PEER REVIEW HISTORY

BMJ Paediatrics Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	A protocol for whole exome sequencing in newborns with congenital deafness – a prospective population based cohort.
AUTHORS	Downie, Lilian; Halliday, Jane; Burt, Rachel; Lunke, Sebastian; Lynch, Ella; Martyn, Melissa; Poulakis, Zeffie; Gaff, Clara; Sung, Valerie; Wake, Melissa; Hunter, Matthew; Saunders, Kerryn; Rose, Elizabeth; Rehm, Heidi; Amor, David

VERSION 1 - REVIEW

REVIEWER	Azaiez, Hela University of Iowa USA Competing interests: No
REVIEW RETURNED	01-Jul-2017

GENERAL COMMENTS	In this manuscript, Downie et al., describe a protocol for a prospective study to ascertain a cohort of children with congenital hearing loss for genetic screening using WES over 2 year period. The overall goals are to develop an evidence base for implementing WES testing in clinical diagnostics and to determine the genetic etiology of congenital HL in Australia. The authors also intend to survey parents to evaluate their experience with genetic testing as well as their preferences regarding reporting of incidental findings. Hereditary hearing loss is the most frequent sensory disorder and is phenotypically and genetically heterogeneous making its diagnosis using traditional testing methods laborious and time-consuming. The work proposed here is of great value as it will help set up a framework and develop a pipeline for genetic diagnosis in Australia capitalizing on the advances of Next generation sequencing. The manuscript provide great details regarding the identification and recruitment procedures of patients, however it is lacking clarification and specifics regarding genetic data analysis and variant interpretation. My specific comments are listed below. 1- The authors chose WES as a platform for their genetic testing; however, they have not explained the rationale behind using WES versus whole genome sequencing or targeted gene panels. It would be great if they could elaborate on the rationale behind this as more and more labs are either using targeted gene panels because of the increased depth of coverage and sensitivity or WGS because of its comprehensive coverage. 2- The authors do not provide an estimation of the number of patients that will be recruited over the 2 year period. At one point they mention “the sample size being less than 100” but this is vague. Additionally, they indicate the population sample will have significant diversity of ethnicity, language, etc, without providing proportions or percentages. It would be great if the authors could estimate a more

	specific number of patients and their ethnicities based on their previous experience. 3- The small sample size (even if it were 100 patients) combined with an ethnic diversity of the study population will hinder any effort to draw any conclusion regarding the genetic etiology of hearing loss in Australia. I would suggest the authors remove this aim from their study as it is not realistic. 4- The authors estimate that they will provide a diagnosis for half of patients enrolled. It was not clear where they got this estimation. Published studies investigating the genetic etiology of hearing loss using next generation sequencing provided diagnostic rates around 40% for heterogeneous diverse populations. Moreover, those studies also included copy number variations (accounting for up to 20% of diagnoses) in their analysis which this study does not intend to because of the challenges on identifying CNVs with WES. I suggest the authors propose a more realistic and conservative diagnostic rate of 30-40%. 5- WES data analysis and interpretation it at the heart of this project, yet not enough details were provided on how the authors intend to do that. The authors need to provide detailed description of their bioinformatic pipeline and data analysis: depth of coverage, quality control metrics and thresholds for variant quality, variant filtering, and variant interpretation. How many variants in average are estimated to be identified in each person after filtering? Please organize all the above in separate paragraphs with distinct titles: -Whole exome sequencing, -Bioinformatic analysis and variant filtering, -Variant interpretation. 6- The authors need to provide the list of genes they intend to analyze in tier 1, 2, etc. They also need to provide the list of genes in Choice B and C. 7- Please change the terminology ”genomic sequencing” used throughout the manuscript to WES. Genomic sequencing refers to a wide range of technologies: WES, WGS and targeted gene panels. 8- It was not clear what the authors meant by “Truncating and nonsense variants are curated in accordance with the standard procedures”. Description of these procedures is needed.

VERSION 1 – AUTHOR RESPONSE

1-*The authors chose WES as a platform for their genetic testing; however, they have not explained the rationale behind using WES versus whole genome sequencing or targeted gene panels. It would be great if they could elaborate on the rationale behind this as more and more labs are either using targeted gene panels because of the increased depth of coverage and sensitivity or WGS because of its comprehensive coverage.*

Response: This is the technology that is available to us currently. We have altered the wording of the introduction to include the rationale behind using WES. Paragraph 4 of the introduction now has the following additional text; This project is part of the Alliance’s broader program of work to develop an evidence base to inform the cost-effective use of genomics in clinical practice and to build workforce capability in Victoria, Australia(7). Alliance processes/procedures/systems are technically agnostic, that is, they are designed to be modifiable as utility in routine clinical practice is demonstrated for

other'-omic' technologies. WES with targeted analysis the first of these technologies ready for clinical application.

2- *The authors do not provide an estimation of the number of patients that will be recruited over the 2 year period. At one point they mention "the sample size being less than 100" but this is vague. Additionally, they indicate the population sample will have significant diversity of ethnicity, language, etc, without providing proportions or percentages. It would be great if the authors could estimate a more specific number of patients and their ethnicities based on their previous experience.*

Response: We have now included an estimate of eligible participants and provided an explanation of uncertainty around sample size as we do not have previous data to estimate the uptake of the test. We have included descriptive data of the ethnicity and socio-demographic status of a 2014-2015 cohort of patients diagnosed with congenital hearing loss in the state of Victoria. The methods/design section of the paper now includes the following text; Based on VIHSP data from previous years, it is estimated that 140 patients will meet the criteria for entry into the study over this 2 year period(12). There are no similar studies or empirical data on which to predicted uptake sample size. And Over the two year period prior to this project commencing VicCHILD identified that of babies born with congenital hearing loss, 40% of their parents were born overseas with the majority of these being of Asian ethnicity and 20% of families identified a primary language other than English. Families came from a range of socioeconomic backgrounds. This provides an insight into the diversity of the cohort that will be recruited in this project.

3- *The small sample size (even if it were 100 patients) combined with an ethnic diversity of the study population will hinder any effort to draw any conclusion regarding the genetic etiology of hearing loss in Australia. I would suggest the authors remove this aim from their study as it is not realistic.*

Response: We agree that the sample size is insufficient to accurately determine the genetic aetiology of deafness in our population; rather, this is a descriptive element of the study and we have altered the wording accordingly, to 'describe' the genetic aetiologies in our cohort rather than define, throughout the manuscript.

4- *The authors estimate that they will provide a diagnosis for half of patients enrolled. It was not clear where they got this estimation. Published studies investigating the genetic etiology of hearing loss using next generation sequencing provided diagnostic rates around 40% for heterogeneous diverse populations. Moreover, those studies also included copy number variations (accounting for up to 20% of diagnoses) in their analysis which this study does not intend to because of the challenges on identifying CNVs with WES. I suggest the authors propose a more realistic and conservative diagnostic rate of 30-40%.*

Response: We agree that to achieve diagnosis rate of about 50% probably will require CNV detection as well as WES. We have now included CNV detection, by microarray, in our protocol and have left the anticipated detection rate at 50%.

5- *WES data analysis and interpretation it at the heart of this project, yet not enough details were provided on how the authors intend to do that. The authors need to provide detailed description of their bioinformatic pipeline and data analysis: depth of coverage, quality control metrics and thresholds for variant quality, variant filtering, and variant interpretation. How many variants in average are estimated to be identified in each person after filtering? Please organize all the above in separate paragraphs with distinct titles: -Whole exome sequencing, -Bioinformatic analysis and variant filtering, -Variant interpretation.*

Response: We have added the details of sequencing and analysis using the suggested headings, please see the revised section 'Lab Protocol', however, please note that the WES is being performed in a clinically accredited laboratory in accordance with national guidelines and the test itself is not the focus of the research. The paper is aimed at describing the protocol around the use of the test in a specific patient cohort.

6- *The authors need to provide the list of genes they intend to analyze in tier 1, 2, etc. They also need to provide the list of genes in Choice B and C.*

Response: These are now included in Appendix 1.

7- *Please change the terminology "genomic sequencing" used throughout the manuscript to WES. Genomic sequencing refers to a wide range of technologies: WES, WGS and targeted gene panels.*

Response: This has been altered in one instance, however, the authors have intentionally used the terminology 'WES' when referring to the study and this protocol, but used the term 'genomic sequencing' when discussing concepts or other studies as other research in this area has used a range of next generation technology.

8- *It was not clear what the authors meant by "Truncating and nonsense variants are curated in accordance with the standard procedures". Description of these procedures is needed.*

Response: Please see revised section 'Lab protocol'.