (page 2, line 40 ff), were these pre-specified by you, or are they reported as such in the included studies. 20. Discussion. This should start with your main results, rather than those of case series. A more detailed summary of results would also be useful, i.e. radiosurgery was found to be more beneficial in terms of xxx, over xxx time period. Did quality of life deteriorate in both treatment arms? (not clear in which direction the different qol scales work) This needs to be stated in the results. 21. Page 5, line 12. *Overall, these patients are more similar to the general disease population than those obeying to the strict inclusion and exclusion criteria of a randomised clinical trial.* Better: those complying with the strict.... But there aren't any RCTs in this area? Do you mean the patients in the observational studies are more similar to a general population than those of hypothetical trials would be? 22. Page 5, line 27. *A major scientific hazard...* Better: a major risk of

	bias.. 23. Do the results differ between the good and poorer quality studies? And how do results fit with those of the Arthurs review? 24. Page 6, line 26. *The overall assessment of study quality gave confidence in four studies, because no relevant biases were signalled*. (better: identified) It is not accurate to say that no biases were found in the four studies. 25. *Looking for best practice, one should realise indeed that the results of various health-related quality of life studies after surgery called for modesty.* Do you mean called for caution? Needs to be considered?

VERSION 2 – AUTHOR RESPONSE

Response to the comments of Dr. Dretzke

1. We intent to say that numerous 'classical' reviews are summarized in a recent article by Arthurs et al. We clarified the text accordingly. To the best of our knowledge the present study is the first study that assesses the quality of intervention studies for vestibular schwannoma in a systematic and transparent way.

2. The main confounders for this specific disease are listed in Appendix 3. We emphasized in the text that the professional team for the treatment of vestibular schwannoma judged all quality criteria that were considered relevant for this specific disease. The term 'confidence that the effect is due to the intervention' ensues from the 'Note on the use of methodology checklist 3 for cohort studies' and this Sign50 reference have been provided in the manuscript. We changed the phrase to: no relevant bias, outcome due to intervention.
We thought it more appropriate to argue in the discussion paragraph on the importance of confounders and other quality criteria in relation to the two specific interventions reported on. See also point 11.
We agree that the phrase on Oxford EBM criteria does not really add, because of the used more detailed checklist of Appendix 3. We omitted the phrase, also in table 1 and the abstract.
We added 'the upper four of table 2' in the text, being the better quality studies.

3-25. Gratefully we took advantage of the various suggestion of Dr Dretzke for rephrasing the text.

10. In table 2 the rows called RS provide the numbers on complication after radiosurgery.

11. Yes, the imbalance of confounders works in favour of microsurgery.

12. Yes, about the unimportance of the disease progression.

18. From the flowchart we can see that 6 controlled comparative intervention studies were included for quality assessment in more detail. These 6 studies are elucidated in the manuscript. Two of which subsequently failed to give confidence that the difference in outcome was due to the intervention and not to bias, as reported on in Appendix 3. Earlier, seven other studies were excluded already from this detailed quality assessment, because lacking an intervention control arm (previously called Oxford EBM level 4) or because the predefined clinical outcome measures were not reported.

19. Predefined by us when formatting table 2. Typically, in clinical studies on the treatment of vestibular schwannoma the clinical outcome measures are quite the same: mortality, cranial nerves and more general (surgical) complications.

20. Our main result is stated in the sentence: We found evidence of greater clinical effectiveness of radiosurgery compared to microsurgery in medium-sized tumours. Arthurs et al. summarised recently and in detail treatment results and needed 15 pages.
On QoL: Arrows up is better, same is same and arrows down is worse. In some studies QoL was worse in both arms (Pollock 1995 and Regis 2002). The results of interventions are best learned from table 2 and appendix 4.

21. Yes, it is a general remark. As a rule RCT do study (more or less heavily) selected patients and intent to exclude selection bias, taking care that both arms are similar, while common clinical practice possibly involve patient with different characteristics.

23. Also in the studies of lesser quality radiosurgery is better.(see table 2)
The results of the controlled studies are within the ranges reported by Arthurs et al. We stated in the discussion paragraph: …..patients' outcome in the assessed comparative studies is in accord with the long-term outcome in sizeable contemporary radiosurgery series as summarised in Appendix 4. Radiosurgery for vestibular schwannoma is a day care with 2% (median) of patients requiring additional treatment; less than 1% (median) experienced some facial neuropathy and trigeminal neuropathy occurred in 5% (median). It has no direct mortality and the risk of incapacitating complications is negligible or not existing. The comprehensive review of Arthurs et al. showed that after microsurgery less than 2% requires additional treatment. Varying with tumour size the rates of facial nerve palsy are as high as 10-30%.4 These numbers are of the same range in the comparative studies on tumours limited to a size of 3cm in table 2. Not mentioned in any detail by Arthurs et al. are other surgical morbidities, which are not trivial at all, being between 14-47% in the comparative studies. Major adverse events like mortality and discharge to long-term care may occur after microsurgery in about 0.5% and 1.2%, respectively.
Apart from this phrase also appendix 4 is provided.

25. No. In the referred articles is reported that QoL is disappointing and doctors are asked to be modest/humble on the outcome after microsurgery. Modesty is the word used in their text.