Determining the sample size for future trials of hearing instruments for unilaterally deaf adults: an application of network meta-analysis

Adam Christopher Gaunt[1] *and Pádraig Thomas Kitterick*[2,3]

[1] Nottingham University Hospitals NHS Trust, Queen's Medical Centre, NG7 2UH,

Nottingham, UK

[2] National Institute for Health Research (NIHR) Nottingham Biomedical Research Centre, Ropewalk House, Nottingham NG1 5DU, United Kingdom

[3] Otology and Hearing group, Division of Clinical Neuroscience, School of Medicine, University of Nottingham, Nottingham NG7 2UH, United Kingdom

Address correspondence to:

Adam Gaunt

Specialty Registrar, Otolaryngology

Queens Medical Centre, Nottingham University Hospitals NHS Trust

Derby Road, Nottingham NG7 2UH

United Kingdom

Telephone:

Email:

Conflicts of Interest and Source of Funding:

This work was supported by infrastructure funding from the National Institute for Health Research (NIHR). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.

Introduction

There have been several prospective studies of the effectiveness of hearing instruments for adults with single sided deafness (SSD), a condition that has been associated with significant psychological and social burden [1,2,3]. A recent meta-analysis examined the evidence for various hearing instruments including devices that re-route signals from the impaired to the non-impaired ear via air conduction (ACD) or bone conduction (BCD), and cochlear implantation (CI) [4]. Comparable outcomes were available across studies on a limited set of measures: the Speech Spatial and Qualities of hearing scale (SSQ) [5], the Abbreviated Profile of Hearing Aid Benefit (APHAB) [6], and the Hearing In Noise Test (HINT) [7].

The systematic review identified that there was a paucity of data for comparisons between certain hearing instruments (Figure 1). For example, three studies directly compared ACD to the Unaided condition[8–10] and three studies directly compared BCD and ACD[8–10], whereas comparable outcomes for BCD vs the unaided condition were available from eight studies[8–15]. Few studies compared these interventions to CI. The meta-analysis (MA) of data extracted from those studies was therefore limited by the specific comparisons that had been reported in the published literature. Revised estimates of the relative effects of these different treatment alternatives could be obtained using network meta-analysis (NMA) to fully utilise all available evidence, both direct and indirect.

To understand NMA in lay-terms the following analogy is useful. We have 3 hypothetical treatments: A, B and C. In this scenario there is a lot of data comparing A vs B and A vs C but little to none comparing B vs C [16]. These data form a network from which inferences can be made on the basis of the indirect relationships formed by the data. In other words, NMA allows us to draw meaningful conclusions about the relationship between intervention B vs C

even though we do not have little if any direct evidence to rely upon. It is therefore a statistical meta-analytic technique that incorporates both direct and indirect evidence [17].

We believe that NMA has a role to play in reducing research waste by utilising both indirect and direct evidence to provide best estimates of treatment effects based on all the available evidence. The application of this meta-analytical approach allows for research effort to be targeted where there is the greatest amount of treatment uncertainty; i.e. where there are few direct comparisons and the incorporation of indirect evidence has a notable impact on the estimated size of the treatment effect. Conducting clinical trials is costly, and poorly targeted studies risk wasting scarce research budgets, moreover, errors in design can lead to an inability to draw meaningful clinical conclusions. Chalmers and Glasziou[18] have estimated that 85% of all research effort translates to no meaningful or reproducible output. This may be due to various reasons including the underreporting of studies with disappointing results, selective publication of results or inappropriate study design. A cross-sectional analysis has demonstrated that up to half of all National Institutes of Health funded trial results remain unpublished at 30 months after trial completion [19]. Furthermore, Glasziou states that 'studies of published trial reports showed that the poor description of interventions meant that 40-89% were non-replicable' [20]. This poor conversion from research activity to real clinical benefit to patients is of great concern to all those involved in and relying upon clinical research.

Kitterick et al's MA identified 30 papers of an original 778 that met the criteria to be included in their review [4]. These were identified using PICOS (participants, intervention(s), comparators, outcomes, and study designs) framework [21] to set parameters that were of interest. These can be summarised as (P) Patients with average PTA of ≤30dB loss in the better ear and ≥70dB loss in the better ear, (I) hearing instruments used in SSD, (C) hearing instruments, placebo and no intervention, (O) speech perception in quiet and in noise, sound

localization, hearing- and health-related quality of life, complications and adverse events, (S) Controlled trials and prospective observational studies. The studies from which data were extracted are outlined in Table 1 and show the interventions assessed and the outcome measures used. One of the main observations arising from the systematic review and MA was the lack of data on comparisons between certain interventions (i.e. BCD vs ACD) and a lack of controlled trials that had been designed prospectively to have sufficient statistical power to detect treatment effects. The authors suggested that the effect sizes from the MA could be used to inform the sample sizes of future studies [4].

NMA is an attractive prospect in this context as it allows one to use all the available evidence to obtained revised estimates of treatment effects. With better estimates of effect size come better knowledge of where the greatest uncertainty lies, and better estimates of the sample sizes required to detect such effects in the context of future prospective clinical trials. The current study subjected data from the previous meta-analysis to NMA to examine whether the incorporation of indirect evidence changed the size and direction of treatment effects. The resulting changes were also assessed to identify the outcomes and comparisons with the greatest level of uncertainty, and to determine whether the required sample sizes based on the revised effect sizes would be feasible to recruit in future clinical trials.

Materials and Methods

The original meta-analysis synthesised data obtained using a variety of outcome measures that followed a prescribed methodology (and thus were likely to have been administered consistently) and were used across multiple studies [4]. The Speech Spatial and Qualities of Hearing Scale (SSQ) measures hearing difficulties across several domains including speech perception, spatial awareness of sound and sound qualities. It is designed to measure hearing disability across a range of scenarios including those that are affected by binaural function [5]. The Abbreviated Profile of Hearing Aid Benefit (APHAB) measures listening ability across four six-item subscales: aversion to sounds, background noise, ease of communication and reverberation [6]. The Hearing In Noise Test (HINT) assesses the ability of participants to understand sentences with a degree of background noise either presented directly ahead (co-incident with the speech, S_0N_0) or presented towards the impaired ear (S_0N_{ie}) or the non-impaired ear (S_0N_{ne}) [7]. The test can also be conducted in the absence of background noise (SIQ).

A network meta-analysis of data obtained using these outcome measures was conducted in four steps. First, the raw data obtained using the measures described above and the number of patients for whom data were available were extracted from each study listed in Table 1 and organised into a spreadsheet using Microsoft Excel. Second, these data were loaded into the R statistical programming environment and effect sizes were calculated for each individual study. As all studies used before-after designs, effect sizes were computed by dividing the observed pre-post treatment change on each outcome measure by the standard deviation of that change [22] using the 'metafor' package [23]. The resulting values expressed the size of each effect in units of standard deviations. Third, the effect sizes for each outcome measure were subjected to traditional random-effects meta-analyses separately for each treatment comparison (e.g. BCD vs unaided). The analyses determined the pooled treatment effect on

each outcome measure for each treatment comparison [23] and represent the meta-analysis approach used in the original systematic review [4]. We refer to these pooled effects as the 'direct evidence'.

Fourth, and finally, all effect sizes for each outcome measure were subjected to a network meta-analysis to determine pooled effects based on indirect evidence using the graph-theoretical method described by Rücker [24] as implemented in the 'netmeta' package for the R statistical programming environment [25]. A simple explanation for the general approach is that the analysis determines the *indirect* evidence for a particular treatment comparison of interest (i.e. A vs B) based on the difference between the *direct* evidence for other treatment comparisons that involve one of the treatments of interest (e.g. A vs C and B vs C). The general approach to determining the 'indirect evidence' can be expressed mathematically in the following form adapted from Cipriani et al.[26]: $AB_{indirect} = AC_{direct} - BC_{direct}$. We refer to the treatment effects produced by the network meta-analyses as the 'network evidence' as they combine both direct and indirect evidence.

For each treatment comparison, the direct and indirect evidence and the 'network evidence' (the result of synthesising both direct and indirect evidence) are reported in terms of the mean effects and their 95% confidence intervals [27]. Given the complexity of the network-based approach to determining treatment estimates based on direct and indirect evidence, metrics and tests have been proposed to aid interpretation of the resulting estimates of treatment effect. We report the proportion of direct evidence that contributes to the network evidence and a statistical test to compare the direct and indirect evidence to assess whether the assumption of consistency was violated [28]. The pooled effects resulting from the use of direct and indirect evidence were also compared by noting whether the direction of the effect had

changed and the whether the size of the effect had changed. Effect sizes were categorized as ‘small’ (<0.20), ‘medium’ (0.21 to 0.80), or ‘large’ (>0.8) [29].

A sample size calculation was conducted for treatment effects based on direct and network evidence using G*Power [30], a free to use cross-platform statistical tool that is available as a download for Windows and Macintosh operating systems from the Hienrich Heine University, Dusseldorf [31]. The sample size calculation determined the number of participants required to detect a given effect size with 80% power (probability of a false-negative of 0.2) and an alpha of 5% (probability of a false-positive of 0.05). The calculations were based on the assumption that future trials would know the expected direction of the effect (beneficial or harmful) and would be powered to detect changes in mean outcome scores between intervention and a control/comparator groups. Therefore, the sample size estimates were based on a one-tailed independent-samples t-test with an allocation ratio of 1:1 to the two groups.

Results

Tables 2 and 3 lists the estimates of effect size for comparisons between the unaided condition, ACD, BCD, and CI for the self-reported outcomes (APHAB, SSQ) and speech perception outcomes (HINT), respectively. Inconsistency between direct and indirect evidence was not identified for the self-reported outcomes but was identified for the S_0N_0 and S_0N_{NE} conditions of the HINT (*Z* and *p* values in Table 3). The size of the change resulting from the incorporation of indirect evidence varied from negligible (0.01 standard deviations, SD) to notable (0.38 SD). In one case the incorporation of indirect evidence altered the direction of the mean effect of ACD from being detrimental to listening ability to being beneficial (ACD vs Unaided; SSQ). However, in all cases the 95% confidence intervals of the effect sizes estimated from direct and network evidence overlapped, with the incorporation of indirect evidence widening confidence intervals around the treatment effects.

Table 3 reports sample size calculations performed using effect sizes based on direct and network evidence for comparisons between ACD, BCD, and the unaided condition. The incorporation of indirect evidence reduced the required sample size to detect changes in SSQ scores when comparing CI to unaided reduced from 36 to 26. However, the inclusion of indirect evidence increased the sample size required to detect changes in SSQ scores when comparing CI to ACD (46 to 48) and CI to BCD (42 to 204).

Discussion

Network meta-analysis is a useful adjunct to standard meta-analytical techniques in cases where there are multiple treatment options for a condition and few studies that directly compare certain pairs of interventions. It is a technique that is yet to be widely adopted in the otological sciences; for example, at the time of publication the only example in the field indexed on PubMed is a protocol for a NMA in sudden sensorineural hearing loss

[32]. The application of NMA in the context of hearing instruments for adults with SSD resulted in some notable changes in terms of both the direction and size of treatment effects. For example, when using SSQ to measure listening abilities with ACD compared to the unaided condition, the incorporation of indirect evidence revised the mean treatment effect on listening ability from being a small detrimental effect to a medium beneficial effect. Such cases highlight areas where there is considerable uncertainty over treatment effects.

Differences in treatment effects based on direct and network evidence could arise due to a variety of factors. There may be an imbalance in the quantity of direct and indirect evidence. For example, the effect size associated with the difference in SSQ scores for CI vs BCD decreased substantially from 0.79 to 0.35, a 56% reduction, once the indirect evidence was considered and the network evidence had the lowest proportion of direct evidence (67%) across all the comparisons examined in the current study. In that case, direct evidence was available from only one trial that reported a large positive treatment effect[8] whereas two studies reported comparisons of BCD with unaided condition using that outcome measure. Differences in study methodology or population could have could also have resulted in varying effect sizes across these studies. For example, the study that compared BCD to CI provided bone-conduction devices on a softband/tension clamp whereas all of the studies comparing BCD with the unaided condition used osseo-integrated implants. Osseo-integrated (percutaneous) implants are more effective at transducing high frequencies than transcutaneous devices such as softband-mounted devices[33]. These factors and other differences in study designs, such as how the treatments were delivered and the duration of follow up, could account for the significant inconsistency between direct and indirect evidence noted for some of the treatment comparisons.

The substantial reduction in the estimated size of treatment effect of CI vs BCD with and without indirect evidence increased the required sample size by a factor of 5 (42 to 204). A similar implication arose when comparing ACD to the unaided condition using the HINT sentence test in a frequently-used testing configuration for patients with SSD; i.e. speech from in front and noise towards non-impaired (good) ear (S_0N_{NE}). The sample size increased from 50 patients to 150 when indirect evidence was considered. For certain outcome measures, the numbers needed to power studies adequately became infeasible if one considered the indirect evidence (e.g. using APHAB to measure outcomes in ACD vs unaided) or conversely were reduced to potentially-feasible levels (e.g. using SSQ to measure outcomes in ACD vs unaided). These examples illustrate how NMA could prevent an underpowered trial being conducted or avoid unnecessary burden by over-recruitment, and in doing so prevent wastage of scarce research resources.

When discussing sample size calculations for future clinical trials, a distinction must be made between an observed difference reported in a published study, such as those incorporated into the current meta-analyses, and a clinically-important difference. In areas where researchers are unsure of a treatment effect it is expedient to determine the minimal clinically important difference (MCID) in the primary outcome of interest [34] . The MCID can be defined as the minimum change in outcome that is deemed clinically-significant. For example, an increase in the SSQ score of a few points may be statistically significant following an intervention but may or may not give a patient a clinically-important (perceptible) benefit. It is therefore relevant not only to consider what effect sizes may be the subject of uncertainty (as indicated by large changes in Tables 2 and 3) and whether it feasible to conduct a trial based on the required sample size (Table 4), but also whether the estimated sample size is likely to be meaningful to the clinician and patient alike.

Integral to the challenge of meta-analysis is the difficulty of comparing differing methodologies and outcome measures and synthesising this into a meaningful discourse about the benefits of interventions. Comparison between trials can be facilitated by the development of Core Outcome Sets (COS) that offer the prospect of a uniform way of measuring interventions in the context of clinical trials [35]. They can also inform the choice of primary outcome for future trial design by identifying outcomes that are important to patients. By adopting a COS it will be possible to directly compare interventions trials. There is currently a COS in development for adults with SSD [36, 37].

Limitations

The prospect of being able to use all available data to provide new evidence of treatment effectiveness and therefore inform clinical trial design and clinical decision making is undoubtedly attractive. Whilst NMA is able to adjust for bias when used in conjunction with conventional direct comparison techniques[38], as with any statistical procedure there are limitations to the NMA technique. To carry out NMA, as with traditional MA, assumptions have to made to allow the grouping and comparison of studies that include but are not limited to: (1) the study populations are likely to respond in comparable ways to the treatments under consideration; (2) the interventions are delivered in a similar way; and (3) the study designs are broadly similar – this is the concept of 'transivity'[26,39].An example of this is that when we compare populations that have received a CI to those that have received ACD or BCD. There are likely to be subtle variations between these groups, including differences in the characteristics of those eligible for implantation and fit for a surgical procedure versus those unable, ineligible, or unwilling to receive a cochlear implant. In addition, data were only available from a small number of studies with small sample sizes, restricting the evidence upon which any

inferences can be made about treatment effects, whether based on direct or network evidence.

Indirect evidence such as that provided by NMA is not afforded the same status as direct evidence found in head to head comparisons, in part due to the fact that the application of this technique is still an emerging field [26]. Donegan et al's review of reporting and methodological quality in indirect analyses drew attention to the fact that the 'underlying assumptions are not routinely explored or reported when undertaking indirect comparisons.'[40]. Chou also cites the limitation of indirect comparisons when comparing 'complex and rapidly evolving interventions'[41]. However, there is a move towards placing additional weight on indirect analyses as since 2015 10% of Cochrane reviews have utilised NMA[42] , with some calling for a re-evaluation of the evidential status accorded to NMA[42]. This technique is therefore illustrated here as an adjunct to conventional head to head comparisons that may be useful in situations where some treatment comparisons are under-represented in the published literature.

Conclusion

The application of network meta-analysis to extend existing analyses published alongside systematic reviews or to supplement the conduct of future reviews can aid the design of future trials of interventions for hearing-related interventions. The current results suggest that there is considerable uncertainty surrounding some published estimates of treatment effects associated with hearing instruments for adults with SSD. These results, together with further research to establish MCIDs and ongoing work to define a COS for SSD, will help ensure that future trials are targeted to reduce known uncertainty around treatment alternatives and make effective use of limited research resources.

REFERENCES

1. Lucas L, Katiri R, Kitterick PT. The psychological and social consequences of single-sided deafness in adulthood. *Int J Audiol*. 2018;57(1):21-30. doi:10.1080/14992027.2017.1398420

2. Noble W, Gatehouse S. Interaural asymmetry of hearing loss, Speech, Spatial and Qualities of Hearing Scale (SSQ) disabilities, and handicap. *Int J Audiol*. 2004;43(2):100-114.

3. Giolas TG, Wark DJ. Communication problems associated with unilateral hearing loss. *J Speech Hear Disord*. 1967;32(4):336-343.

4. Kitterick PT, Smith SN, Lucas L. Hearing Instruments for Unilateral Severe-to-Profound Sensorineural Hearing Loss in Adults: A Systematic Review and Meta-Analysis. *Ear Hear*. 2016;37(5):495-507. doi:10.1097/AUD.0000000000000313

5. Gatehouse S, Noble W. The Speech, Spatial and Qualities of Hearing Scale (SSQ). *Int J Audiol*. 2004;43(2):85-99.

6. Cox RM, Alexander GC. The abbreviated profile of hearing aid benefit. *Ear Hear*. 1995;16(2):176-186.

7. Nilsson M, Soli SD, Sullivan JA. Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *J Acoust Soc Am*. 1994;95(2):1085-1099.

8. Arndt S, Aschendorff A, Laszig R, et al. Comparison of pseudobinaural hearing to real binaural hearing rehabilitation after cochlear implantation in patients with unilateral deafness and tinnitus. *Otol Neurotol Off Publ Am Otol Soc Am Neurotol Soc Eur Acad Otol Neurotol*. 2011;32(1):39-47. doi:10.1097/MAO.0b013e3181fcf271

9. Niparko JK, Cox KM, Lustig LR. Comparison of the bone anchored hearing aid implantable hearing device with contralateral routing of offside signal amplification in the rehabilitation of unilateral deafness. *Otol Neurotol Off Publ Am Otol Soc Am Neurotol Soc Eur Acad Otol Neurotol*. 2003;24(1):73-78.

10. Wazen JJ, Spitzer JB, Ghossaini SN, et al. Transcranial contralateral cochlear stimulation in unilateral deafness. *Otolaryngol--Head Neck Surg Off J Am Acad Otolaryngol-Head Neck Surg*. 2003;129(3):248-254. doi:10.1016/S0194-5998(03)00527-8

11. Desmet JBJ, Wouters K, De Bodt M, Van de Heyning P. Comparison of 2 implantable bone conduction devices in patients with single-sided deafness using a daily alternating method. *Otol Neurotol Off Publ Am Otol Soc Am Neurotol Soc Eur Acad Otol Neurotol*. 2012;33(6):1018-1026. doi:10.1097/MAO.0b013e31825e79ba

12. Dumper J, Hodgetts B, Liu R, Brandner N. Indications for bone-anchored hearing AIDS: a functional outcomes study. *J Otolaryngol - Head Neck Surg J Oto-Rhino-Laryngol Chir Cervico-Faciale*. 2009;38(1):96-105.

13. Pai I, Kelleher C, Nunn T, et al. Outcome of bone-anchored hearing aids for single-sided deafness: a prospective study. *Acta Otolaryngol (Stockh)*. 2012;132(7):751-755. doi:10.3109/00016489.2012.655862

14. Saliba I, Nader M-E, El Fata F, Leroux T. Bone anchored hearing aid in single sided deafness: outcome in right-handed patients. *Auris Nasus Larynx*. 2011;38(5):570-576. doi:10.1016/j.anl.2011.01.008

15. Yuen H-W, Bodmer D, Smilsky K, Nedzelski JM, Chen JM. Management of single-sided deafness with the bone-anchored hearing aid. *Otolaryngol--Head Neck Surg Off J Am Acad Otolaryngol-Head Neck Surg*. 2009;141(1):16-23. doi:10.1016/j.otohns.2009.02.029

16. Caldwell DM, Ades AE, Higgins JPT. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *BMJ*. 2005;331(7521):897-900. doi:10.1136/bmj.331.7521.897

17. Mills EJ, Thorlund K, Ioannidis JPA. Demystifying trial networks and network meta-analysis. *BMJ*. 2013;346:f2914.

18. Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. *Lancet Lond Engl*. 2009;374(9683):86-89. doi:10.1016/S0140-6736(09)60329-9

19. Ross JS, Tse T, Zarin DA, Xu H, Zhou L, Krumholz HM. Publication of NIH funded trials registered in ClinicalTrials.gov: cross sectional analysis. *BMJ*. 2012;344:d7292. doi:10.1136/bmj.d7292

20. Glasziou P, Altman DG, Bossuyt P, et al. Reducing waste from incomplete or unusable reports of biomedical research. *Lancet Lond Engl*. 2014;383(9913):267-276. doi:10.1016/S0140-6736(13)62228-X

21. pmhdev. PICO Framework - National Library of Medicine. PubMed Health. https://www.ncbi.nlm.nih.gov/pubmedhealth/PMHT0029906/. Accessed September 9, 2018.

22. Morris SB, DeShon RP. Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychol Methods*. 2002;7(1):105-125.

23. Conducting Meta-Analyses in R with the metafor Package | Viechtbauer | Journal of Statistical Software. doi:10.18637/jss.v036.i03

24. Rücker G. Network meta-analysis, electrical networks and graph theory. *Res Synth Methods*. 2012;3(4):312-324. doi:10.1002/jrsm.1058

25. Rücker G, Schwarzer G, Krahn U, König J. *Netmeta: Network Meta-Analysis Using Frequentist Methods*.; 2018. https://CRAN.R-project.org/package=netmeta. Accessed December 7, 2018.

26. Cipriani A, Higgins JPT, Geddes JR, Salanti G. Conceptual and technical challenges in network meta-analysis. *Ann Intern Med*. 2013;159(2):130-137. doi:10.7326/0003-4819-159-2-201307160-00008

27. A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis | The BMJ. https://www.bmj.com/content/349/bmj.g5630.full.print. Accessed December 7, 2018.

28. König J, Krahn U, Binder H. Visualizing the flow of evidence in network meta-analysis and characterizing mixed treatment comparisons. *Stat Med*. 2013;32(30):5414-5429. doi:10.1002/sim.6001

29. Cohen J. *Statistical Power Analysis for the Behavioral Sciences.* Burlington: Elsevier Science; 2013. http://public.eblib.com/choice/publicfullrecord.aspx?p=1882849. Accessed February 3, 2018.

30. Faul F, Erdfelder E, Lang A-G, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods*. 2007;39(2):175-191.

31. Universität Düsseldorf: G*Power. http://www.gpower.hhu.de/. Accessed September 9, 2018.

32. Ahmadzai N, Kilty S, Wolfe D, et al. A protocol for a network meta-analysis of interventions to treat patients with sudden sensorineural hearing loss. *Syst Rev*. 2018;7(1):74. doi:10.1186/s13643-018-0736-3

33. Reinfeldt S, Håkansson B, Taghavi H, Eeg-Olofsson M. New developments in bone-conduction hearing implants: a review. *Med Devices Auckl NZ*. 2015;8:79-93. doi:10.2147/MDER.S39691

34. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. *Control Clin Trials*. 1989;10(4):407-415. doi:10.1016/0197-2456(89)90005-6

35. Williamson PR, Altman DG, Bagley H, et al. The COMET Handbook: version 1.0. *Trials*. 2017;18(3):280. doi:10.1186/s13063-017-1978-4

36. Hall DA, Kitterick PT. Towards a Consensus on Outcome Measures for Interventions that Seek to Restore Bilateral and Binaural Hearing in Adults with Unilateral Severe-to-Profound Hearing Loss: The CROSSSD (Core Rehabilitation Outcome Set for Single Sided Deafness). http://www.comet-initiative.org/studies/details/1084. Accessed March 6, 2018.

37. Hall DA, Kitterick PT, Katiri R. Systematic review of outcome domains and instruments used in designs of clinical trials for interventions that seek to restore bilateral and binaural hearing in adults with unilateral sensorineural severe to profound hearing loss ('Single Sided Deafness'). http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42018084274.

38. Song F, Harvey I, Lilford R. Adjusted indirect comparison may be less biased than direct comparison for evaluating new pharmaceutical interventions. *J Clin Epidemiol*. 2008;61(5):455-463. doi:10.1016/j.jclinepi.2007.06.006

39. Faltinsen EG, Storebø OJ, Jakobsen JC, Boesen K, Lange T, Gluud C. Network meta-analysis: the highest level of medical evidence? *BMJ Evid-Based Med*. 2018;23(2):56-59. doi:10.1136/bmjebm-2017-110887

40. Donegan S, Williamson P, Gamble C, Tudur-Smith C. Indirect Comparisons: A Review of Reporting and Methodological Quality. *PLoS One San Franc*. 2010;5(11):e11054. doi:http://dx.doi.org/10.1371/journal.pone.0011054

41. Chou R, Fu R, Huffman LH, Korthuis PT. Initial highly-active antiretroviral therapy with a protease inhibitor versus a non-nucleoside reverse transcriptase inhibitor: discrepancies between direct and indirect meta-analyses. *Lancet Lond Engl*. 2006;368(9546):1503-1515. doi:10.1016/S0140-6736(06)69638-4

42. Kanters S, Ford N, Druyts E, Thorlund K, Mills EJ, Bansback N. Use of network meta-analysis in clinical guidelines. *Bull World Health Organ*. 2016;94(10):782-784. doi:10.2471/BLT.16.174326

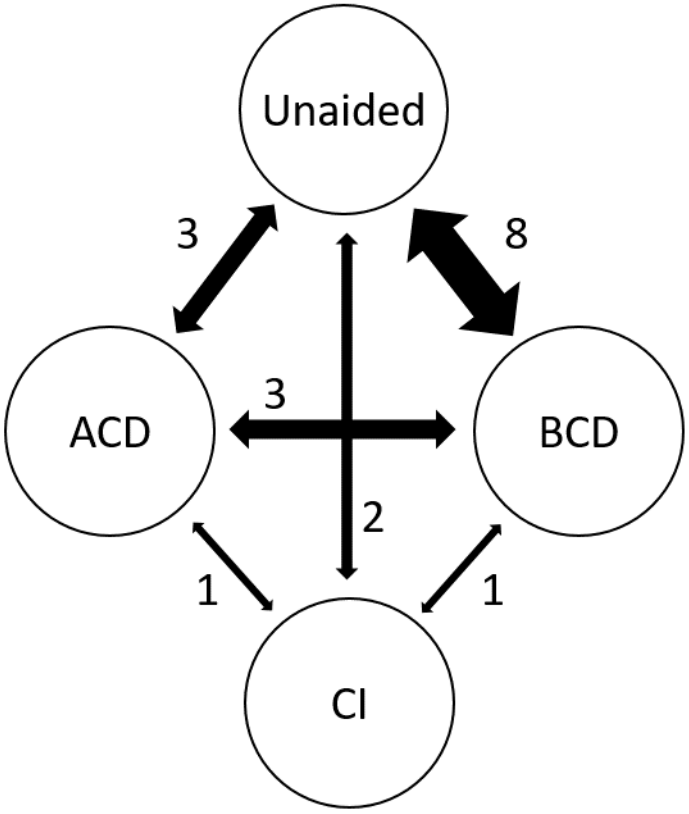

Figure 1: A schematic representation of the network of comparisons reported in the published literature of hearing instruments for SSD. The size of the arrows are approximately proportional to the frequency with which the various comparisons have been reported, with the actual numbers alongside the arrows (ACD: Air-conduction device; ; BCD: Bone-conduction device; CI: Cochlear Implant).

Table 1: Characteristics of the studies and data used to conduct the network meta-analysis.

Study	Total study sample size	Comparisons	Follow-up duration	Outcome measures extracted	Extracted data sample size
Arndt et al. 2011a,b	11	ACD vs Unaided	3 weeks	SSQ	10
		BCD vs Unaided	3 weeks	SSQ	11
		ACD vs BCD	3 weeks vs 3 weeks	SSQ	11
		CI vs Unaided	12 months	SSQ	11
		ACD vs CI	12 mo. vs 3 weeks	SSQ	11
		BCD vs CI	12 mo. vs 3 weeks	SSQ	11
Desmet et al. 2012	10	BCD vs Unaided	18 days	APHAB	10
Dumper et al. 2009	15	BCD vs Unaided	Not reported	HINT (SIQ, S_0N_0, S_0N_{IE}, S_0N_{NE})	15
Niparko et al 2003	10	ACD vs Unaided	1 month	APHAB, HINT (SIQ, S_0N_0, S_0N_{IE}, S_0N_{NE})	10
		BCD vs Unaided	4 months	APHAB, HINT (SIQ, S_0N_0, S_0N_{IE}, S_0N_{NE})	10
		ACD vs BCD	4 mo. vs 1 mo.	APHAB, HINT (SIQ, S_0N_0, S_0N_{IE}, S_0N_{NE})	10
Pai et al. 2012	25	BCD vs Unaided	At least 6 mo.	SSQ	25
Saliba et al. 2011	21	BCD vs Unaided	6 months	HINT (S_0N_0, S_0N_{IE}, S_0N_{NE})	21
Vermeire et al. 2009	20	CI vs Unaided	12 months	SSQ	9*
Wazen et al. 2003	18	ACD vs Unaided	4 months	APHAB, HINT (SIQ, S_0N_0, S_0N_{IE}, S_0N_{NE})	13 (APHAB) 16 (HINT SIQ) 11 (HINT SiN)
		BCD vs Unaided	4 months	APHAB, HINT (SIQ, S_0N_0, S_0N_{IE}, S_0N_{NE})	13 (APHAB) 12 (HINT)
		ACD vs BCD	4 mo. vs 4 mo.	APHAB, HINT (SIQ, S_0N_0, S_0N_{IE}, S_0N_{NE})	13 (APHAB) 12 (HINT)
Yuen et al. 2009	13	BCD vs Unaided	3 months	APHAB	13

APHAB: Abbreviated Profile of Hearing Aid Benefit; SSQ: Speech Spatial and Qualities of hearing scale; HINT: Hearing In Noise Test; *10 of the 20 participants had single-sided deafness, of which pre- and post-CI data were available for 9.

Table 2: Effect sizes based on direct, indirect, and network evidence for the self-reported outcome measures. Positive effect sizes indicate a more favourable outcome with the first intervention in each comparison. *Z* and *p* values relate to a test comparing direct with indirect evidence.

Comparison	*Direct evidence*	*Indirect evidence*	*Network evidence*	*w*	*Z*	*p*
APHAB						
ACD vs BCD	-0.63 [-1.08; -0.18]	0.00 [-0.84; 0.84]	-0.49 [-0.89; 0.10]	0.78	1.30	0.193
ACD vs Unaided	0.14 [-0.27; 0.55]	-0.58 [-1.61; 0.45]	0.04 [-0.34; 0.42]	0.86	-1.27	0.205
BCD vs Unaided	0.51 [0.26; 0.77]	—	0.53 [0.28; 0.79]	>0.99	—	—
SSQ						
ACD vs BCD	0.00 [-1.66; 1.67]	-2.92 [-7.21; 1.37]	-0.38 [-1.93; 1.17]	0.87	1.25	0.213
ACD vs CI	-0.75 [-2.44; 0.95]	-0.61 [-4.88; 3.66]	-0.73 [-2.30; 0.85]	0.86	-0.06	0.953
ACD vs Unaided	-0.07 [-1.74; 1.61]	1.85 [-1.52; 5.22]	0.31 [-1.19; 1.81]	0.80	-1.00	0.318
BCD vs CI	-0.79 [-2.49; 0.91]	0.54 [-1.87; 2.94]	-0.35 [-1.73; 1.04]	0.67	-0.88	0.379
BCD vs Unaided	0.71 [-0.47; 1.89]	0.46 [-4.12; 5.04]	0.69 [-0.44; 1.83]	0.94	0.10	0.918
CI vs Unaided	0.86 [-0.36; 2.08]	3.62 [-0.96; 8.20]	1.04 [-0.13; 2.22]	0.93	-1.14	0.253

APHAB: Abbreviated Profile of Hearing Aid Benefit; SSQ: Speech Spatial and Qualities of hearing scale; *w*: Proportion of direct evidence. Effect size categories: ‘small’ (<0.20), ‘medium’ (0.21 to 0.80), ‘large’ (>0.8).

Table 3: Effect sizes based on direct, indirect, and network evidence for the speech perception outcome measures from the Hearing In Noise Test. Negative effect sizes indicate a more favourable outcome with the first intervention in each comparison. *Z* and *p* values relate to a test comparing direct with indirect evidence.

Comparison	*Direct evidence*	*Indirect evidence*	*Network evidence*	*w*	*Z*	*p*
SIQ						
ACD vs BCD	0.38 [-0.10; 0.87]	0.21 [-0.97; 1.39]	0.36 [-0.09; 0.81]	0.86	0.27	0.790
ACD vs Unaided	0.26 [-0.17; 0.69]	0.02 [-1.52; 1.56]	0.24 [-0.17; 0.65]	0.93	0.29	0.773
BCD vs Unaided	-0.11 [-0.47; 0.24]	-4.07 [-14.98; 6.85]	-0.12 [-0.47; 0.24]	>0.99	0.71	0.478
S_0N_0						
ACD vs BCD	1.04 [0.40; 1.68]	-0.70 [-1.96; 0.55]	0.68 [0.11; 1.25]	0.79	2.44	0.015
ACD vs Unaided	0.11 [-0.45; 0.67]	1.90 [0.20; 3.60]	0.29 [-0.25; 0.82]	0.90	-1.96	0.049
BCD vs Unaided	-0.38 [-0.76; 0.00]	-3.36 [-8.74; 2.02]	-0.39 [-0.77; -0.01]	>0.99	1.08	0.278
S_0N_{IE}						
ACD vs BCD	0.36 [-0.07; 0.79]	0.16 [-1.13; 1.45]	0.34 [-0.07; 0.75]	0.90	0.29	0.771
ACD vs Unaided	0.78 [0.29; 1.27]	0.99 [0.07; 1.91]	0.83 [0.40; 1.26]	0.78	-0.39	0.695
BCD vs Unaided	0.50 [0.23; 0.78]	—	0.49 [0.21; 0.76]	>0.99	—	—
S_0N_{NE}						
ACD vs BCD	0.33 [-0.25; 0.91]	-1.39 [-2.89; 0.12]	0.11 [-0.43; 0.65]	0.87	2.09	0.037
ACD vs Unaided	-0.72 [-1.34; -0.09]	0.89 [-0.39; 2.16]	-0.41 [-0.97; 0.15]	0.81	-2.21	0.027
BCD vs Unaided	-0.49 [-0.88; -0.09]	-46.42 [-62.11; -30.73]	-0.52 [-0.91; -0.12]	>0.99	5.74	0.000

SIQ: Speech in quiet; S_0N_x: Speech presented from the front and noise presented from the front (x=0), the side of the impaired ear (x=IE) or the normal-hearing ear (x=N). *w*: Proportion of direct evidence. Effect size categories: ‘small’ (<0.20), ‘medium’ (0.21 to 0.80), ‘large’ (>0.8).

Table 4: Sample size calculations based on direct and network evidence.

	ACD vs Unaided			BCD vs Unaided			ACD vs BCD			CI vs Unaided			ACD vs CI			BCD vs CI		
Measure	D	N	Δ	D	N	Δ	D	N	Δ	D	N	Δ	D	N	Δ	D	N	Δ
SSQ	5050	360	*-4690*	52	54	*+2*	—*	174	—*	36	26	*-10*	46	48	*+2*	42	204	*+162*
APHAB	1264	15458	*+14194*	98	90	*-8*	64	106	*+42*	—	—	—	—	—	—	—	—	—
HINT																		
SIQ	368	432	*+64*	2046	1720	*-326*	174	194	*+20*	—	—	—	—	—	—	—	—	—
S_0N_0	2046	296	*-1750*	174	164	*-10*	26	56	*+30*	—	—	—	—	—	—	—	—	—
S_0N_{IE}	44	38	*-6*	102	106	*+4*	194	216	*+22*	—	—	—	—	—	—	—	—	—
S_0N_{NE}	50	150	*+100*	106	94	*-12*	230	2046	*+1816*	—	—	—	—	—	—	—	—	—

*Could not be calculated as direct effect size ≈ 0.