AWARD NUMBER: CDMRPL-17-0-H1602900

TITLE: Individual Exposure Health Risk Profile (IEHRP)

PRINCIPAL INVESTIGATOR: Dr. Dirk Yamamoto

CONTRACTING ORGANIZATION: 711th Human Performance Wing (WPAFB)
Wright-Patterson AFB, OH

REPORT DATE: May 2021

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Development Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
May 2021	Final	01Feb2018-31Jan2021

4. TITLE AND SUBTITLE

Individual Exposure Health Risk Profile (IEHRP)

5a. CONTRACT NUMBER

5b. GRANT NUMBER
CDMRPL-17-0-H1602900

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Dr. Dirk Yamamoto

email: dirk.yamamoto.2@us.af.mil

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESSES

711th Human Performance Wing (WPAFB)
Wright-Patterson AFB, OH 45433

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Development Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The US Air Force is developing a comprehensive understanding of an individual's exposures over a lifetime, and how those exposures interact with genetic factors to affect health risks. A pilot model focused on sensorineural hearing loss was developed using a mathematical framework incorporating genetic, demographic and exposure data into a prediction of relative risk for individuals in a population. The initial structure is based on a published cumulative genetic risk score algorithm and for key risk factors used public data from the UK Biobank and National Health and Nutrition Examination Survey for 41 single nucleotide polymorphisms (SNPs), gender, age, race/ethnicity, highest level of education, and self-reported exposure to occupational and firearm noise. The model estimates approximately 10.5% of the US population with elevated (8.1%) or high (2.4%) risk (measured prevalence: 20-30%). Age is the primary determinant of risk (OR for 60-69 yr old: 39.5), followed by education (OR: 2.4 - 4.2), self-reported noise exposure OR (1.0 – 2.16) and gender (OR:1.8). Genetic predisposition had negligible effect on overall risk. To facilitate model access, we developed a user-friendly application for identification of at-risk individuals and prioritization of risk factors for accelerated implementation of reasonable hazard protection plans as needed.

15. SUBJECT TERMS

exposomics, exposome, total exposure, Total Exposure Health, TEH, health risk, exposure risk, exposure science

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRDC
Unclassified	Unclassified	Unclassified	Unclassified	66	19b. TELEPHONE NUMBER *(include area code)*

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

PREFACE

Funding for this research was provided by Defense Health Program (DHP) Military Operational Medicine Research Program (MOMRP) Joint Program Committee-1 (JPC-1). Support was also provided through the Total Exposure Health Team in the Air Force Research Laboratory, 711th Human Performance Wing (711 HPW), Airman Systems Directorate, Airman Biosciences Division, Product Development Branch, Force Health Protection Section at Wright-Patterson Air Force Base, OH. This research was conducted under the cooperative agreement FA8650-15-2-6608 with the Henry M. Jackson Foundation for the Advancement of Military Medicine (HJF) and under UES contract FA8650-17-2-6834, which was the vehicle for the subcontract to 21st Century Tox Consulting, LLC. David R. Mattie, PhD was the technical manager for this project.

ABSTRACT

The US Air Force is developing a comprehensive understanding of an individual's exposures over a lifetime, and how those exposures interact with genetic factors to affect health risks. A pilot model focused on sensorineural hearing loss was developed using a mathematical framework incorporating genetic, demographic and exposure data into a prediction of relative risk for individuals in a population. The initial structure is based on a published cumulative genetic risk score algorithm and for key risk factors used public data from the UK Biobank and National Health and Nutrition Examination Survey for 41 single nucleotide polymorphisms (SNPs), gender, age, race/ethnicity, highest level of education, and self-reported exposure to occupational and firearm noise. The model estimates approximately 10.5% of the US population with elevated (8.1%) or high (2.4%) risk (measured prevalence: 20-30%). Age is the primary determinant of risk (OR for 60-69 yr old: 39.5), followed by education (OR: 2.4 - 4.2), self-reported noise exposure OR (1.0 – 2.16) and gender (OR:1.8). Genetic predisposition had negligible effect on overall risk. To facilitate model access, we developed a user-friendly application for identification of at-risk individuals and prioritization of risk factors for accelerated implementation of reasonable hazard protection plans as needed.

KEY WORDS
hearing loss, risk assessment, computational modeling

1.0 SUMMARY

The goal of the Individual Exposure Health Risk Profile (IEHRP) project is the development of mathematical models that will support prediction of individual risk to workplace and environmental exposures. This approach is intended to promote the health and well-being of Air Force personnel through effective early detection, intervention, and risk mitigation through the development of a predictive model for individual susceptibility to disease based on occupational, environmental, demographic, and genetic factors. Ultimately, the goal of IEHRP is to translate these findings into valuable, clinically actionable recommendations for individuals and their health care providers for illness/disease prevention and improving human performance, ensuring a ready, fit personnel force for combatant commanders. This first IEHRP model will be the beginning of an Operational Health Risk and Readiness Profile (OHR&RP) to reach this goal.

This initial IEHRP modeling effort focused on sensorineural hearing loss (SNHL) as a proof of concept for the broader OHR&RP program. The IEHRP model structure is based on a published statistical model, the Risk Estimation for Genetic and Environmental Traits (REGENT) model, which calculates an aggregate risk score by combining multiple risk factors from both environmental and genetic sources. Key contributing factors to individual risk were incorporated into the predictive health risk profile, including: occupational exposure to noise, susceptibility due to individual demographics (e.g., age, gender, race/ethnicity, level of education), behavioral characteristics (recreational firearm use), occupational noise exposure, and genetic susceptibility. Model predictions were derived from demographic data from the National Health and Nutrition Examination Survey (NHANES) and genetic data from published genome-wide association studies (GWAS) in human populations. The model builds a virtual population based on incidence rates and odds ratios derived from the input data; using confidence intervals from the data, the model then bins the population into average, reduced, elevated or high-risk categories. Using the described risk factors, the model predicts approximately 10% of the US population to be at elevated or high risk for hearing loss. This is within a factor of three from the measured hearing loss prevalence of 24% in the NHANES study. However, the under prediction of the prevalence does indicate that additional risk factors may need to be considered in further iterations of the model. In particular, chemical exposures and additional sources of noise should be further explored

To facilitate access to the IEHRP model, an application (app) was developed which provides a graphical interface. This IEHRP app may be used by combatant commanders or physicians to quickly and efficiently evaluate susceptibility of an individual or group of combatants to hearing loss, as well as fitness for duty. It is expected that this user interface can be easily adapted for use as technological platforms are updated and as the model is expanded.

2.0 INTRODUCTION

In response to Public Law 105-85, Presidential Review Directive 5, and the 2013 National Defense Appropriations Act (NDAA) Section 313 to collect, document, and act on long-term environmental health risks, the Air Force Medical Service (AFMS) introduced the Total Exposure Health (TEH) Initiative. TEH is intended to capture workplace, environmental, and lifestyle exposures to the individual using advances in science, technology, and informatics to help the Military Health System (MHS) reduce short- and long-term health risks to individuals by preventing disease, enhancing human performance, supporting force health protection, and ultimately developing precision health solutions for the warfighter in garrison and in theater.

To support the technical capabilities of TEH, it is necessary to identify and integrate various big data sources and emerging data sets for genetics, exposures, demographics, and various biomarkers into a common framework for clinical and operational action. We refer to this common framework as the IEHRP. Ultimately, this framework will be implemented as part of a future Clinical Decision Support System (CDSS) intended to enrich clinical and performance assessments of personnel by physicians and commanders, and to facilitate enhanced documentation. The aspiration for this future state would also include full integration with each individual military member's electronic health record (EHR) to create an individual longitudinal exposure record (ILER) per 2013 NDAA Section 313. As efforts to develop and increase accessibility to such records advance, incorporation of real-world exposure, medical, and demographic data into a computational framework will support high-resolution predictions of individual risk to various health outcomes and performance status throughout the member's lifetime.

In 2018 the development of a prototype model began to demonstrate the possible framework for incorporation of these different data streams and development of mathematical models that will support prediction of individual risk to workplace and environmental exposures. The IEHRP model concept was intended to promote health and well-being of Air Force personnel through effective early detection, intervention, and risk mitigation through development of a predictive model for individual susceptibility to disease based on occupational, environmental, demographic, and genetic factors. Ultimately, the goal is to translate these findings into valuable, clinically actionable recommendations for individuals, their supervisors, and their health care providers for illness/disease prevention and evaluating/improving human performance, and for ensuring a ready, fit force for combatant commanders. This first IEHRP model will be the beginning of an OHR&RP to reach this goal.

2.1 Objectives

This initial IEHRP modeling effort focused on SNHL as a proof of concept for the broader OHR&RP program. SNHL is a general term for hearing loss due to damage to the inner ear or the auditory nerve, whether due to age, noise or chemical toxicity, or conductive hearing loss (when sound cannot reach the inner ear, e.g. due to fluid in the middle ear or a hole in the eardrum). Key contributing factors to individual risk for SNHL were incorporated into the predictive health risk profile, including: occupational exposure to physical and chemical

stressors, environmental exposure to physical and chemical stressors, susceptibility due to individual demographics (e.g., age, weight, race), behavioral characteristics (e.g., smoking, alcohol consumption, risk enhancing hobbies), and genetic susceptibility (Figure 1). The goal of this project was to develop a prototype model to demonstrate the feasibility of combining exposure, demographic, and genetic information in a risk prediction model.

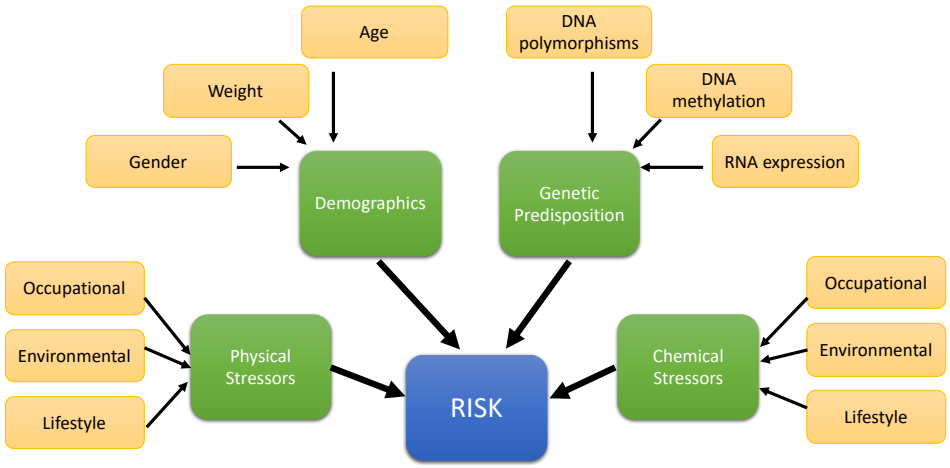

Figure 1. Key variables in the predictive model for individual susceptibility to SNHL. *Green boxes indicate key variables in the model predicting overall risk (blue box). Gold boxes indicate sources contributing to each variable.*

3.0 APPROACH

3.1 Model structure

The initial structure of the pilot model was based on the REGENT model package in R. The REGENT model code (in R script) is available through the Comprehensive R Archive Network (https://cran.r-project.org/), and implementation of the model is described in published papers (Crouch *et al*., 2013) and the associated manual (https://cran.r-project.org/web/packages/REGENT/REGENT.pdf). The REGENT model calculates relative risk for disease outcomes based on genetic susceptibility (single nucleotide polymorphism (SNP) data) and "environmental" factors (Crouch *et al*., 2013). Original developers of the model used the term "environmental" to refer to any non-genetic factors, including demographics (e.g., age, race), lifestyle risk factors (e.g., smoking), and chemical or physical exposures. Inputs are categorical and based on traditional analysis of epidemiological data. Thus, categories of risk factors (age, smoking status, SNP presence/absence) are associated with an odds ratio (OR) calculated from the epidemiological or GWAS data. The OR is used in epidemiological studies to quantify the relationship between a particular risk factor and a health outcome. OR is calculated using the number of case-patients who did or did not have a risk factor and the number of controls who did or did not have the same risk factor. The odds ratio tells us how much higher the odds of the outcome are among case-patients/subjects than among controls. The confidence interval (CI) provides a measure of confidence around the OR.

The REGENT model uses a statistical approach to identify individuals with low, average, elevated, or high risk for a disease outcome based on the relative risk (or risk ratio, RR) conferred by the existence of genetic or environmental factors (each expressed as an OR or RR). Using measured incidence rate and RR information for various risk factors reported in epidemiological studies, the model builds a virtual population and assigns the various risk factors randomly across the population. Assuming that the risk factors are independent, total relative risk (TRR) is then calculated by multiplying the OR for each risk factor. Finally, the model ranks the virtual individuals by TRR and assigns a classification of low, average, elevated or high risk based on the confidence intervals of the relative risk for the input factors (Crouch *et al.,* 2013). Figure 2 illustrates the model process.

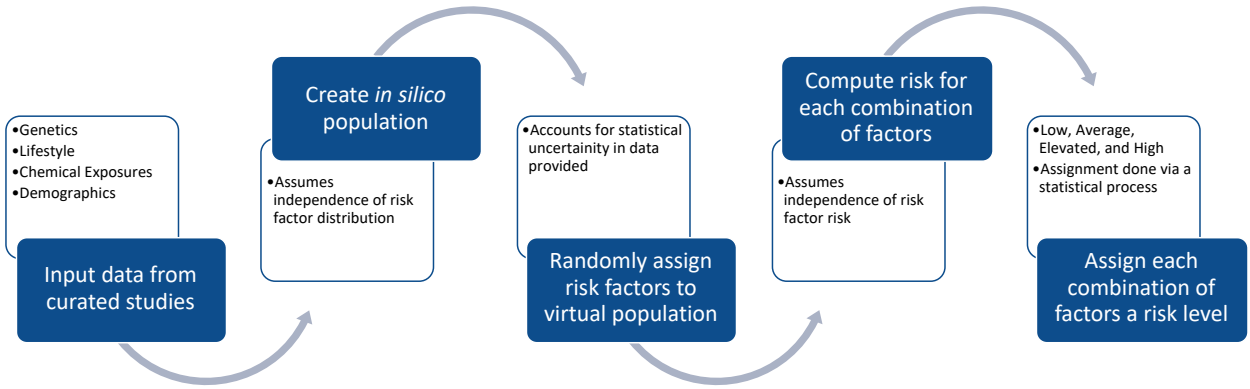

Figure 2. REGENT model process.
Blue boxes describe the model process. White boxes provide information on model parameters and assumptions.

The original REGENT code can be found at https://cran.r-project.org/. The modified code for the IEHRP hearing loss model is provided in Appendix A.

A number of edits were made to the original REGENT code in order to improve utility of the model for IEHRP, including:
- Improved efficiency in running the model
- Improved ease of use for novice users
- Allowing for input of larger data sets
- Allowing for evaluation of a greater number of risk factors

In addition to these technical aspects of the model, the graphical output of the model was modified to support better communication of the model predictions with the end user. The graphs of the model output were modified to improve the labeling of the axes and the colors used in the graphs. The new colors are more distinct, in order to better visualize the different risk groups. The color palette is also color blind-friendly in an effort to communicate the results with the broadest possible audience, as approximately 8.5% of the population is color blind. A comparison of the pre-programmed REGENT graphs vs the improved graph is shown in Figure 3 The final model code as edited for the IEHRP pilot model is provided in Appendix A.

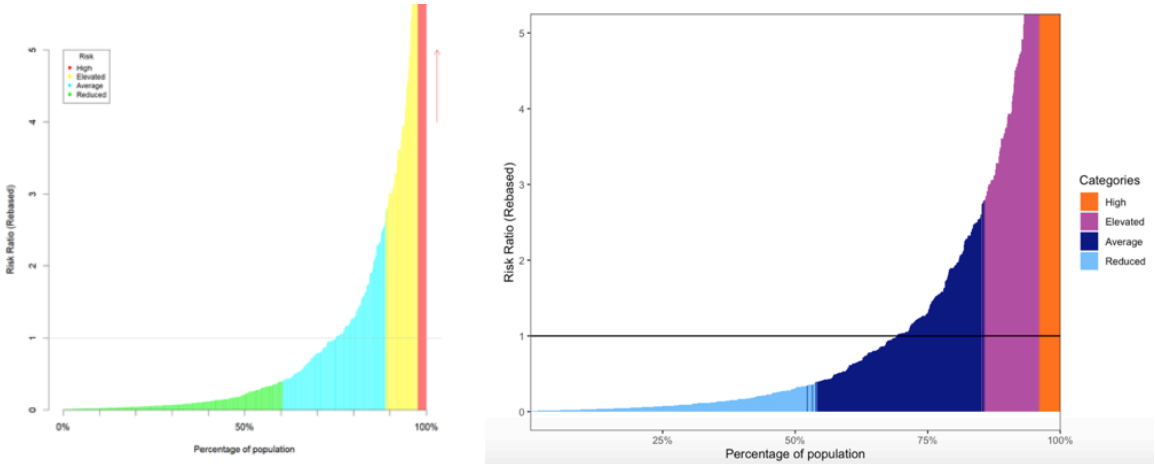

Figure 3. Pre-existing REGENT model output (left) and modified graphical output of the IEHRP hearing loss model (right).
The colors on the improved plot (right) are color blind-friendly. Predictions are based on the NHANES hearing loss data set (see Section 4).

3.2 Collection of published data to support model development

Initial model development and evaluation were performed using publicly available data sets. Scientific literature searches were performed using PubMed and Scopus, and internet searches

were performed using the Safari browser and Google web search engine to identify useful data sets.

Areas of interest for model development were defined *a priori* during discussions with project managers and USAF hearing loss experts:
- Background information on the biology and physiology of hearing
- Demographic factors related to risk for hearing loss
    - Age
    - Gender
    - Race/Ethnicity
    - Disease status (cardiovascular disease, diabetes, obesity, etc.)
    - Behavioral factors (smoking, alcohol consumption, etc.)
- Genetic susceptibility to hearing loss
- Mechanism of action studies (identify key pathways/proteins/genes)
- Human studies (occupational, epidemiological, defined exposure) with RNA, DNA data
- Chemical-induced ototoxicity
    - Identifying compounds with ototoxic effects and their mode of action
    - Dose-response information in animals and humans

More than 100 papers were identified to contain information that could support the hearing loss model. The citations for these publications are provided in Appendix B. Data that could be used to support model predictions were extracted from the literature.

Key attributes of a useful data set were defined *a priori*:
- Accessibility of data
- Concurrent measure of hearing loss (quantitative evaluation, e.g., pure tone air conduction/audiogram, etc.) and risk factors
- Large study populations for statistical power
- Diversity of sampled subjects (e.g., race, gender, socioeconomic status, etc.)

Demographic data

Of the identified epidemiological studies with publicly available data, diverse study populations, and quantitative measures of hearing loss, only three studies would be considered "large" (n > 1,000) – the NHANES, Epidemiology of Hearing Loss Study (EHLS), and Beaver Dam Offspring Study (BOSS).

NHANES is a program of cross-sectional studies conducted under the purview of the Centers for Disease Control and Prevention (CDC) to assess the health and nutritional status of adults and children in the United States. To ensure statistical power, NHANES employs a weighted sampling approach with oversampling of persons 60 and older, African Americans, and Hispanics. The survey combines interviews and physical examinations and is conducted as a series of surveys focusing on different population groups or health topics. A nationally representative sample of about 5,000 persons is examined each year across locations in counties across the country. The NHANES interview includes demographic, socioeconomic, dietary, and health-related questions. The examination component consists of medical, dental, and

physiological measurements, as well as laboratory tests administered by highly trained medical personnel. Complete audiogram threshold measurements are available for a total of 3,831 participants (1,953 men and 1,878 women) in the 1999 – 2004 and the 2011 – 2012 NHANES cycles.

The EHLS is a longitudinal study conducted by the University of Wisconsin. The study was initiated in 1993 in Beaver Dam, Wisconsin with 3,753 participants ranging in age from 48–92 years, who were then followed up every 5 years. Hearing loss was evaluated using pure tone air conduction and word recognition in quiet and in competing message (WRCM).

The BOSS is a longitudinal study conducted as a follow-up to the EHLS in 2005 – 2008. Hearing loss data were collected for 2,837 offspring of the EHLS. The hearing examination included otoscopy, tympanometry, pure-tone air- and bone-conduction audiometry, as well as WRCM.

The NHANES, EHLS, and BOSS studies had a similar number of subjects and similar level of racial, gender, and socioeconomic diversity. Similar risk factors were evaluated in each of the three studies as well (income, education, gender, age, cardiovascular disease, smoking status, occupational noise exposure, etc.). However, the NHANES study had a greater diversity in age and geographical location. For this reason, the IEHRP hearing loss model was based on the NHANES hearing loss data.

Three published analyses of the NHANES data were identified by the literature search: Agrawal *et al.* (2009), Carroll *et al.* (2017), and Hoffman *et al.* (2017). The Agrawal and Hoffman papers used similar definitions of hearing loss (> 25 dB change at 1-2 or 3-8 kilohertz; kHz) to define speech frequency and high frequency hearing loss, respectively. These studies were evaluating age-related hearing loss (i.e., not specifically noise-induced). Carroll *et al.* (2017) used very different criteria for hearing loss (high-frequency audiometric notch when any threshold at 3, 4, or 6 kHz exceeded the average threshold at 0.5 and 1 kHz by ≥15 decibels hearing level (db HL) and the 8 kHz threshold was at least 5 dB HL lower (better) than the maximum threshold at 3, 4, or 6 kHz), as they were attempting to evaluate specifically noise-induced hearing loss. Both approaches were evaluated with the model, using the data of Carroll *et al.* (2017), Agrawal *et al.* (2017) and Hoffman *et al.* (2017).

Genetic data

Literature searches were performed to identify publicly available data for genetic polymorphisms that confer susceptibility to SNHL. Initial searches included all nonsyndromic hearing loss (congenital, age-related, noise-induced). This returned studies using targeted sequencing using either Sanger sequencing or Next Generation Sequencing (NGS) in the human; GWAS in the mouse; and GWAS in the human using DNA databanks such as the UK Biobank (https://www.ukbiobank.ac.uk). More than 30 publications have been identified with information for human or mouse single nucleotide polymorphisms (SNPs) associated with nonsyndromic hearing loss, providing a total of 183 genes. Each gene can have more than one reported SNP associated with hearing loss, making the total number of SNPs well over 200. Information about specific SNPs will be collected in future efforts. These genes have been summarized in Appendix C. In the majority of studies, the aim was to identify putative SNPs or genes that could

confer susceptibility of hearing loss. For data collected in mice to be useful for human prediction, additional confirmation is needed in the human. For the current pilot model project, we focused on genetic data sets collected in humans, though additional SNPs could be later added to the model if confirmed to be relevant.

Few studies provide both a quantitative evaluation of the relative risk associated with specific DNA mutations (SNPs) and a comprehensive evaluation of the genome for hearing loss. The majority of human studies typically suffer from insufficient sample sizes to allow calculation of statistically sound ORs. Of the studies identified in the literature, only two publications allowed calculation of ORs for the putative SNPs: Zhang *et al.* (2019), and Wells *et al*. (2019).

In Zhang *et al*. (2019), 476 subjects with noise-induced hearing loss (NIHL; defined as > 1 year occupational noise exposure and high frequency hearing threshold > 45 dB) and 476 matched controls were recruited from a cross-sectional survey of NIHL in China. A total of 83 candidate SNPs were genotyped using nanofluidic dynamic arrays. Seven SNPs in the CDH23, PCDH15, EYA4, MYO1A, KCNMA1, and OTOG genes were significantly ($P < 0.05$) associated with the risk of NIHL, whereas seven other SNPs were marginally ($P > 0.05$ and $P < 0.1$) associated with the risk of NIHL. Zhang *et al.* (2019) provides ORs and CIs for each of these SNPs, which were then used as input in the IEHRP hearing loss model for initial evaluations. In general, ORs for individual SNPs were < 1.5, though two SNPs had ORs > 2.

Wells *et al.* (2019) performed a GWAS analysis of the UK Biobank (UKBB) DNA data set. The UKBB contains DNA SNP data and health and lifestyle questionnaires for > 250,000 subjects from the United Kingdom (https://www.ukbiobank.ac.uk). The UKBB does not have audiogram data. Hearing loss was defined by self-reported hearing loss or reported use of a hearing aid based on participants' questionnaires. However, the study size provides a tremendous advantage in terms of statistical power, particularly for the large number of comparisons made in GWAS studies. Both the Zhang *et al.* (2019) and Wells *et al.* (2019) studies were used with the IEHRP model.

Wells *et al.* (2019) did not report ORs or CIs for the SNPs associated with hearing loss in their GWAS study. To model the genetic risk for hearing loss, we obtained the summary statistics of Wells *et al.* (2019) study from UK Biobank. From the summary statistics, we selected the 41 SNPs that reached genome-wide significance ($p < 5x10^{-8}$) and represented independent signals, based on linkage disequilibrium (LD) clumping using the PLINK program (v1.07; Purcell *et al*., 2007). Effect size estimates were converted to ORs and CIs for use in the REGENT model R package.

3.3 Evaluating utility of public data sets to predict relative risk for Air Force personnel

To compare the distribution of the demographic characteristics in the NHANES population to those of current active duty Air Force personnel, we obtained access to the anonymized (de-identified) individual subject data in the in NHANES study. The anonymized demographic data for current Air Force personnel (enlisted and officers) were also retrieved (16 September, 2020, 15 January 2021) from the Air Force Personnel Center Personnel Statistics (PERS-STAT) Analytics Site.

3.4 Development of an application (app)

An app was developed to facilitate end user access to the model. The graphical user interface was programmed in R with the Shiny plugin. Currently the app is accessed through R Studio, but will ultimately be accessed via existing 711th Human Performance Wing (711 HPW) applications that can be downloaded onto mobile devices.

4.0 RESULTS

4.1 IEHRP model structure and parameters

To determine which risk factors should be included in the model, the supporting literature was evaluated to identify risk factors that are consistently associated with increased risk (ORs > 1) for hearing loss across diverse data sets. Key risk factors for hearing loss that were identified from the literature search are discussed below. The data chosen for use in model development are summarized in Tables 1 - 2.

4.1.1. Age as a risk factor for hearing loss

Age is the predominant risk factor for elevated risk to hearing loss in the general population. NHANES data collated from the 1999 – 2004 surveys demonstrated increased risk for bilateral high-frequency hearing loss (> 25 dB change at 3-8 kHz) from ages 20-69, with odds ratios ranging from 2.1 at ages 30-39 to 50 at ages 60-69 (Figure 4; Agrawal *et al*., 2008), when adjusted for sex, race, smoking cardiovascular disease, diabetes, socioeconomic factors, and occupational noise. At speech frequencies (0.5 – 2 kHz), even greater risk was associated with age and bilateral hearing loss (OR: 3.3 – 101). Similar results have been reported by independent evaluations of NHANES data (Hoffman *et al*., 2017) and in other epidemiological studies (Cruickshanks *et al*., 1998; Nash *et al*., 2011; Pearson *et al*., 1995).

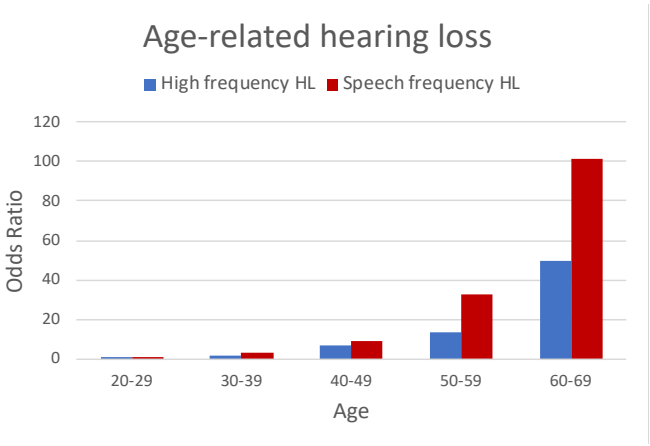

Figure 4. Age-related hearing loss from 1999-2004 NHANES studies.

4.1.2. Gender as a risk factor for hearing loss

According to various evaluations of the NHANES data, men are approximately 2-fold more likely to experience hearing loss (Agrawal *et al*., 2008; Carroll *et al*., 2017; Hoffman *et al*., 2017). Other studies have confirmed this gender-dependence (Cruickshanks *et al*., 1998; Pearson *et al*., 1995). More in-depth analyses have found that the gender differences are more pronounced at high frequencies, with older men showing drastic decline in ability to hear high

frequencies (Figure 5). Interestingly, women are slightly more susceptible to loss of hearing at low frequencies as they age compared to men.

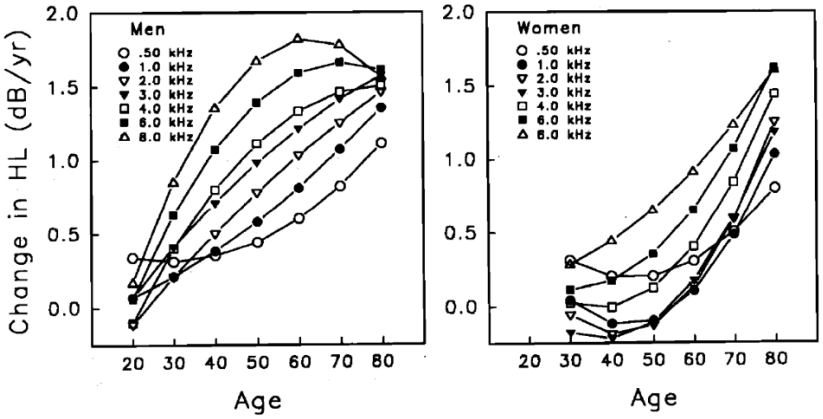

Figure 5. Average age-related changes in hearing according to gender.
Reproduced from Pearson et al. (1995).

4.1.3. Race/ethnicity as a risk factor for hearing loss

Data from the NHANES studies demonstrate differences in hearing loss associated with race/ethnicity (Agrawal *et al*., 2008; Hoffman *et al.,* 2017; Lin *et al*., 2012). In the NHANES studies, participants self-identify as "white, non-Hispanic", "black, non-Hispanic", "Asian, non-Hispanic", "Mexican American", "other Hispanic" or "other". Non-whites tend to show decreased risk for hearing loss, particularly "black, non-Hispanic" (Figure 6). It has been suggested that this difference is a result of the increased protection that melanin provides against reactive oxygen species (ROS), which act as biological mediators of hearing loss (Lin *et al*., 2012).

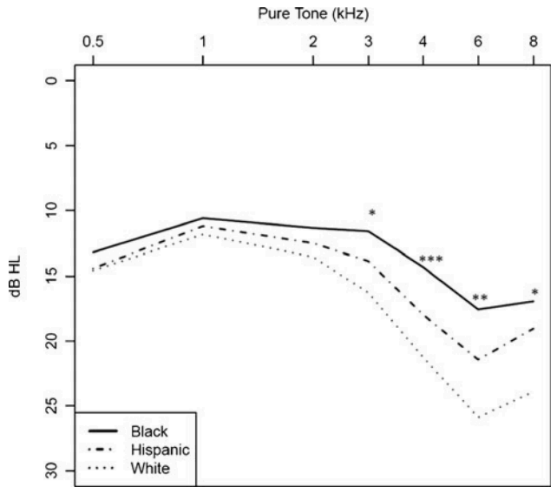

Figure 6. Mean audiograms by race/ethnicity.
The significance of the association of racial/ethnic category with each pure tone is denoted by asterisks ($p < 0.05$; ** $p < 0.01$; *** $p < 0.001$) based on regression models adjusting for all demographic (age, sex, education, income), medical (diabetes, smoking, hypertension, stroke), and noise (work, leisure, firearm) exposure covariates.*

4.1.4. Noise as a risk factor for hearing loss

NHANES included occupational noise exposure as a risk factor based on self-reported noise exposure. The NHANES survey questionnaire defined "loud" noise as when "you had to speak in a raised voice to be heard" and "very loud" noise as when "you have to shout in order to be understood by someone standing 3 feet away from you." When this same data set was evaluated for the contribution of noise to hearing loss, the odds ratio ranged from 1.2 – 2.4 when adjusted for age, sex, and other demographic factors (Hoffman *et al.*, 2017). Other studies have shown that the contribution of occupational noise to overall hearing loss is independent of age and gender, and calculated odds ratios are typically ≤ 2 (Carroll *et al.*, 2017; Krishnamurti, 2009).

Also included in the NHANES survey was noise from recreational firearm use. Participants were asked to estimate the number of rounds fired from a gun during the previous year. Firearm use was associated with elevated risk of hearing loss, with 1-999 yearly rounds fired yielding an OR = 1.4 and with 1,000 or more rounds annually yielding an OR = 1.8 (Hoffman *et al*., 2017).

<u>Other demographic factors</u>
Level of education was evaluated as potential risk factor by the three independent analyses of the NHANES hearing loss data (Agrawal *et al*., 2008; Carroll *et al*., 2017; Hoffman *et al*., 2017). In each of the studies, level of education was significantly associated with hearing loss. Increased level of education was associated with a reduced risk of hearing loss, when adjusting for other risk factors including age, gender, race, and occupational noise exposure. Hoffman *et al*. (2017) evaluated the association of the putative comorbidities (Besser *et al*., 2008) hypertension, cardiac disease and diabetes with hearing loss in the NHANES population. ORs for these conditions,

adjusting for age, race, gender, education, and noise, ranged from 0.9-1 are were not statistically significant (Hoffman *et al*., 2017). Likewise, despite published studies suggesting that smoking is associated with hearing loss, Hoffman *et al*. (2017) found no association between cigarette pack years and hearing loss at speech frequency in the NHANES population when adjusting for cofactors. Thus, education, but not hypertension, cardiovascular disease, diabetes or smoking, was included in the pilot IEHRP model for hearing loss.

Genetic susceptibility as risk factor for hearing loss

Two studies were identified with potentially useful data on genetic susceptibility for hearing loss: Zhang *et al*. (2019) and Wells *et al*., (2019). Zhang *et al*. evaluated 83 candidate SNPs in 476 subjects with NIHL (defined as > 1 year occupational noise exposure and high frequency hearing threshold > 45 dB) and 476 matched controls in China. Seven SNPs in the CDH23, PCDH15, EYA4, MYO1A, KCNMA1, and OTOG genes were significantly ($P < 0.05$) associated with the risk of NIHL, whereas seven other SNPs were marginally ($P > 0.05$ and $P < 0.1$) associated with the risk of NIHL. In general, ORs for individual SNPs were < 1.5, though two SNPs had ORs > 2.

Wells *et al.* (2019) performed a GWAS analysis of the UKBB DNA data set and identified 41 independent SNPs with a statistical association to self-reported hearing loss from a population of > 250,000 subjects. The SNPs, and their ORs and associated CIs are provided in Table 2.

Final model structure

Based on the data described above, the risk factors for the initial model were designated to be: exposure (occupational noise, firearm noise), demographic factors (race/ethnicity, gender, age) and genetic susceptibility (SNPs). The primary data sets used for model development and evaluation were the NHANES hearing loss studies by Carroll *et al.* (2017), Agrawal *et al.* (2008), and Hoffman *et al.* (2017) and the genetic susceptibility studies of Zhang *et al.* (2019) and Wells *et al.* (2019) summarized in Tables 1-2.

Table 1. Demographic factors included in IEHRP pilot model for hearing loss

	SNHL				NIHL	
	Agrawal *et al.* (2008) n = 5742		**Hoffman *et al.* (2017)** n = 3831		**Carroll *et al.* (2017)** n = 3583	
	OR	**95% CI**	**OR**	**95% CI**	**OR**	**95% CI**
Age						
20-29	1		1		1	
30-39	3.3	0.9-12.0	1.1	0.3-4.4	1.4	0.98-2
40-49	9.5	3.3-28.0	3.3	0.8-13.3	1.72	1.28-2.31
50-59	33	10-112	13.4	2.8-63.5	1.58	1.04-2.42
60-69	101	29-344	39.5	10.5-149	1.09	0.66-1.82
Gender						
Female	1		1		0.35	0.19-0.66
Male	2.4	1.7-3.5	1.8	1.1-3.0	1	
Race/Ethnicity						
White, non-Hispanic	1		2.3	1.3-3.9	1	
Black, non-Hispanic	0.4	0.2-0.5	1		0.54	0.35-0.82
Mexican American	0.7	0.4-1.1	1.4	0.8-2.3	1.93	1.09-3.42
Asian, non-Hispanic			2.1	1.1-4.2		
Other Hispanic			1.2	0.6-2.3		
Other	1.3	0.7-2.1	1.4	0.6-3.3		
Education						
< High school	1		4.2	2.1-8.5	1.75	0.89-3.42
High school/equivalency	0.4	0.3-0.6	2.8	1.2-6.9	1.78	1.26-2.51
Some college/AA			2.4	1.2-4.9		
College graduate or above			1			
> High school (combined)	0.3	0.2-0.5			1	
Occupational Noise						
No			1			
Loud, < 5 yr			1	0.5-2.2		
Loud,, > 5 yr			1.2	0.5-2.6		
Very loud, < 5 yr			1.5	0.7-3.3		
Very loud, >5 yr			1.5	0.9-2.7		
Unknown			0.5	0.1-3.1		
No					1	
Yes (combined)					1.91	1.17-3.11
Firearms						
No			1			
1- 9999 rounds			1.4	0.8-2.2		
>=1,000 rounds			1.8	1.1-3.0		

Data shown for speech frequency hearing loss as defined in the respective publications of Agrawal *et al*. (2008), Carroll *et al*. (2017), and Hoffman *et al*. (2017).

Table 2. Genetic factors included in IEHRP pilot model for hearing loss.

SNPs identified by Wells *et al*. (2019)							
CHR	SNP	Gene	Effect Allele	OR	Lower	Upper	P-value
22	rs36062310	*KLDHC7B*	G	1.032	1.026	1.038	1.90E-22
5	rs6453022	*ARHGEF28*	C	1.013	1.011	1.015	1.70E-21
6	rs759016271	*ZNF318*	AGTAGTCC	1.013	1.011	1.015	6.10E-21
5	rs6890164	*ARHGEF28*	A	1.012	1.010	1.014	3.30E-19
11	rs7951935	*TYR*	G	1.011	1.009	1.013	7.80E-17
6	rs35186928	*HLA-DQA1*	G	1.011	1.009	1.013	1.70E-15
6	rs9493627	*EYA4*	G	1.010	1.008	1.012	1.40E-13
22	rs132929	*BAIAP2L2*	G	1.010	1.008	1.012	2.20E-13
22	rs5756795	*TRIPOBP*	T	1.009	1.007	1.011	5.10E-12
14	rs1566129	*NID2*	T	1.009	1.007	1.011	1.40E-11
4	rs35414371	*CLRN2*	T	1.013	1.009	1.017	1.60E-11
3	3:182069497 TA T	*ATP11B*	TA	1.012	1.008	1.016	4.10E-11
11	rs12225399	*PHLDB1*	G	1.009	1.007	1.011	8.60E-11
11	rs55635402	*TUB*	A	1.011	1.007	1.015	2.90E-10
16	rs62033400	*FTO*	A	1.009	1.007	1.011	2.90E-10
8	rs13277721	*AGO2*	G	1.008	1.006	1.010	3.30E-10
2	rs62188635	*KLF7*	C	1.008	1.006	1.010	4.70E-10
6	rs2236401	*SYNJ2*	C	1.008	1.006	1.010	9.30E-10
7	rs4947828	*GRB10*	T	1.010	1.006	1.014	1.00E-09
10	rs6597883	*CTBP2*	T	1.011	1.007	1.015	1.00E-09
5	rs34442808	*MCTP1,SLF1*	T	1.008	1.006	1.010	1.30E-09
10	rs835267	*EXOC6*	A	1.008	1.006	1.010	1.60E-09
10	rs4948502	*ARID5B*	T	1.008	1.006	1.010	1.70E-09
10	rs10824108	*ADK*	G	1.008	1.006	1.010	3.00E-09
1	rs12027345	*MAST2*	G	1.008	1.006	1.010	3.60E-09
6	rs217289	*SNAP91*	G	1.008	1.006	1.010	4.90E-09
3	rs13093972	*ZBTB20*	A	1.008	1.006	1.010	5.50E-09
15	rs62015206	*MAPK6*	C	1.008	1.006	1.010	7.70E-09
5	rs10475169	*IRX2*	A	1.012	1.008	1.016	9.30E-09
17	rs17671352	*ACADVL*	T	1.008	1.006	1.010	1.00E-08
1	rs7525101	*LMX1A*	C	1.008	1.006	1.010	1.50E-08
17	rs12938775	*PAFAH1B1*	G	1.008	1.006	1.010	1.60E-08
8	rs76837345	*CHMP4C*	A	1.015	1.009	1.021	1.90E-08
6	rs9366417	*SOX4*	G	1.009	1.005	1.013	2.10E-08
8	rs3890736	*GFRA2*	G	1.008	1.006	1.010	2.20E-08
10	rs143282422	*CDH23*	G	1.036	1.023	1.048	2.40E-08
7	rs9691831	*TMEM213*	A	1.007	1.005	1.009	3.10E-08
11	rs141403654	*AGBL2*	A	1.032	1.020	1.044	3.50E-08
18	rs4611552	*CCDC68*	T	1.009	1.005	1.013	3.60E-08
13	rs12552	*OLFM4*	A	1.007	1.005	1.009	4.80E-08
1	rs10927035	*AKT3*	C	1.008	1.006	1.010	4.90E-08
SNPs identified by Zhang *et al*. (2019)							
CHR	SNP	Gene	Effect Allele	OR	Lower	Upper	P-value
10	rs2394795	*CDH23*	T	1.44	1.01	2.05	0.043
6	rs212769	*EYA4*	A	1.43	1.01	2.02	0.042
6	rs3777781	*EYA4*	A	0.72	0.52	0.99	0.049
10	rs696211	*KCNMA1*	C	0.48	0.26	0.87	0.016
12	rs1552245	*MYO1A*	A	0.7	0.51	0.96	0.028
11	rs7106021	*OTOG*	A	1.43	1.02	2.00	0.036
10	rs11004085	*Pcdh15*	C	0.59	0.41	0.84	0.004

CHR = chromosome

Data shown for hearing loss as defined in the respective publications of Wells *et al*. (2019) and Zhang *et al*., (2019).

4.2 Evaluating IEHRP pilot model with published data

4.2.1. Noise-induced hearing loss using the Carroll *et al.* (2017) and Zhang *et al.* (2019) data

The genetic data for the 14 statistically significant SNPs identified in Zhang *et al*. (2019), together with demographic data (age, gender, race) from Carroll *et al*. (2017), were input into the published REGENT model, in a first effort to test the utility of the model structure. The model output is shown in Figure 7. The model predicts that the majority of the population is indistinguishable from the average using this input data. Only 3% of the population is predicted to have elevated risk, while none of the population is predicted to be in the high risk category.

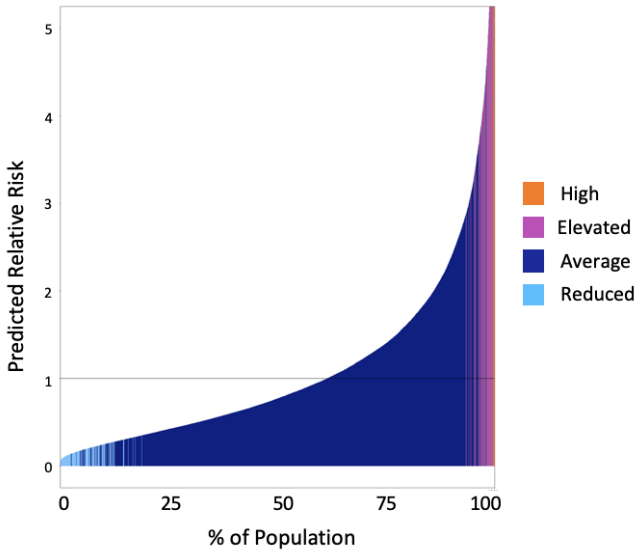

Figure 7. IEHRP pilot model for NIHL using data from Carroll *et al*. (2017) for noise-induced hearing loss, including age, race, gender and genetic susceptibility from Zhang *et al*. (2019).
Demographic data from Carroll et al. (2017). Genetic data includes 7 statistically significant SNPs from Zhang et al. (2019). Orange indicates high-risk individuals. Purple indicates elevated risk. Dark blue indicates average risk. Light blue indicates reduced risk individuals.

The lack of resolution between categories and the lack of representation in the elevated and high risk categories indicates either that the risk factors do not confer very high risk (i.e., low odds ratios) or that the uncertainty is very high (i.e., large confidence intervals) for the associated risk factors.

Regarding these possibilities, we note that in their evaluations, Carroll *et al.* (2017) defined noise-induced hearing loss as an audiogram with a "high-frequency audiometric notch when any threshold at 3, 4, or 6 kHz exceeded the average threshold at 0.5 and 1 kHz by ≥ 15 decibel (dB) hearing level (HL), and the 8 kHz threshold was at least 5 dB HL lower (better) than the maximum threshold at 3, 4, or 6 kHz." Thus, they filtered the NHANES data to only consider

tests that fell within this range as positive responders. This data set, which excludes any sensorineural hearing loss associated with factors other than noise, showed much lower odds ratios (OR of ~2) than the papers that evaluated overall hearing loss (ORs of 3-100 when also including other factors such as age). Because relevant risk factors are excluded, we may not have sufficient resolution in these data to discriminate between average and elevated risk, except in the most extreme cases. To test this, we then evaluated the model against NHANES data using odds ratios from Hoffman *et al*. (2017) developed for non-specific (age-related, sensorineural) hearing loss which have higher ORs (e.g., up to 100 for age).

4.2.2. Sensorineural hearing loss using the Hoffman *et al.* (2017) and Zhang *et al.* (2019) data

As opposed to Carroll *et al.* (2017), Hoffman *et al.* (2017) analyzed the association of various risk factors with non-specific (or sensorineural) hearing loss. The odds ratios for some risk factors (e.g., age) were much greater than those reported for noise-induced hearing loss in Carroll *et al.* (2017) (e.g., 100 vs 2 for ages 60-69). We tested whether these data would better support definition of average vs. elevated risk in the tested populations; the genetic data for the 14 statistically significant SNPs identified in Zhang *et al.* (2019), together with demographic data (age, gender, race) from Hoffman *et al.* (2017), were input into the IEHRP pilot model. The model output is shown in Figure 8.

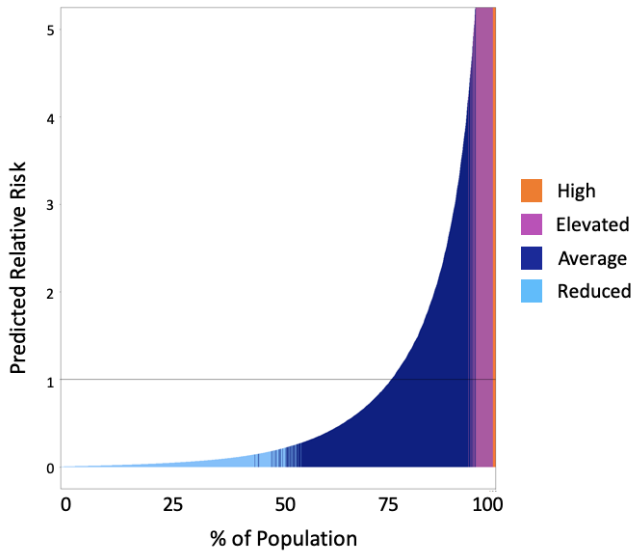

Figure 8. IEHRP pilot model for SNHL using risk factors from Hoffman *et al.* (2017) for sensorineural hearing loss, including age, race, gender and genetic susceptibility from Zhang *et al.* (2019).
Demographic and exposure data from Hoffman et al. (2017). Genetic data from Zhang et al. (2019). Orange indicates high-risk individuals. Purple indicates elevated risk. Dark blue indicates average risk. Light blue indicates reduced risk individuals.

Interestingly, while the model's discrimination of reduced vs. average risk was improved, there was still only a small portion of the population (6%) identified as elevated/high risk. While we are only in the first spiral stage of development for this model and expect that expanded efforts to describe chemical and noise exposures will improve model predictivity, we have expectation that a larger portion of the population should be at elevated risk when age, race, education and gender are considered. In the population that was used for the Hoffman paper, approximately 24% of the population tested positive for hearing loss using pure tone testing.

To explore why the model appears to be under predicting risk, we evaluated the model predictions using only the demographic data from Hoffman *et al.* (2017) (excluding any genetic component). Using only the demographic data from Hoffman, it is reasonable to assume that if all of the key risk factors were accounted for in the Hoffman study, the model should then predict an "at risk" population that is consistent with the number of subjects identified to have experienced hearing loss (~24%). The model prediction is shown in Figure 9.

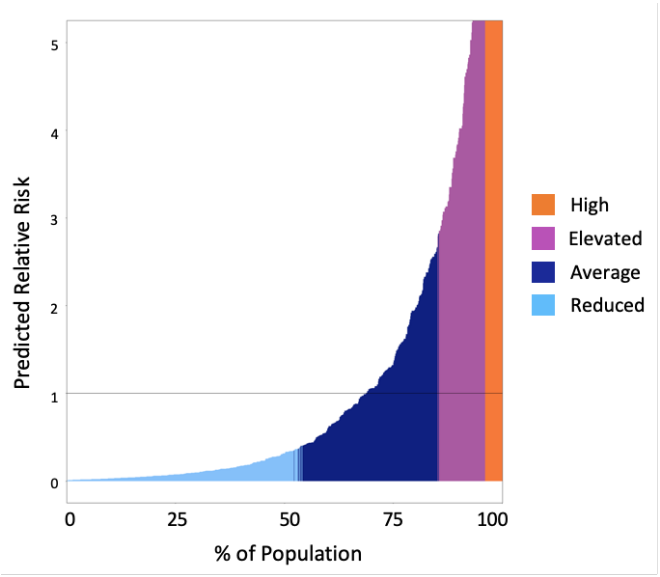

Figure 9. IEHRP pilot model prediction using risk factors from Hoffman *et al.* (2017) for sensorineural hearing loss using only demographic factors - excluding genetic susceptibility.
Demographic data from Hoffman et al. (2017). Orange indicates high-risk individuals. Purple indicates elevated risk. Dark blue indicates average risk. Light blue indicates reduced risk individuals.

By excluding the genetic data, the model is better able to discriminate between reduced, average, elevated, and high-risk populations. This model also predicts that 11% of the population would be at elevated or high risk, which is more consistent with the Hoffman study (assuming that most or all of the ~24% of subjects that experience hearing loss would be at elevated or high risk).

The fact that the model is improved by excluding the genetic data indicates that the genetic data are impairing the model prediction. To explore this, we ran a number of scenarios with the

model. First, the model was run using only the genetic data as risk factors. As shown in Figure 10, the model cannot discriminate between average and reduced or elevated risk except at the most extreme cases. We also ran several iterations of the model using the Hoffman *et al.* (2017) demographic data together with subsets of the genetic data (e.g., only the 7 highly associated SNPs, only the 3 SNPs with the highest ORs, etc.). These runs showed that the model's ability to discriminate between average and elevated risk improved as the number of genetic factors were reduced (data not shown). Thus, based on all of these runs, we would conclude that the current genetic susceptibility data are not sufficiently strong to add value to the model.

This is likely due to the fact that the study from which these data are taken was relatively small (952 subjects). The reported odds ratios are low (typically < 1.5), and the confidence intervals are relatively large. Thus, a priority for improving this model would be finding – or creating – a more robust genetic data set.

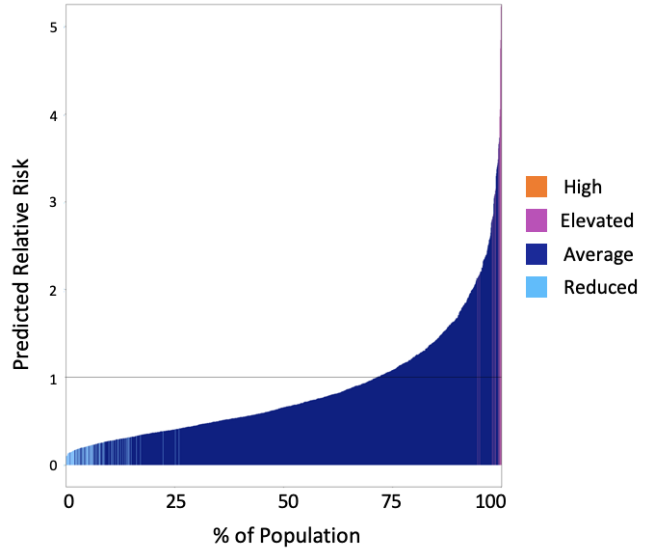

Figure 10. IEHRP pilot model prediction for sensorineural hearing loss using only genetic susceptibility from Zhang *et al.* (2019).
Orange indicates high-risk individuals. Purple indicates elevated risk. Dark blue indicates average risk. Light blue indicates reduced risk individuals.

A recent study by Wells *et al.* (2019) evaluated the role of genetics in hearing loss using the UK Biobank database. Wells and co-authors identified 41 independent SNPs that had a statistically significant association with self-reported hearing loss (Table 2). This study had the advantage of a vastly increased study size (n > 250,000) compared to the Zhang *et al.* (2019) study (n < 1,000). The Wells analysis also evaluated a broader range of SNPs and included a more diverse population compared to the Zhang study. Because of the large sample size, the Wells data set is likely to provide a higher level of resolution and improved statistical power compared to Zhang *et al.* (2019), which may be the reason more SNPs were identified in the Wells study. However,

like Zhang et al. (2019), the odds ratios for the individual SNPs identified in the Wells study were quite small (< 1.2).

We ran the IEHRP model with the demographic and exposure data from each of the NHANES analyses for SNHL (Hoffman *et al.,* 2017; Agrawal *et al*., 2008) and NIHL (Carroll *et al*., 2017) with and without the genetic data of Wells *et al*., 2019) (Figures 11-13). Because access to the raw SNP data from the UK Biobank requires a time-intensive application process (approximately 0.5 – 1 year), the following analyses used the summary data provided in the Wells *et al*. (2019) publication. These summary data allow us to replicate the analyses performed by Wells and co-authors, but does not support further analysis of additional risk factors or gene by environment interactions that may exist.

As opposed to the Zhang data, the addition of the Wells genetic data does not negatively affect the ability of the model to distinguish between reduced, average, elevated, and high risk categories. However, the genetic factors appear to slightly reduce the overall risk of hearing loss.

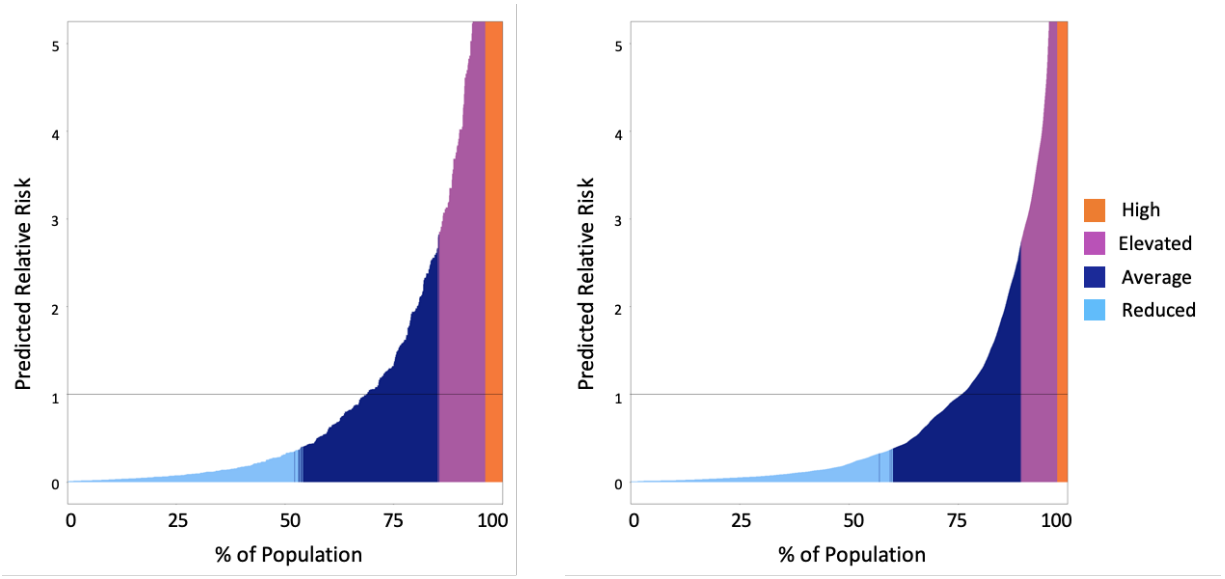

Figure 11. IEHRP pilot model for SNHL using the demographic and exposure risk factors from Hoffman *et al.* (2017) (left) or the demographic, exposure, and genetic factors from Hoffman *et al.* (2017) and Wells *et al.* (2019) (right).
Orange indicates high-risk individuals. Purple indicates elevated risk. Dark blue indicates average risk. Light blue indicates reduced risk individuals.

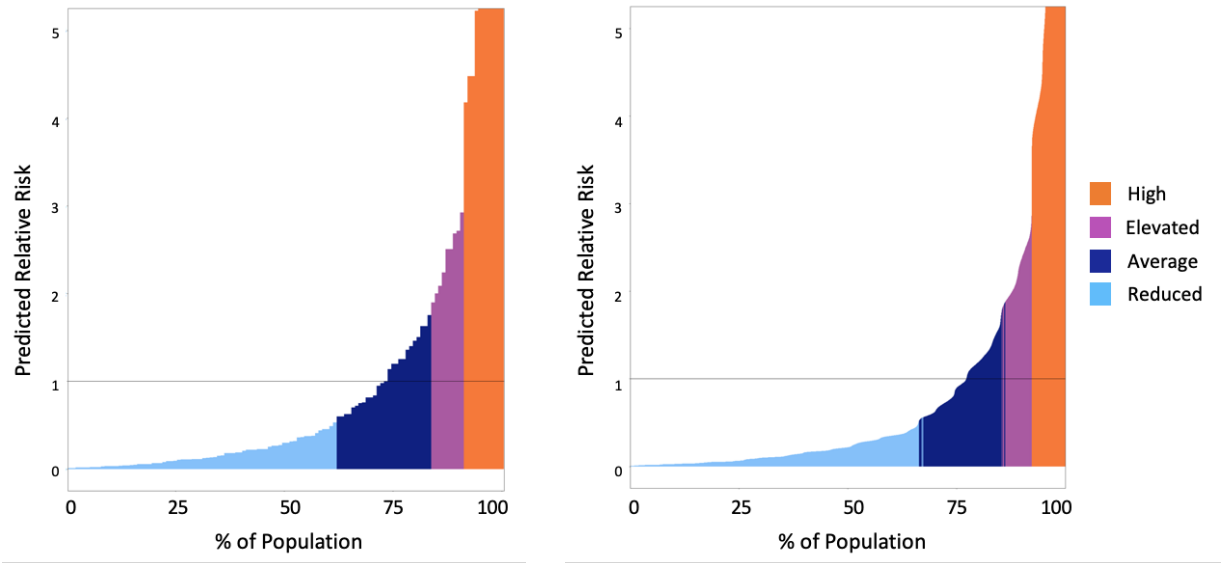

Figure 12. IEHRP pilot model for SNHL using the demographic and exposure risk factors from Agrawal *et al.* (2008) (left) or the demographic, exposure, and genetic factors from Agrawal *et al.* (2008) and Wells *et al.* (2019) (right). *Orange indicates high-risk individuals. Purple indicates elevated risk. Dark blue indicates average risk. Light blue indicates reduced risk individuals.*

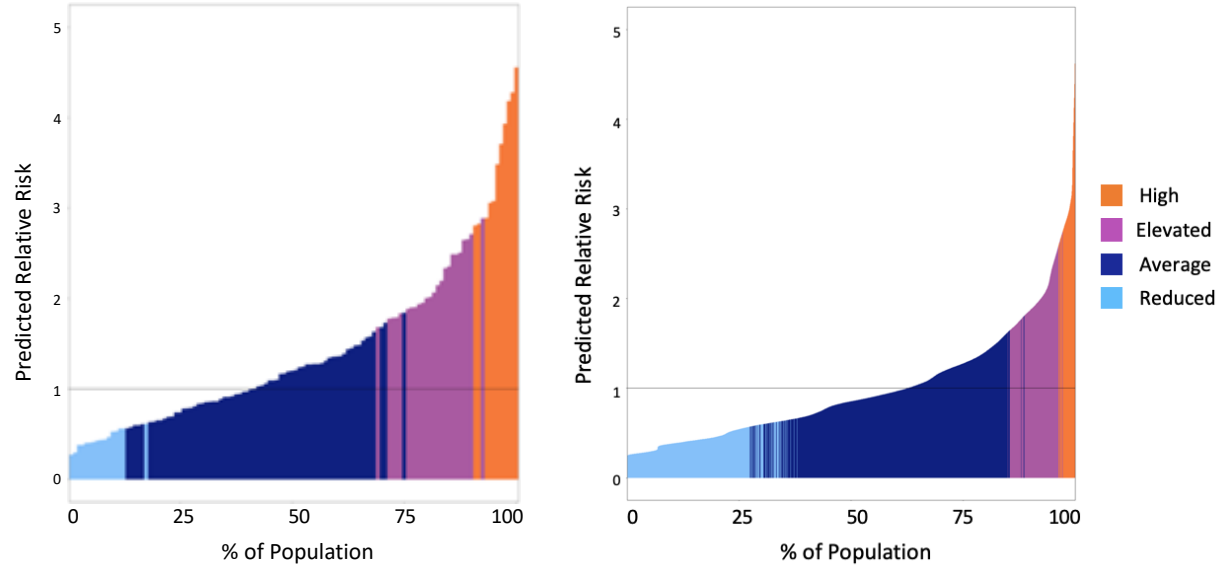

Figure 13. IEHRP pilot model for NIHL using the demographic and exposure risk factors from Carroll *et al.* (2017) (left) or the demographic, exposure, and genetic factors from Carroll *et al.* (2017) and Wells *et al.* (2019) (right).

The population distribution of the four predicted risk categories are summarized in Table 3 for each combination of demographic and genetic risk data. In each case, addition of the Zhang *et al*. (2019) genetic data reduced model resolution compared to using only the demographic and

environmental factors from Carroll *et al*., (2017), Agrawal *et al*., (2008) or Hoffman *et al*. (2017), as evidenced by larger population of "moderate risk" and reduced populations in all other categories. In contrast, addition of the Wells *et al*. (2019) genetic data from the much larger and more diverse population (n > 250,000 vs. 952) had negligible effect on population distribution, and the slight changes observed were in the direction of reduced overall risk.

Table 3. IEHRP model predicted population distributions of risk categories based on different data sources.

	NIHL			**SNHL**					
Demographic data[a]	Carroll			Hoffman			Agrawal		
Genetic data[b]	---	Zhang	Wells	---	Zhang	Wells	---	Zhang	Wells
Risk Category (% population)									
Reduced	30%	13%	32%	60%	53%	60%	66%	57%	67%
Average	57%	84%	54%	29%	42%	30%	19%	37%	19%
Elevated	11%	3%	11%	9%	5%	8%	7%	4%	7%
High	3%	0%	3%	2%	1%	2%	8%	3%	8%

[a]Demographic ORs and CIs from Agrawal *et al*. (2008), Carroll *et al*. (2017) and Hoffman *et al*. (2017).
[b]Genetic ORs and CIs from Wells *et al*. (2019) and Zhang *et al*. (2019). Dashed line indicates that genetic data was not used in the model.

4.3 Evaluating utility of public data sets to predict relative risk for Air Force personnel

It is expected that significant differences exist between the NHANES and Air Force populations, as NHANES employs a weighted sampling technique to ensure sufficient data are collected in minority populations of all ages. Further, the NHANES survey specifically excludes active duty military personnel. Thus, in order to evaluate whether the model built using the NHANES data can be useful for predicting relative risk in the Air Force population, the distribution of the demographic characteristics in the NHANES population was compared to those of current Air Force personnel.

The comparison of the distribution of three risk factors (age, gender, race/ethnicity) in the NHANES (weighted) and Air Force populations is shown in Figure 14.
A B

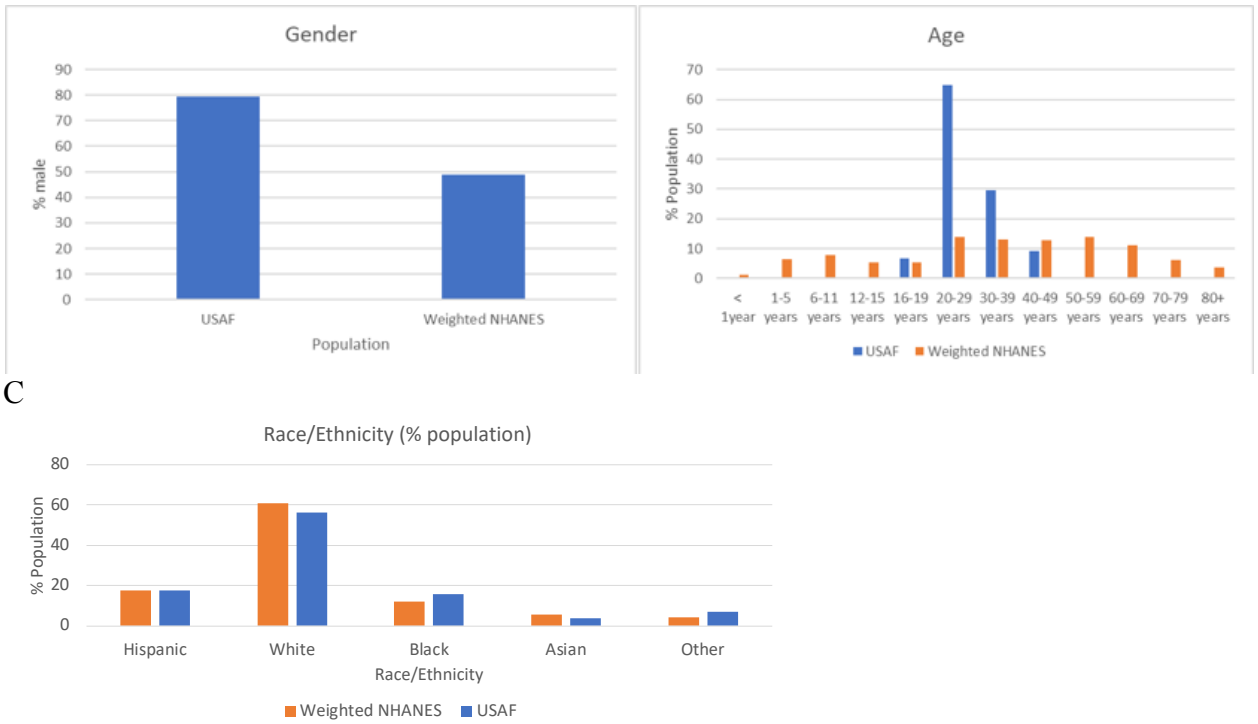

Figure 14. Comparison of the distribution of gender (A), age (B), and race/ethnicity (C) in the NHANES (weighted) and Air Force populations (actual).

While there are differences in the distributions, the NHANES data set clearly encompasses the demographics of the Air Force personnel. This indicates that a model built based-on the NHANES data should be applicable to the Air Force population.

4.4 Development of a model app

To facilitate user access to the IEHRP model, a graphical user interface was developed in R, using the Shiny plugin. Currently the app is accessed through R Studio but will ultimately be accessed via the internet, like existing 711 HPW applications, and potentially downloaded onto mobile devices. Screen captures of the current version of the app are shown below. For each figure, the numbers and notes on the side, which are in red font, have been added on top of the screen capture to highlight various features of the app for this report.

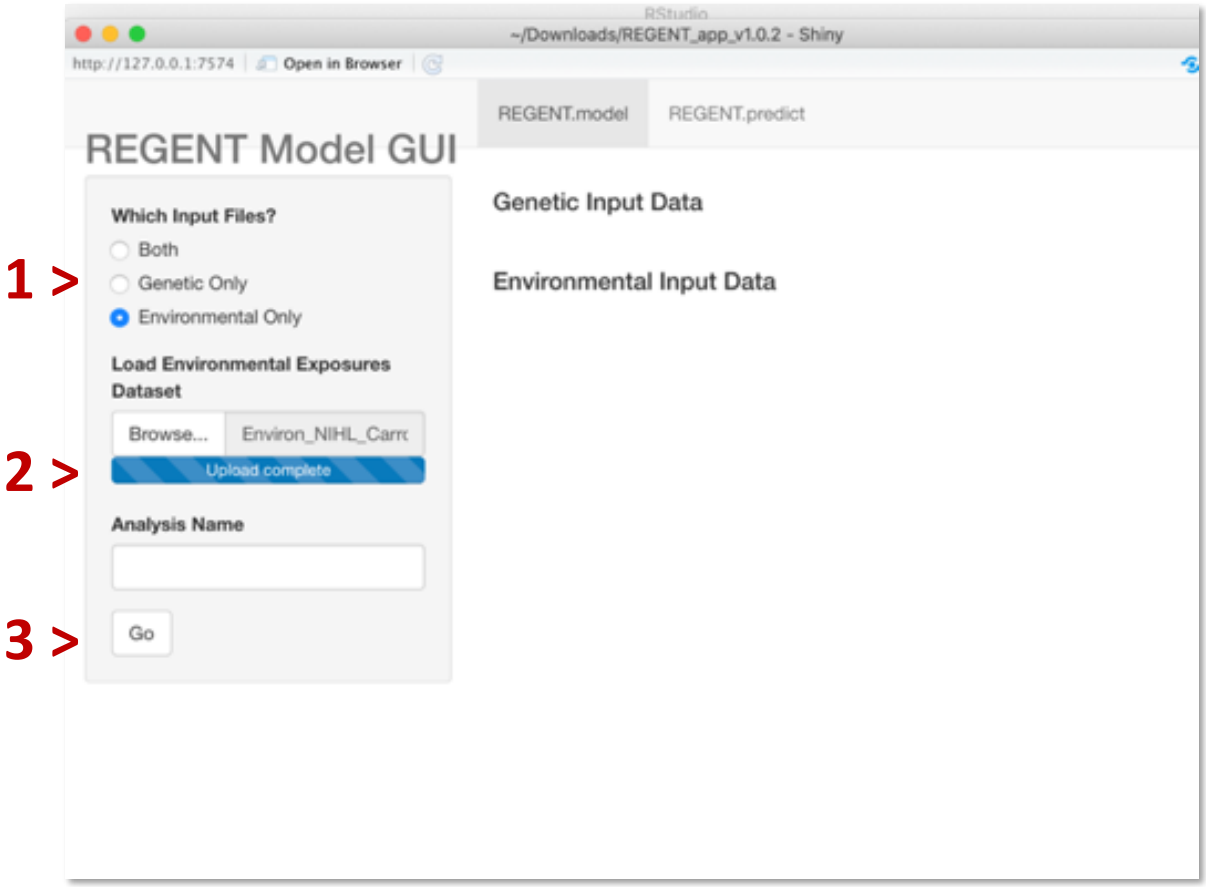

Figure 15. Screen capture of the landing page for the IEHPR model app.
In the above figure, "1 >" (in red) denotes the tabs that allow the end user to choose the kind of data they want to consider in the model. The end user may choose genetic data, environmental risk factors (demographic and exposure), or both. "2 >" denotes the feature that allows the user to choose a data set from which to build their model. This could be using different data sets for the same outcome (e.g., NHANES vs Air Force specific data) or in later versions of the model could also be used to choose which health outcome is being modeled. "3 >" denotes the action tab which tells the model to run.

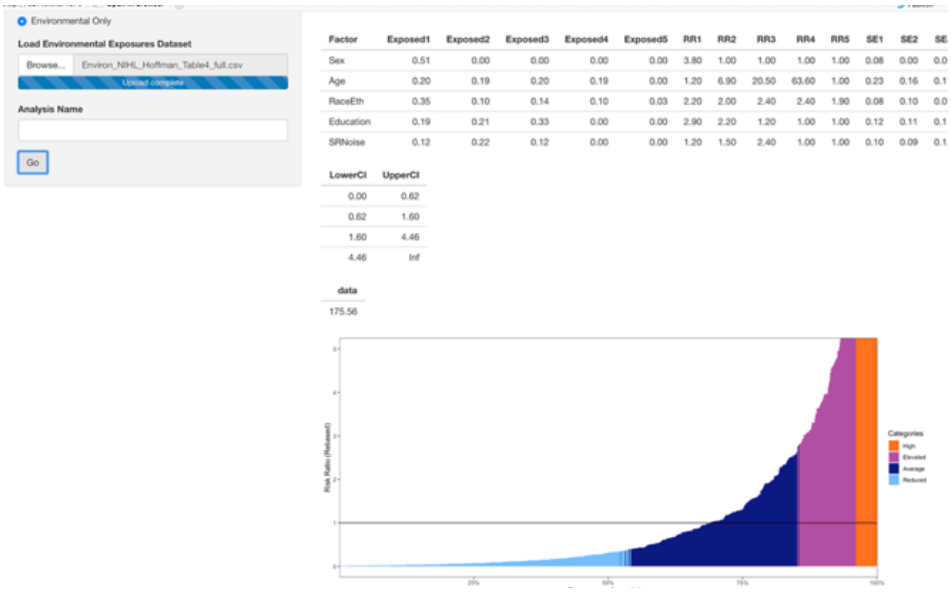

Figure 16. Screen capture of IEHRP hearing loss model app output for population level data using the NHANES data.
App output of the overall distribution of the risk categories for the modeled population.

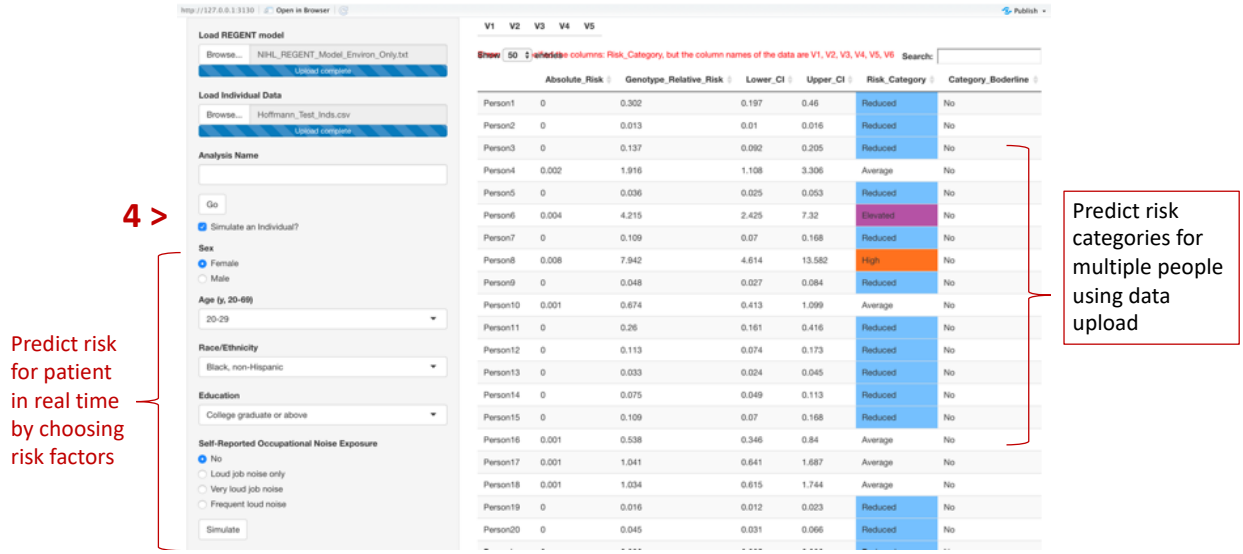

Figure 17. Screen capture of the model utility for predicting risk of individuals.
The IEHRP app allows the model to be used to predict the risk categories for multiple individuals simultaneously in batch runs, by importing data (denoted by the text box and brackets on the right). The model may also be used to evaluate a single patient in real time (denoted by the text box and brackets on the left). "4 >" (in red) denotes dropdown menus that allow the end user to choose pre-set levels for the risk factors included in the model. An example is shown here for a theoretical patient.

5.0 DISCUSSION

5.1 Model structure

Currently, it appears that the REGENT model structure is sufficient to support the first phase of a model for hearing loss (Spiral 1 development). Through modifications of the model code, it is possible to allow larger data sets, different types of data (continuous vs. discrete), and different analyses/comparisons. The computing power available to the Air Force will likely allow the model to be run relatively efficiently, even though the increased size of the IEHRP model compared to the original REGENT model package will significantly increase runtime. However, we may need to revisit the issue of using the current code as the backbone for the model as the OHR&RP program is expanded to include other health endpoints. One possible approach would be to run and store the virtual population predictions, and then save the output into static files that can be accessed for individual predictions within the model App. This would bypass the need for a time-consuming step of running the population level calculations each time the end user wanted to access predictions for individuals.

In additional spirals of model development, the current hearing loss model could be expanded to consider additional risk factors, including chemical exposure and additional sources of noise (earbuds, concerts, flightline, etc.) that are likely to be risk factors for the Air Force population. Further, issues resulting from acute (impulse) noise could be included in an effort to help commanders evaluate Airman readiness for return to duty following impulse noise incidents. In addition, this model platform can be used to address any health outcomes that are both environmentally and genetically driven. Ultimately, the goal of this and future spirals of work is to develop a comprehensive model of the health of airmen, that includes the diverse health risks that may affect operational readiness of Air Force personnel.

Another approach that would be of value is the development of a machine learning model. This type of model could more efficiently utilize the large data sets of databanks such as the UKBB and would be amenable to supporting an ever-evolving model in additional spiral developments.

5.2 Supporting data

It is clear from the initial results that the ability of the model to discriminate between risk categories (average vs. reduced or elevated) is highly dependent on the sample size of the data that used to develop the ORs that are used in the model. Sample size drives the size of the confidence intervals for each OR and therefore the ability of the model to predict whether a particular risk profile may lie outside the average range. While the genetic data from the Wells study (Wells *et al.,* 2019) is based on over 250,000 subjects, the data we are currently using for demographic and exposure factors are taken from a combined NHANES population of less than 6,000 subjects. Clearly, there is room for improvement in these data. We are currently undergoing the application process to access the raw data from the UK Biobank. The analyses performed to date have relied on summary data from a published analysis (Wells *et al*., 2019) and, therefore, are limited to the factors studied by the authors of the publication (genetic risk and self-reported hearing loss). Obtaining the raw data from the UK Biobank, though a more time-intensive process than using summary data from published analyses, will enable access to

questionnaire data associated with the > 250,000 subjects with hearing loss and genetic data that were analyzed in the Wells study. The questionnaire data includes information on demographic factors already considered in the model (age, gender, ethnicity, education) as well as many other possible risk factors, including possible noise and chemical exposures and medical conditions. With such a large data set, the resolution of the model would be improved. One drawback to using this data set for the model, however, is that the UKBB does not have audiogram data – meaning that hearing loss can only be defined by self-reported hearing loss or the use of a hearing aid. Thus, in order to use the UKBB data, we must ensure a robust relationship between self-reported hearing loss and audiogram measurements.

A preliminary literature search was performed and > 20 papers were identified that address this topic. The studies are generally consistent, reporting approximately 70% agreement between self-reported hearing loss and audiograms (using audiogram as the comparator and a 25 dB HL cutoff). This specificity increases with larger dB HL cutoffs, which is reasonable considering larger dB HL cutoffs indicate more severe, and therefore noticeable, hearing loss. Specificity – or odds of self-reporting hearing loss when no change is seen in audiograms – is more variable, with reported values ranging from ~30 - 70%. This may be due to the type of questions asked. It appears that the simpler the question, the better correlation with the audiogram. For example, "Have you experienced hearing loss?" is likely to correlate more closely to audiograms than a more specific question. Methods for estimating "true" hearing loss from self-reported hearing loss have also been described in the literature (Ikeda *et al*., 2009) and could be useful for incorporating self-reported hearing loss into the IEHRP model.

Overall, the literature is relatively supportive of using a resource such as the UKBB to parameterize the model, even if the study did not include audiogram data. If the correlation of self-reported hearing loss to audiogram is consistent enough, it may be reasonable to use self-reported hearing loss based on the advantages gained in sample size/predictive power of the various risk factors.

An additional advantage to using the UK Biobank for model building is the addition of several risk factors that may be important for hearing loss. The UK Biobank questionnaires included extensive polling of work history, workplace conditions, health history, noise exposure, and even some pertinent chemical exposure (e.g., occupational volatile chemical exposures) that may improve the model predictions of overall risk for hearing loss. We have requested 176 data fields from the UK Biobank questionnaire and intend to evaluate their role in hearing loss risk in future iterations of the IEHRP hearing loss model.

5.3 Model output and communication

The model app provides a user-friendly means of accessing the model and its predictions. While evolution of the model and the data is expected, it is also expected that whatever application is used would be easily adapted to incorporate any such changes. We will continue to update the accessibility to the model based on feedback from the customer and in response to any changes to the model or data.

6.0 CONCLUSIONS

Our evaluations of the REGENT modeling framework indicate that the approach is appropriate for the purpose of predicting relative risk in a population. Despite limitations in the supporting data, the modeling framework is versatile enough that it can evolve as new data are obtained. It is expected that additional data will improve the accuracy of predictions but will not significantly impact how the results are presented.

Initial model runs showing a lack of resolution between predicted risk of the individuals in the population highlight the need for data that can be used to ground-truth (i.e., validate) the model. In particular, access to information from the ILER or DOEHRS-IH databases are paramount. We are also working to access the UK Biobank raw data, which will strengthen both the demographic and genetic data sets. As data are made available, we will continue to incorporate them into a suitable model structure.

We have identified the assumption of independence in the model as a key issue to be evaluated going forward. We have engaged mathematicians at the Air Force Institute of Technology to help with evaluation of parameter dependence.

The app provides an initial user-friendly means of accessing the model and the model predictions. While evolution of the model and the data is expected, it is likely that the app will also be easily adapted to incorporate any such changes. The model may also be incorporated into existing 711 HPW applications. Both will continue to be updated based on feedback from customers and in response to any changes to the model or data.

7.0 REFERENCES

Agrawal Y, Platz EA, Niparko JK. 2009. Risk factors for hearing loss in US adults: data from the National Health and Nutrition Examination Survey, 1999 to 2002. Otol Neurotol. 30(2):139-45.

Carroll YI, Eichwald J, Scinicariello F, Hoffman HJ, Deitchman S, Radke MS, Themann CL, Breysse P. 2017. Vital Signs: Noise-Induced Hearing Loss Among Adults - United States 2011-2017. MMWR Morb Mortal Wkly Rep. 66(5):139-144.

Crouch DJ, Goddard GH, Lewis CM. 2013. REGENT: a risk assessment and classification algorithm for genetic and environmental factors. Eur J Hum Genet. 21(1):109-11.

Cruickshanks KJ, Wiley TL, Tweed TS, Klein BE, Klein R, Mares-Perlman JA, Nondahl DM. 1998. Prevalence of hearing loss in older adults in Beaver Dam, Wisconsin. The Epidemiology of Hearing Loss Study. Am J Epidemiol. 148(9):879-86.

Hoffman HJ, Dobie RA, Losonczy KG, Themann CL, Flamme GA. 2017. Declining Prevalence of Hearing Loss in US Adults Aged 20 to 69 Years. JAMA Otolaryngol Head Neck Surg. 143(3):274-285.

Krishnamurti S. 2009. Sensorineural hearing loss associated with occupational noise exposure: effects of age-corrections. Int J Environ Res Public Health. 6(3):889-99.

Lin FR, Maas P, Chien W, Carey JP, Ferrucci L, Thorpe R. 2012. Association of skin color, race/ethnicity, and hearing loss among adults in the USA. J Assoc Res Otolaryngol. 13(1):109-17.

Nash SD, Cruickshanks KJ, Klein R, Klein BE, Nieto FJ, Huang GH, Pankow JS, Tweed TS. 2011. The prevalence of hearing impairment and associated risk factors: the Beaver Dam Offspring Study. Arch Otolaryngol Head Neck Surg. 137(5):432-9.

Pearson JD, Morrell CH, Gordon-Salant S, Brant LJ, Metter EJ, Klein LL, Fozard JL. 1995. Gender differences in a longitudinal study of age-associated hearing loss. J Acoust Soc Am. 97(2):1196-205.

Wells HRR, Freidin MB, Zainul Abidin FN, Payton A, Dawes P, Munro KJ, Morton CC, Moore DR, Dawson SJ, Williams FMK. 2019. GWAS Identifies 44 Independent Associated Genomic Loci for Self-Reported Adult Hearing Difficulty in UK Biobank. Am J Hum Genet. 105(4):788-802. PMID: 31564434

Zhang X, Ni Y, Liu Y, Zhang L, Zhang M, Fang X, Yang Z, Wang Q, Li H, Xia Y, Zhu Y. 2019. Screening of noise-induced hearing loss (NIHL)-associated SNPs and the assessment of its genetic susceptibility. Environ Health. 18(1):30.

APPENDIX A. IEHRP MODEL CODE

```
.onAttach <- function(...) {
  startMessage <- "User manual in ../library/REGENT/doc/REGENT_usermanual.pdf"
  packageStartupMessage(startMessage)
}

library(dplyr)

#Set the function and the default values
REGENT.model <- function(AnalysisName,
                         LocusFile = NULL,
                         EnvFile = NULL,
                         prev = 0.001,
                         cv = 0.05,
                         alpha = 0.05,
                         sims = 1e+05,
                         indsims = 1e+05,
                         SmallSampAdjust = 0.5,
                         BaseRange = 0.01,
                         PlotMax = 5,
                         Block = 100,
                         lowerLimit = 1,
                         upperLimit = 1,
                         convertRR = FALSE) {
  print(paste("Analysis started at", date(), sep = " "))
  log <- paste("Analysis started at", date(), sep = " ")
##############################################################################

#Read in the model and locus files/objects

##############################################################################

  print("Reading input...")
  log <- c(log, "Reading input...")
 # Make sure data is provided and provide warnings if simulations would fail.
  if (is.null(LocusFile) && is.null(EnvFile)) {
    print("ERROR: No data available - specify LocusFile or EnvFile")
    log = c(log, "ERROR: No data available - specify LocusFile or EnvFile")
    return(NULL)
  }
  if ((sims < 1e+05) && (sims > 9999)) {
    warn <- paste("Warning: low number of simulations,",
                  "confidence intervals may be inaccurate",
                  sep = "")
    print(warn)
    log <- c(log, warn)
```

```
}
if (sims < 10000) {
  warn <- paste("Warning: very low number of simulations,",
                "confidence intervals likely to be inaccurate!!",
                sep = "")
  print(warn)
  log <- c(log, warn)
}
if ((indsims < 1e+05) && (indsims > 9999)) {
  warn <- paste("Warning: low number of individuals simulationed,",
                "relatively common multilocus genotypes may be missed",
                sep = "")
  print(warn)
  log <- c(log, warn)
}
if (indsims < 10000) {
  warn <- paste("Warning: very low number of simulations, relatively",
                " common multilocus genotypes likely to be missed!!",
                sep = "")
  print(warn)
  log <- c(log, warn)
}

#Read in provided data.
Geno <- !is.null(LocusFile)
Env <- !is.null(EnvFile)
if (Geno) {
  GenoIn <- read.table(LocusFile, header = TRUE)
  path <- unlist(strsplit(LocusFile, split = "/"))
}
else {
  path <- unlist(strsplit(EnvFile, split = "/"))
}
path <- path[-length(path)]
if (length(path) > 0) {
  setwd(paste(path, collapse = "/"))
}

#Filtering is complicated by quosure nonsense.
upperLimit <- rlang::enquo(upperLimit)
lowerLimit <- rlang::enquo(lowerLimit)
#Validate, clean, and filter genetic data.
if (Geno) {
  if (sum(c("SNP", "MAF", "Ncase", "Ncontrol") %in% colnames(GenoIn)) < 4) {
    err <- paste("ERROR: Variables missing from input. ",
                 "Ensure SNP name (SNP), Minor Allele Frequency (MAF),",
```

```
            "number of cases (Ncase) and number of controls ",
            "(Ncontrol) are specified in locus file",
            sep = "")
    print(err)
    log <- c(log, err)
    return(NULL)
  }
  if ("RR" %in% colnames(GenoIn)) {
    GenoIn <- GenoIn %>%
      filter(!(RR > 1 & RR ^ 2 < !!upperLimit),
           !(RR < 1 & RR ^ 2 > !!lowerLimit))
      nSNP <- nrow(GenoIn)
    if (nrow(GenoIn) > 0) {
      p <- GenoIn[, match("MAF", colnames(GenoIn))]
      ncase <- GenoIn[, match("Ncase", colnames(GenoIn))]
      ncontr <- GenoIn[,match("Ncontrol", colnames(GenoIn))]
      OR <- GenoIn[, match("RR", colnames(GenoIn))]
      OR[p > 0.5] <- 1 / OR[p > 0.5]
      MltplctvOR <- TRUE
    }
    else {
      Geno <- FALSE
    }
  }
  else {
    if (("RR_het" %in% colnames(GenoIn)) &&
       ("RR_hom" %in% colnames(GenoIn))) {
      GenoIn <- GenoIn %>%
        filter(!(RR_het > 1 & RR_het < !!upperLimit) &
                 !(RR_hom > 1 & RR_hom < !!upperLimit),
               !(RR_het < 1 & RR_het > !!lowerLimit) &
                 !(RR_hom < 1 & RR_hom > !!lowerLimit))
      nSNP <- nrow(GenoIn)
      OR <- cbind(GenoIn[, match("RR_het", colnames(GenoIn))],
                  GenoIn[,match("RR_hom", colnames(GenoIn))])
      if (nrow(GenoIn) > 0) {
        p <- GenoIn[, match("MAF", colnames(GenoIn))]
        ncase <- GenoIn[, match("Ncase", colnames(GenoIn))]
        ncontr <- GenoIn[,match("Ncontrol", colnames(GenoIn))]
        OR[p > 0.5,] <- 1 / OR[p > 0.5,]
        MltplctvOR <- FALSE
      }
      else {
        Geno <- FALSE
      }
    }
```

```
    else {
      err <- paste("ERROR: Ensure risk ratios are present,",
                   " with column name OR for allelic risk ratios or two",
                   " columns RR_het and RR_hom for genotypic risk ratios",
                   sep = FALSE)
      print(err)
      log <- c(log, err)
      return(NULL)
    }
  }
  if (nSNP > 0) {
  p[p > 0.5] <- 1 - p[p > 0.5]
  q <- 1 - p
  }
  else {
    Geno <- FALSE
    nSNP <- 0
    p <- integer(0)
    q <- integer(0)
    MltplctvOR <- TRUE
    OR <- NULL
    GenoIn <- NULL
  }

}
else {
  nSNP <- 0
  p <- integer(0)
  q <- integer(0)
  MltplctvOR <- TRUE
  OR <- NULL
  GenoIn <- NULL
}
#Validate, clean, and filter enviromental data.
MlvlEF <- TRUE
if (Env) {
  EnvIn <- read.table(EnvFile, header = TRUE)
  if (!("Factor" %in% colnames(EnvIn))) {
    err <- paste("ERROR: Ensure factor name (Factor)is specified",
                 "in environmental file",
                 sep = "")
    print(err)
    log <- c(log, err)
    return(NULL)
  }
  EnvIn <- EnvIn %>%
```

```
  filter_at(vars(starts_with("RR")),
            any_vars(!(. < 1 & . > !!lowerLimit) & !(. > 1 & . < !!upperLimit)))
nEnv <- nrow(EnvIn)
ORnames <- paste("RR", 1:((dim(EnvIn)[2] - 1) / 3), sep = "")
if (length(ORnames) == 1) {
  ORnames <- "RR"
  MlvlEF <- FALSE
}
Exnames <- paste("Exposed", 1:((dim(EnvIn)[2] - 1) / 3), sep = "")
if (length(Exnames) == 1) {
  Exnames <- "Exposed"
  MlvlEF <- FALSE
}
SEnames = paste("SE", 1:((dim(EnvIn)[2] - 1) / 3), sep = "")
if (length(SEnames) == 1) {
  SEnames <- "SE"
  MlvlEF <- FALSE
}
if (MltplctvOR) {
  if (!MlvlEF) {
    OR <- c(OR, EnvIn[, match(ORnames, colnames(EnvIn))])
  }
  else {
    genoFill <- matrix(rep(NA, length(OR) * (length(ORnames) - 1)),
                       nrow = length(OR))
    if (length(OR) > 0) {
      OR <- rbind(as.matrix(cbind(OR, genoFill)),
                  as.matrix(EnvIn[, match(ORnames, colnames(EnvIn))]))
    }
    else {
      OR <- as.matrix(EnvIn[, match(ORnames, colnames(EnvIn))])
    }
  }
}
else {
  if (!MlvlEF) {
    OR <- rbind(OR, cbind(EnvIn[, match("RR", colnames(EnvIn))],
                          rep(NA, sum("RR" %in% colnames(EnvIn)))))
  }
  else {
    if (length(OR) > 0) {
      if (length(ORnames) > 2) {
        genoFill <- matrix(rep(NA, length(OR) * ((length(ORnames) - 2))),
                           nrow = length(OR))
      OR <- cbind(OR, genoFill)
      }
```

```
          OR <- rbind(as.matrix(OR),
                      as.matrix(EnvIn[, match(ORnames, colnames(EnvIn))]))
        }
        else {
          OR <- as.matrix(EnvIn[, match(ORnames, colnames(EnvIn))])
        }
      }
    }
    if (nrow(EnvIn) > 0) {
      colnames(OR) <- NULL
      Ex <- as.matrix(EnvIn[, match(Exnames, colnames(EnvIn))])
      nEx <- 1 - Ex
      SE <- as.matrix(EnvIn[, match(SEnames, colnames(EnvIn))])
      nEnv <- nrow(EnvIn)
      if (sum(rowSums(Ex) > 1) > 0) {
        err <- "ERROR: sum of environmental exposure levels greater than 1"
        print(err)
        log <- c(log, err)
        return(NULL)
      }
    }
    else {
      nEnv <- 0
      Ex <- integer(0)
      nEx <- integer(0)
      EnvIn <- NULL
      Env <- FALSE
    }
  }
  else {
    nEnv <- 0
    Ex <- integer(0)
    nEx <- integer(0)
    EnvIn <- NULL
  }
  nVar <- nSNP + nEnv
  if (nVar == 0) {
    err <- "All factors have been filtered out."
    print(err)
    log <- c(log, err)
    return(NULL)
  }
################################################################################

#Calculate the multifactoral risks for the population
```

```
##############################################################################

  print("Simulating population of genotypes...")
  log <- c(log, "Simulating population of genotypes...")
  pop <- matrix(nrow = indsims, ncol = nVar, rep(0, nVar * indsims))
  if (Geno) {
    for (i in 1:nSNP) {
      rand <- runif(indsims, 0, 1)
      pop[rand <= p[i], i] <- 1
      rand <- runif(indsims, 0, 1)
      pop[rand <= p[i], i] <- pop[rand <= p[i], i] + 1
    }
  }
  if (Env) {
    for (i in 1:nEnv) {
      rand <- runif(indsims, 0, 1)
      for (j in dim(Ex)[2]:1) {
        pop[rand <= sum(Ex[i, 1:j]), nSNP + i] <- j
      }
    }
  }
  Mp <- table(apply(X = pop, MARGIN = 1, FUN = paste, collapse = "")) / indsims
  Combos <- t(matrix(as.numeric(unlist(strsplit(rownames(Mp),
                                    split = ""))), nrow = nVar))
  nGeno <- dim(Combos)[1]
  rownames(Mp) <- NULL
  pop <- NULL
  print("Calculating point estimates...")
  log <- c(log, "Calculating point estimates...")
  MGRR <- rep(1, nGeno)
  rsim <- rnorm(sims, 1, sd = cv)
  if (Geno) {
    gsim <- array(dim = c(sims, nSNP, 3))
    gsim[, , 1] <- 1
    if (MltplctvOR) {
      for (i in 1:nSNP) {
        MGRR <- MGRR * OR[i] ^ Combos[, i]
        #Create therorical experiment based on values provided
        #Note: SamllSampAdjust is there to avoid divison by 0
        con0 <- (q[i] ^ 2) * ncontr[i]
        con1 <- 2 * p[i] * q[i] * ncontr[i]
        con2 <- (p[i] ^ 2 * ncontr[i]) + SmallSampAdjust
        denom <- (1 + (OR[i] - 1) * p[i]) ^ 2
        case0 <- ((q[i] ^ 2) * ncase[i]) / denom
        case1 <- (2 * OR[i] * p[i] * q[i] * ncase) / denom
        case2 <- ((((OR[i] * p[i]) ^ 2) * ncase) / denom) + SmallSampAdjust
```

```
      #Compute SD of the RR from theorical experiment above
      sdLog1 <- sqrt((1 / case0) + (1 / case1) + (1 / con0) + (1 / con1))
      sdLog2 <- sqrt((1 / case0) + (1 / case2) + (1 / con0) + (1 / con2))
      #Simulate OR
      gsim[, i, 2] <- rlnorm(sims, meanlog = log(OR[i]), sdlog = sdLog1)
      gsim[, i, 3] <- rlnorm(sims, meanlog = log(OR[i]^2), sdlog = sdLog2)
      sdFreq <- sqrt((p[i] * q[i]) / ncontr[i])
      lsim <- rnorm(sims, mean = p[i], sd = sdFreq)
      lsim <- (1 + (gsim[, i, 2] - 1) * lsim) ^ 2
      lsim <- mean(lsim)/lsim
      gsim[, i, ] <- gsim[, i, ] * lsim
    }
  }
  else {
    for (i in 1:nSNP) {
      MGRR <- MGRR * cbind(rep(1, nSNP),
                    as.matrix(OR)[1:nSNP,])[i, Combos[, i] + 1]
      #Create therorical experiment based on values provided
      #Note: SamllSampAdjust is there to avoid divison by 0
      con0 <- (q[i] ^ 2) * ncontr[i]
      con1 <- 2 * p[i] * q[i] * ncontr[i]
      con2 <- (p[i] ^ 2 * ncontr[i]) + SmallSampAdjust
      denom <- (1 + (OR[i, 1] - 1) * p[i]) ^ 2
      case0 <- ((q[i] ^ 2) * ncase[i]) / denom
      case1 <- (2 * OR[i, 1] * p[i] * q[i] * ncase) / denom
      case2 <- (((OR[i ,2] * (p[i]) ^ 2) * ncase) / denom) +
        SmallSampAdjust
      #Compute SD of the RR from theorical experiemnt above
      sdLog1 <- sqrt((1 / case0) + (1 / case1) + (1 / con0) + (1 / con1))
      sdLog2 <- sqrt((1 / case0) + (1 / case2) + (1 / con0) + (1 / con2))
      gsim[, i, 2] <- rlnorm(sims, meanlog = log(OR[i, 1]),
                      sdlog = sdLog1)
      gsim[, i, 3] <- rlnorm(sims, meanlog = log(OR[i, 2]),
                      sdlog = sdLog2)
      lsim <- rnorm(sims, mean = p[i],
                sd = sqrt((p[i] * q[i]) / ncontr[i]))
      lsim <- (1 + (gsim[, i, 2] - 1) * lsim) ^ 2
      lsim <- mean(lsim) / lsim
      gsim[, i, ] <- gsim[, i, ] * lsim
    }
  }
}
if (Env) {
  esim <- array(dim = c(sims, nEnv, dim(Ex)[2] + 1))
  esim[, , 1] <- 1
  for (i in 1:nEnv) {
```

```
      MGRR <- MGRR * (c(1, as.matrix(OR)[nSNP + i, ])[!is.na(c(1, as.matrix(OR)[nSNP + i,
]))][Combos[, nSNP + i] + 1])
      for (j in 2:(dim(Ex)[2] + 1)) {
        esim[, i, j] = rlnorm(sims,
                              meanlog = log(as.matrix(OR)[nSNP + i, j - 1]),
                              sdlog = as.matrix(SE)[i, j - 1])
      }
    }
  }

#############################################################################

#Obtain confidence intervals for multilocus GRRs by simulation

#############################################################################

  print("Simulating for confidence intervals...")
  log <-  c(log, "Simulating for confidence intervals...")
  upper <- vector(length = nGeno)
  lower <- upper
  for (i in seq(0, nGeno - 1, by = Block)) {
    if (i <= (nGeno) - Block) {
      risksim <- matrix(rep(rsim, Block), nrow = sims, ncol = Block)
      if (Geno) {
        for (j in 1:nSNP) {
          risksim <- risksim * gsim[, j,
                                    Combos[(i + 1):(i + Block), j] + 1]
        }
      }
      if (Env) {
        for (j in 1:nEnv) {
          risksim <- risksim * esim[, j,
                                    Combos[(i + 1):(i + Block),
                                           nSNP + j] + 1]
          }
        }
        sorted <- as.matrix(apply(X = risksim,
                                  MARGIN = 2,
                                  FUN = quantile,
                                  probs = c(alpha/2, 1 - (alpha/2))),
                            nrow = 2)
        lower[(i + 1):(i + Block)] = sorted[1, ]
        upper[(i + 1):(i + Block)] = sorted[2, ]
      }
      else {
        risksim <- matrix(rep(rsim, nGeno - i),
```

```
                    nrow = sims,
                    ncol = nGeno - i)
        if (Geno) {
          for (j in 1:nSNP) {
            risksim <- risksim * gsim[, j, Combos[(i + 1):nGeno, j] + 1]
          }
        }
        if (Env) {
          for (j in 1:nEnv) {
            risksim <- risksim * esim[, j,
                                      Combos[(i + 1):nGeno, nSNP + j] + 1]
          }
        }
        sorted <- as.matrix(apply(X = risksim,
                                  MARGIN = 2,
                                  FUN = quantile,
                                  probs = c(alpha/2, 1 - (alpha/2))),
                            nrow = 2)
        lower[(i + 1):nGeno] <- sorted[1, ]
        upper[(i + 1):nGeno] <- sorted[2, ]
      }
    }

    #Order all values properly from least to greatest value
    ranks <- order(MGRR, decreasing = FALSE)
    MGRR <- MGRR[ranks]
    Mp <- Mp[ranks]
    upper <- upper[ranks]
    lower <- lower[ranks]
    Combos <- Combos[ranks, ]

    b <- order(abs(MGRR - sum(Mp * MGRR)), decreasing = FALSE)[1]
    tot <- Mp[b]
    b2 <- b
    i <- 1
    #Loop adjusts for baseline risk if it's greater than the current risk
    while (tot < BaseRange) {
      tot <- tot + Mp[b + i]
      b2 <- c(b2, b + i)
      tot <- tot + Mp[b - i]
      b2 <- c(b - i, b2)
      i <- i + 1
    }

    #Set up baseline variables
    baseline <- (sum(MGRR[b2] * Mp[b2]) / sum(Mp[b2]))
```

```
  zeroGenType <- rep(0, nSNP)
  zeroGenType[GenoIn$MAF > 0.5] <- 2
  zeroGenType <- c(zeroGenType, rep(0, nEnv))
  baseline2 <- 1
  if (MltplctvOR) {
    if (Geno) {
      baseline2 <- prod(as.matrix(OR)[1:nSNP, 1]^zeroGenType[1:nSNP])
    }
    if (Env) {
      for (i in 1:nEnv) {
        baseline2 <- baseline2 * (c(1, as.matrix(OR)[nSNP + i, ])[!is.na(c(1,
as.matrix(OR)[nSNP + i, ]))][zeroGenType[nSNP + i] + 1])
      }
    }
    else {
      if (Geno) {
        baseline2 <- baseline2 * prod(as.vector(t(cbind(rep(1, nSNP),
OR[1:nSNP]))))[(0:2)[zeroGenType[1:nSNP] + 1] + seq(1, 3 * nSNP, by = 3)])
      }
      if (Env) {
        for (i in 1:nEnv) {
          baseline2 <- baseline2 * (c(1, as.matrix(OR)[nSNP + i, ])[!is.na(c(1,
as.matrix(OR)[nSNP + i, ]))][zeroGenType[nSNP + i] + 1])
        }
      }
    }
  }
  #Adjust values for baseline risk
  baseline2 <- baseline / baseline2
  MGRRb <- MGRR / baseline
  upper <- upper / baseline
  lower <- lower / baseline

#########################################################################

#Find risk categories

#########################################################################

  print("Calculating risk categories...")
  log <- c(log, "Calculating risk categories...")
  LowerAv <- sum(lower[b2] * Mp[b2])/sum(Mp[b2])
  UpperAv <- sum(upper[b2] * Mp[b2])/sum(Mp[b2])
  LowerHi <- upper[(1:nGeno)[lower > UpperAv][1]]
  bins <- seq(log(min(MGRR)),
          log(max(MGRR)), by = (log(max(MGRR)) - log(min(MGRR))) / 1000)
```

```
  binfreq <- vector(length = length(bins) - 1)
  binlow <- binfreq
  binhigh <- binfreq
  for (i in 1:(length(bins) - 2)) {
    index <- (1:nGeno)[log(MGRR) >= bins[i]]
    binfreq[i] <- sum(Mp[index[log(MGRR)[index] < bins[i + 1]]])
    binhigh[i] <- sum(upper[index[log(MGRR)[index] < bins[i + 1]]] *
Mp[index[log(MGRR)[index] < bins[i + 1]]]) / binfreq[i]
    binlow[i] <- sum(lower[index[log(MGRR)[index] < bins[i + 1]]] *
Mp[index[log(MGRR)[index] < bins[i + 1]]])/binfreq[i]
  }
  i <- i + 1
  index <- (1:nGeno)[log(MGRR) >= bins[i]]
  binfreq[i] <- sum(Mp[index[log(MGRR)[index] <= bins[i + 1]]])
  binhigh[i] <- sum(upper[index[log(MGRR)[index] <= bins[i + 1]]] *
Mp[index[log(MGRR)[index] <= bins[i + 1]]])/binfreq[i]
  binlow[i] <- sum(lower[index[log(MGRR)[index] <= bins[i + 1]]] *
Mp[index[log(MGRR)[index] <= bins[i + 1]]])/binfreq[i]
  write.table(cbind(exp(bins[-length(bins)])/baseline, exp(bins[-1])/baseline,
              binfreq, binlow, binhigh),
            file = paste(AnalysisName, "_RRDist.txt", sep = ""),
            col.names = c("Lower_risk",
                          "Upper_risk",
                          "Frequency",
                          "Mean_lowerCI",
                          "Mean_UpperCI"),
            row.names = FALSE,
            quote = FALSE)
  #Table with risk categories
  out <- matrix(nrow = 4, ncol = 2)
  rownames(out) <- c("Reduced", "Average", "Elevated", "High")
  colnames(out) <- c("LowerCI", "UpperCI")
  out[1, ] <- c(0, LowerAv)
  out[2, ] <- c(LowerAv, UpperAv)
  out[3, ] <- c(UpperAv, LowerHi)
  out[4, ] <- c(LowerHi, Inf)
  out <- round(out, digits = 3)
  options(warn = -1) #dafuq??
  write.table(x = paste("##REGENTv1.0 by G. Goddard, D. Crouch & C. Lewis##", sep = ""),
            file = paste(AnalysisName, ".txt", sep = ""),
            quote = FALSE,
            row.names = FALSE,
            col.names = FALSE)
  write.table(x = out,
            file = paste(AnalysisName, ".txt", sep = ""),
            quote = FALSE,
```

```
            row.names = TRUE,
            col.names = TRUE,
            append = TRUE)

#Cleaning up anything extraneous in genetic dataframe
  if (Geno) {
    GenoIn <- GenoIn[, colnames(GenoIn) %in% c("SNP",
                                    "MAF",
                                    "RR",
                                    "RR_het",
                                    "RR_hom",
                                    "Ncase",
                                    "Ncontrol")]
    if ("RR" %in% colnames(GenoIn)) {
      GenoIn <- GenoIn[, match(colnames(GenoIn), c("SNP",
                                      "MAF",
                                      "RR",
                                      "Ncase",
                                      "Ncontrol"))]
    }
    else {
      GenoIn <- GenoIn[, match(colnames(GenoIn), c("SNP",
                                      "MAF",
                                      "RR_het",
                                      "RR_hom",
                                      "Ncase",
                                      "Ncontrol"))]
    }
  }

   #Cleaning up environmental exposure dataframe
  if (Env) {
    RRvec <- grep("RR", colnames(EnvIn))
    SEvec <- grep("SE", colnames(EnvIn))
    EXvec <- grep("Exposed", colnames(EnvIn))
    EnvIn <- EnvIn[, c(match("Factor", colnames(EnvIn)), EXvec, RRvec, SEvec)]
   }

   #Code that presents the eventual output
  out <- list(categories = as.matrix(out),
            baseline = round(baseline2,digits = 3),
            LocusFile = GenoIn,
            EnvFile = EnvIn)

   #Writing Files
  write.table(x = paste(rep("#", 50), collapse = ""),
```

```
          file = paste(AnalysisName, ".txt", sep = ""),
          quote = FALSE,
          row.names = FALSE,
          col.names = FALSE,
          append = TRUE)
  EnvName <- EnvFile
  if (is.null(EnvName)) {
    EnvName <- "-"
  }

  GenoName <- LocusFile
  if (is.null(GenoName)) {
    GenoName <- "-"
  }
  write.table(cbind(c("Analysis name:",
                      "Locus file:",
                      "Environment file:",
                      "Disease prevalence:",
                      "Coefficient of variation:",
                      "alpha value:",
                      "Number of RR simulations:",
                      "Size of simulated population:",
                      "Genotypes in RAM ('Block'):",
                      "Small sample adjustment:",
                      "Proportion of population used to calculate baseline confidence intervals:",
                      "Baseline RR:"), c(AnalysisName, GenoName, EnvName, prev,
                                         cv, alpha, sims, indsims, Block,
                                         SmallSampAdjust, BaseRange,
                                         round(baseline2, digits = 3))),
          file = paste(AnalysisName, ".txt", sep = ""),
          quote = FALSE,
          col.names = FALSE,
          row.names = FALSE,
          append = TRUE)
  write.table(x = paste(rep("#", 50), collapse = ""),
          file = paste(AnalysisName, ".txt", sep = ""),
          quote = FALSE,
          row.names = FALSE,
          col.names = FALSE,
          append = TRUE)
   #prev = prev/(1 - prev)
  vy <- c(0, cumsum(rev(Mp * MGRR/sum(MGRR * Mp))))
  vx <- c(0, cumsum(rev(Mp * (1 - prev * MGRR/sum(MGRR * Mp))/(1 - prev))))
  auc <- round(sum((vx[2:length(vx)] -
                      vx[1:(length(vx) - 1)]) *
                     (vy[2:length(vy)] + vy[1:(length(vy) - 1)]) * 0.5),
```

```
          digits = 3)
  Categories <- rep(2, nGeno)
  Categories[upper < LowerAv] <- 1
  Categories[lower > UpperAv] <- 3
  Categories[lower > LowerHi] <- 4
  write.table(cbind(c("Reduced", "Average", "Elevated", "High","", "AUC"),
                c(round(sum(Mp[Categories == 1]), digits = 3),
                  1 - sum(c(round(sum(Mp[Categories == 1]), digits = 3),
                        round(sum(Mp[Categories == 3]), digits = 3),
                        round(sum(Mp[Categories == 4]), digits = 3))),
                  round(sum(Mp[Categories == 3]), digits = 3),
                  round(sum(Mp[Categories == 4]), digits = 3), "", auc)),
            file = paste(AnalysisName, ".txt", sep = ""),
            quote = FALSE,
            col.names = c("Risk_Category", "Proportion_of_population"),
            row.names = FALSE,
            append = TRUE)
  write.table(x = paste(rep("#", 50), collapse = ""),
            file = paste(AnalysisName, ".txt", sep = ""),
            quote = FALSE,
            row.names = FALSE,
            col.names = FALSE,
            append = TRUE)
  if (Geno) {
    write.table(GenoIn, file = paste(AnalysisName, ".txt",sep = ""),
              append = TRUE,
              col.names = TRUE,
              row.names = FALSE,
              quote = FALSE)
    write.table(x = paste(rep("#", 50), collapse = ""),
              file = paste(AnalysisName, ".txt", sep = ""),
              quote = FALSE,
              row.names = FALSE,
              col.names = FALSE,
              append = TRUE)
  }
  if (Env) {
    write.table(EnvIn, file = paste(AnalysisName, ".txt", sep = ""),
              append = TRUE,
              col.names = TRUE,
              row.names = FALSE,
              quote = FALSE)
    write.table(x = paste(rep("#", 50), collapse = ""),
              file = paste(AnalysisName, ".txt", sep = ""),
              quote = FALSE,
              row.names = FALSE,
```

```
            col.names = FALSE,
            append = TRUE)
}
options(warn = 0)

#Generating plots
print("Printing graphs...")
log <- c(log, "Printing graphs...")
PlotMax2 <- PlotMax
if (PlotMax > max(MGRRb)) {
  PlotMax2 <- max(MGRRb)
}
colvec <- rep(4, nGeno)
colvec[Categories == 1] <- 3
colvec[Categories == 3] <- 2
colvec[Categories == 4] <- 1
colvec2 <- rainbow(6, alpha = 0.5)[colvec]
colvec3 <- rainbow(6)[colvec]
colkey <- rainbow(6)[c(1, 2, 4, 3)][c(4, 3, 2, 1) %in% Categories]
tiff(filename = paste(AnalysisName, "_RRDistCol.TIF", sep = ""),
    pointsize = 30,
    width = 1500,
    height = 1500,
    units = "px",
    antialias = "default")
barplot(MGRRb,
       width = Mp,
       border = NA,
       ylab = "Risk Ratio (Rebased)",
       ylim = c(0, PlotMax2),
       space = 0,
       xlab = "Percentage of population",
       col = colvec2, density = -100)
legend(x = 0,
      y = PlotMax2,
      as.factor(c("High",
                  "Elevated",
                  "Average",
                  "Reduced")[c(4, 3, 2, 1) %in% Categories]),
      cex = 0.8,
      col = colkey,
      pch = 19,
      title = "Risk")
axis(1,
    seq(0, 1, by = 0.1),
    c("0%", "", "", "", "", "50%", "", "", "", "", "100%"))
```

```
  abline(1, 0, col = grey(0.75))
  if (sum(MGRRb > PlotMax) > 0) {
    par(xpd = TRUE)
    arrows(1.03, PlotMax - (PlotMax/5), 1.03, PlotMax, col = colkey[1])
  }
  quiet <- dev.off()
  tiff(filename = paste(AnalysisName, "_RRDistGrey.TIF", sep = ""),
       pointsize = 30,
       width = 1500,
       height = 1500,
       units = "px",
       antialias = "default")
  barplot(MGRRb,
          width = Mp,
          border = NA,
          ylab = "Risk Ratio (Rebased)",
          ylim = c(0, PlotMax2),
          space = 0,
          xlab = "Percentage of population",
          density = -100)
  axis(1,
       seq(0, 1, by = 0.1),
       c("0%", "", "", "", "", "50%","", "", "", "", "100%"))
  abline(1, 0, col = 1)
  if (sum(MGRRb > PlotMax) > 0) {
    par(xpd = TRUE)
    arrows(1.03, PlotMax - (PlotMax/5), 1.03, PlotMax, col = grey(0.75))
  }
  quiet <- dev.off()
  print(paste("Analysis completed at", date(), sep = " "))
  log <- c(log, paste("Analysis completed at", date(), sep = " "))
  write.table(as.matrix(log),
              file = paste(AnalysisName, ".txt", sep = ""),
              quote = FALSE,
              row.names = FALSE,
              col.names = FALSE,
              append = TRUE)
  return(out)
}
 (sum(MGRRb > PlotMax) > 0) {
    par(xpd = TRUE)
    arrows(1.03, PlotMax - (PlotMax/5), 1.03, PlotMax, col = grey(0.75))
  }
  quiet <- dev.off()
  print(paste("Analysis completed at", date(), sep = " "))
  log <- c(log, paste("Analysis completed at", date(), sep = " "))
```

```
  write.table(as.matrix(log),
              file = paste(AnalysisName, ".txt", sep = ""),
              quote = FALSE,
              row.names = FALSE,
              col.names = FALSE,
              append = TRUE)
  return(out)
}
          sum(Mp[Categories == 2]) * 100,
                               sum(Mp[Categories == 3]) * 100,
                               sum(Mp[Categories == 4]) * 100)
  simplePlotData <- as_tibble(cbind(categoryNames, values))
  names(simplePlotData) <- c("Risk", "Percent")
  simplePlotData$Percent <- as.numeric(simplePlotData$Percent)
  simplePlotData$Risk <- factor(simplePlotData$Risk,
                                ordered = TRUE,
                                levels = categoryNames)
  simplePLot <- ggplot(simplePlotData,
                       aes(Risk, Percent, fill = Risk)) +
    geom_bar(stat = "identity") +
    scale_fill_manual(values = c("#85C0F9", "#0F2080","#A95AA1", "#F5793A")) +
    xlab("Risk category") +
    ylab("Percentage of simulated population") +
    scale_y_continuous(limits = c(0, max(simplePlotData$Percent) + 5)) +
    theme(legend.position = "none") +
    ggsave(simplePlotName, width = 7, height = 7, units = "in")
  quiet <- dev.off()
  print(paste("Analysis completed at", date(), sep = " "))
  log <- c(log, paste("Analysis completed at", date(), sep = " "))
  write.table(as.matrix(log),
              file = paste(AnalysisName, ".txt", sep = ""),
              quote = FALSE,
              row.names = FALSE,
              col.names = FALSE,
              append = TRUE)

  return(out)
}
```

APPENDIX B. LIST OF CITATIONS TO SUPPORT HEARING LOSS MODEL

Agrawal, Y., Platz, E. A., & Niparko, J. K. (2008). Prevalence of hearing loss and differences by demographic characteristics among US adults: Data from the National Health and Nutrition Examination Survey, 1999-2004. *Archives of Internal Medicine*, *168*(14), 1522–1530. https://doi.org/10.1001/archinte.168.14.1522

Ahmed, H., Shubina-Oleinik, O., & Holt, J. R. (2017, October 1). Emerging Gene Therapies for Genetic Hearing Loss. *JARO - Journal of the Association for Research in Otolaryngology*, Vol. 18, pp. 649–670. https://doi.org/10.1007/s10162-017-0634-8

Atik, T., Bademci, G., Diaz-Horta, O., Blanton, S. H., & Tekin, M. (2015, March 31). Whole-exome sequencing and its impact in hereditary hearing loss. *Genetics Research*, Vol. 97, p. e4. https://doi.org/10.1017/S001667231500004X

Basner, M., Brink, M., Bristow, A., De Kluizenaar, Y., Finegold, L., Hong, J., … Sörqvist, P. (2015, March 1). ICBEN review of research on the biological effects of noise 2011-2014. *Noise and Health*, Vol. 17, pp. 57–82. https://doi.org/10.4103/1463-1741.153373

Baum-Baicker, C. (1985). The health benefits of moderate alcohol consumption: A review of the literature. *Drug and Alcohol Dependence*, Vol. 15, pp. 207–227. https://doi.org/10.1016/0376-8716(85)90001-8

Bhatt, I. S., & Guthrie, O. (2017). Analysis of audiometric notch as a noise-induced hearing loss phenotype in US youth: data from the National Health And Nutrition Examination Survey, 2005–2010. *International Journal of Audiology*, *56*(6), 392–399. https://doi.org/10.1080/14992027.2017.1278799

Bhatt, I., Phillips, S., Richter, S., Tucker, D., Lundgren, K., Morehouse, R., & Henrich, V. (2016). A polymorphism in human estrogen-related receptor beta (ESRRβ) predicts audiometric temporary threshold shift. *International Journal of Audiology*, *55*(10), 571–579. https://doi.org/10.1080/14992027.2016.1192693

Bowl, M. R., & Brown, S. D. M. (2018, August 1). Genetic landscape of auditory dysfunction. *Human Molecular Genetics*, Vol. 27, pp. R130–R135. https://doi.org/10.1093/hmg/ddy158

Bowl, M. R., Simon, M. M., Ingham, N. J., Greenaway, S., Santos, L., Cater, H., … Brown, S. D. M. (2017). A large scale hearing loss screen reveals an extensive unexplored genetic landscape for auditory dysfunction. *Nature Communications*, *8*(1). https://doi.org/10.1038/s41467-017-00595-4

Burr, H., Lund, S. P., Sperling, B. B., Kristensen, T. S., & Poulsen, O. M. (2005). Smoking and height as risk factors for prevalence and 5-year incidence of hearing loss. A questionnaire-based follow-up study of employees in Denmark aged 18-59 years exposed and unexposed to noise. *International Journal of Audiology*, *44*(9), 531–539. https://doi.org/10.1080/14992020500190045

Carlsson, P. I., Van Laer, L., Borg, E., Bondeson, M. L., Thys, M., Fransen, E., & Van Camp, G. (2005). The influence of genetic variation in oxidative stress genes on human noise susceptibility. *Hearing Research*, *202*(1–2), 87–96. https://doi.org/10.1016/j.heares.2004.09.005

Carroll, Y. I., Eichwald, J., Scinicariello, F., Hoffman, H. J., Deitchman, S., Radke, M. S., … Breysse, P. (2017). Vital Signs: Noise-Induced Hearing Loss Among Adults - United States 2011-2012. *MMWR. Morbidity and Mortality Weekly Report*, *66*(5), 139–144. https://doi.org/10.15585/mmwr.mm6605e3

Chang, N. C., Ho, C. K., Lin, H. Y., Yu, M. L., Chien, C. Y., & Ho, K. Y. (2011). Association of polymorphisms of heat shock protein 70 with susceptibility to noise-induced hearing loss in the taiwanese population. *Audiology and Neurotology*, *16*(3), 168–174. https://doi.org/10.1159/000317119

Cocchiarella, L. A., Sharp, D. S., & Persky, V. W. (1995). Hearing threshold shifts, white-cell count and smoking status in working men. *Occupational Medicine*, *45*(4), 179–185. https://doi.org/10.1093/occmed/45.4.179

Crouch, D. J. M., Goddard, G. H. M., & Lewis, C. M. (2015). *Package "REGENT" Title Risk Estimation for Genetic and Environmental Traits*.

Crouch, D. J. M., Goddard, G. H. M., & Lewis, C. M. (2013). REGENT: A risk assessment and classification algorithm for genetic and environmental factors. *European Journal of Human Genetics*, *21*(1), 109–111. https://doi.org/10.1038/ejhg.2012.107

Cruickshanks, K. J., Klein, R., Klein, B. E. K., Wiley, T. L., Nondahl, D. M., & Tweed, T. S. (1998). Cigarette smoking and hearing loss: The epidemiology of hearing loss study. *Journal of the American Medical Association*, *279*(21), 1715–1719. https://doi.org/10.1001/jama.279.21.1715

Cruickshanks, K. J., Nondahl, D. M., Tweed, T. S., Wiley, T. L., Klein, B. E. K., Klein, R., … Nash, S. D. (2010). Education, occupation, noise exposure history and the 10-yr cumulative incidence of hearing impairment in older adults. *Hearing Research*, *264*(1–2), 3–9. https://doi.org/10.1016/j.heares.2009.10.008

Cruickshanks, K. J., Nondahl, D. M., Tweed, T. S., Wiley, T. L., Klein, B. E. K., Klein, R., … Nash, S. D. (2010). Education, occupation, noise exposure history and the 10-yr cumulative incidence of hearing impairment in older adults. *Hearing Research*. https://doi.org/10.1016/j.heares.2009.10.008

Cruickshanks, K. J., Wiley, T. L., Tweed, T. S., Klein, B. E. K., Klein, R., Mares-Perlman, J. A., & Nondahl, D. M. (1998). Prevalence of hearing loss in older adults in Beaver dam, Wisconsin. The epidemiology of hearing loss study. *American Journal of Epidemiology*, *148*(9), 879–886. https://doi.org/10.1093/oxfordjournals.aje.a009713

D'Aguillo, C., Bressler, S., Yan, D., Mittal, R., Fifer, R., Blanton, S. H., & Liu, X. (2019). Genetic screening as an adjunct to universal newborn hearing screening: literature review and implications for non-congenital pre-lingual hearing loss. *International Journal of Audiology*. https://doi.org/10.1080/14992027.2019.1632499

Dawes, P., Cruickshanks, K. J., Moore, D. R., Edmondson-Jones, M., McCormack, A., Fortnum, H., & Munro, K. J. (2014). Cigarette smoking, passive smoking, alcohol consumption, and hearing loss. *JARO - Journal of the Association for Research in Otolaryngology*, *15*(4), 663–674. https://doi.org/10.1007/s10162-014-0461-0

Fallis, A. . (2013). 済無No Title No Title. *Journal of Chemical Information and Modeling*, *53*(9), 1689–1699. https://doi.org/10.1017/CBO9781107415324.004

Fransen, E., Bonneux, S., Corneveaux, J. J., Schrauwen, I., Di Berardino, F., White, C. H., … Friedman, R. A. (2015). Genome-wide association analysis demonstrates the highly polygenic character of age-related hearing impairment. *European Journal of Human Genetics*, *23*(1), 110–115. https://doi.org/10.1038/ejhg.2014.56

Fransen, E., Topsakal, V., Hendrickx, J. J., Van Laer, L., Huyghe, J. R., Van Eyken, E., … Van Camp, G. (2008). Occupational noise, smoking, and a high body mass index are risk factors for age-related hearing impairment and moderate alcohol consumption is protective: A European population-based multicenter study. *JARO - Journal of the Association for Research in Otolaryngology*, *9*(3), 264–276. https://doi.org/10.1007/s10162-008-0123-1

Gates, G. A., Cobb, J. L., D'agostino, R. B., & Wolf, P. A. (1993). The Relation of Hearing in the Elderly to the Presence of Cardiovascular Disease and Cardiovascular Risk Factors. *Archives of Otolaryngology--Head and Neck Surgery*, *119*(2), 156–161. https://doi.org/10.1001/archotol.1993.01880140038006

Genome, M., White, C. H., Ohmen, J. D., Sheth, S., Zebboudj, A. F., Mchugh, R. K., … Friedman, R. A. (2009). *Genome-wide screening for genetic loci associated with noise-induced hearing loss*.

Gilels, F., Paquette, S. T., Beaulac, H. J., Bullen, A., & White, P. M. (2017). Severe hearing loss and outer hair cell death in homozygous Foxo3 knockout mice after moderate noise exposure. *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-017-01142-3

Gopinath, B., Flood, V. M., McMahon, C. M., Burlutsky, G., Smith, W., & Mitchell, P. (2010). The effects of smoking and alcohol consumption on age-related hearing loss: The blue mountains hearing study. *Ear and Hearing*, *31*(2), 277–282. https://doi.org/10.1097/AUD.0b013e3181c8e902

Greco, A., Ralli, M., Falasca, V., Greco, A., & Ralli, M. (2017). Noise induced hearing loss : The role of oxidative stress . *Otollaryngology Open J*, *5*(May), 1–5. https://doi.org/10.17140/OTLOJ-SE-5-101

Grondin, Y., Bortoni, M. E., Sepulveda, R., Ghelfi, E., Bartos, A., Cotanche, D., … Rogers, R. A. (2015). Genetic polymorphisms associated with hearing threshold shift in subjects during first encounter with occupational impulse noise. *PLoS ONE*, *10*(6). https://doi.org/10.1371/journal.pone.0130827

Guo, H., Ding, E., Bai, Y., Zhang, H., Shen, H., Wang, J., … Zhu, B. (2017). Association of genetic variations in FOXO3 gene with susceptibility to noise induced hearing loss in a Chinese population. *PLoS ONE*, *12*(12). https://doi.org/10.1371/journal.pone.0189186

Guo, H., Ding, E., Sheng, R., Cheng, J., Cai, W., Guo, J., … Zhu, B. (2018). Genetic variation in KCNQ4 gene is associated with susceptibility to noise-induced hearing loss in a Chinese population. *Environmental Toxicology and Pharmacology*, *63*, 55–59. https://doi.org/10.1016/j.etap.2018.08.009

Han, B., Yang, X., Li, Y., Hosseini, D. K., Tu, Y., Dong, Y., … Sun, H. (2019). Association of polymorphisms in grainyhead-like-2 gene with the susceptibility to age-related hearing loss: A systematic review and meta-analysis. *Medicine*, *98*(25), e16128. https://doi.org/10.1097/MD.0000000000016128

Han, B., Yang, X., Li, Y., Hosseini, D. K., Tu, Y., Dong, Y., … Sun, H. (2019). Association of polymorphisms in grainyhead-like-2 gene with the susceptibility to age-related hearing loss: A systematic review and meta-analysis. *Medicine*, *98*(25), e16128. https://doi.org/10.1097/MD.0000000000016128

Helzner, E. P., Cauley, J. A., Pratt, S. R., Wisniewski, S. R., Zmuda, J. M., Talbott, E. O., … Newman, A. B. (2005). Race and sex differences in age-related hearing loss: The health, aging and body composition study. *Journal of the American Geriatrics Society*, *53*(12), 2119–2127. https://doi.org/10.1111/j.1532-5415.2005.00525.x

Hoffman, H. J., Dobie, R. A., Ko, C. W., Themann, C. L., & Murphy, W. J. (2012). Hearing threshold levels at age 70 years (65-74 Years) in the unscreened older adult population of the united states, 1959-1962 and 1999-2006. *Ear and Hearing*, *33*(3), 437–440. https://doi.org/10.1097/AUD.0b013e3182362790

Hoffman, H. J., Dobie, R. A., Losonczy, K. G., Themann, C. L., & Flamme, G. A. (2017). Declining prevalence of hearing loss in US adults aged 20 to 69 years. *JAMA Otolaryngology - Head and Neck Surgery*, *143*(3), 274–285. https://doi.org/10.1001/jamaoto.2016.3527

Hoffmann, T. J., Keats, B. J., Yoshikawa, N., Schaefer, C., Risch, N., & Lustig, L. R. (2016). A Large Genome-Wide Association Study of Age-Related Hearing Impairment Using Electronic Health Records. *PLoS Genetics*, *12*(10). https://doi.org/10.1371/journal.pgen.1006371

Holme, R. H., & Steel, K. P. (2004). Progressive Hearing Loss and Increased Susceptibility to Noise-Induced Hearing Loss in Mice Carrying a Cdh23 but not a Myo7a Mutation. *JARO* -

Journal of the Association for Research in Otolaryngology, *5*(1), 66–79. https://doi.org/10.1007/s10162-003-4021-2

Honkura, Y., Matsuo, H., Murakami, S., Sakiyama, M., Mizutari, K., Shiotani, A., … Motohashi, H. (2016). NRF2 is a key target for prevention of noise-induced hearing loss by reducing oxidative damage of cochlea. *Scientific Reports*, *6*. https://doi.org/10.1038/srep19329

Horikawa, C., Kodama, S., Tanaka, S., Fujihara, K., Hirasawa, R., Yachi, Y., … Sone, H. (2013). Diabetes and risk of hearing impairment in adults: A meta-analysis. *Journal of Clinical Endocrinology and Metabolism*, *98*(1), 51–58. https://doi.org/10.1210/jc.2012-2119

Jain, R. K., Pingle, S. K., Tumane, R. G., Thakkar, L. R., Jawade, A. A., Barapatre, A., & Trivedi, M. (2018, May 1). Cochlear proteins associated with noise-induced hearing loss: An update. *Indian Journal of Occupational and Environmental Medicine*, Vol. 22, pp. 60–73. https://doi.org/10.4103/ijoem.IJOEM_43_18

Jakobsdottir, J., Gorin, M. B., Conley, Y. P., Ferrell, R. E., & Weeks, D. E. (2009). Interpretation of genetic association studies: Markers with replicated highly significant odds ratios may be poor classifiers. *PLoS Genetics*, *5*(2). https://doi.org/10.1371/journal.pgen.1000337

Jamesdaniel, S., Rosati, R., Westrick, J., & Ruden, D. M. (2018). Chronic lead exposure induces cochlear oxidative stress and potentiates noise-induced hearing loss. *Toxicology Letters*, *292*, 175–180. https://doi.org/10.1016/j.toxlet.2018.05.004

Jing, W., Zongjie, H., Denggang, F., Na, H., Bin, Z., Aifen, Z., … Ring, B. Z. (2015). Mitochondrial mutations associated with aminoglycoside ototoxicity and hearing loss susceptibility identified by meta-analysis. *Journal of Medical Genetics*, *52*(2), 95–103. https://doi.org/10.1136/jmedgenet-2014-102753

Kaye, C., Bodurtha, J., Edick, M., Ginsburg, S., Keehn, A., Lloyd-Puryear, M., … Taylor, M. (n.d.). Regional models of genetic services in the United States the National Coordinating Center for the Regional Genetic Service Collaboratives Regional Support Service Model Workgroup and Advisory Committee. *GENETICS in MEDICINE*. https://doi.org/10.1038/s41436-019

Kesser, B. W., Hashisaki, G. T., Fletcher, K., Eppard, H., & Holt, J. R. (2007). An in vitro model system to study gene therapy in the human inner ear. *Gene Therapy*, *14*(15), 1121–1131. https://doi.org/10.1038/sj.gt.3302980

Konings, A., Van Laer, L., Michel, S., Pawelczyk, M., Carlsson, P. I., Bondeson, M. L., … Van Camp, G. (2009). Variations in HSP70 genes associated with noise-induced hearing loss in two independent populations. *European Journal of Human Genetics*, *17*(3), 329–335. https://doi.org/10.1038/ejhg.2008.172

Krishnamurti, S. (2009). Sensorineural hearing loss associated with occupational noise exposure: Effects of age-corrections. *International Journal of Environmental Research and Public Health*, *6*(3), 889–899. https://doi.org/10.3390/ijerph6030889

Kujawa, S. G., & Liberman, M. C. (2009). Adding insult to injury: Cochlear nerve degeneration after "temporary" noise-induced hearing loss. *Journal of Neuroscience*, *29*(45), 14077–14085. https://doi.org/10.1523/JNEUROSCI.2845-09.2009

Lanvers-Kaminsky, C., & Ciarimboli, G. (2017). Pharmacogenetics of drug-induced ototoxicity caused by aminoglycosides and cisplatin. *Pharmacogenomics*, *18*(18), 1683–1695. https://doi.org/10.2217/pgs-2017-0125

Lavinsky, J., Crow, A. L., Pan, C., Wang, J., Aaron, K. A., Ho, M. K., … Friedman, R. A. (2015). Genome-Wide Association Study Identifies Nox3 as a Critical Gene for Susceptibility to Noise-Induced Hearing Loss. *PLoS Genetics*, *11*(4). https://doi.org/10.1371/journal.pgen.1005094

Lavinsky, J., Ge, M., Crow, A. L., Pan, C., Wang, J., Salehi, P., … Friedman, R. A. (2016). The genetic architecture of noise-induced hearing loss: Evidence for a gene-by-environment interaction. *G3: Genes, Genomes, Genetics*, *6*(10), 3219–3228. https://doi.org/10.1534/g3.116.032516

Lavinsky, J., Ge, M., Crow, A. L., Pan, C., Wang, J., Salehi, P., … Friedman, R. A. (2016). The genetic architecture of noise-induced hearing loss: Evidence for a gene-by-environment interaction. *G3: Genes, Genomes, Genetics*, *6*(10), 3219–3228. https://doi.org/10.1534/g3.116.032516

Lei, S., Huang, L., Liu, Y., Xu, L., Wang, D., & Yang, L. (2017). Association between polymorphisms of heat-shock protein 70 genes and noise-induced hearing loss: A meta-analysis. *PLoS ONE*, *12*(11). https://doi.org/10.1371/journal.pone.0188539

Lewis, C. M., Whitwell, S. C. L., Forbes, A., Sanderson, J., Mathew, C. G., & Marteau, T. M. (2007). Estimating risks of common complex diseases across genetic and environmental factors: The example of Crohn disease. *Journal of Medical Genetics*, *44*(11), 689–694. https://doi.org/10.1136/jmg.2007.051672

Li, Y. H., Yang, Y., Yan, Y. T., Xu, L. W., Ma, H. Y., Shao, Y. X., … Yang, L. (2018). Analysis of serum microRNA expression in male workers with occupational noise-induced hearing loss. *Brazilian Journal of Medical and Biological Research*, *51*(3). https://doi.org/10.1590/1414-431X20176426

Lie, A., Skogstad, M., Johannessen, H. A., Tynes, T., Mehlum, I. S., Nordby, K. C., … Tambs, K. (2016). Occupational noise exposure and hearing: a systematic review. *International Archives of Occupational and Environmental Health*, *89*(3), 351–372. https://doi.org/10.1007/s00420-015-1083-5

Lin, C. Y., Wu, J. L., Shih, T. S., Tsai, P. J., Sun, Y. M., & Guo, Y. L. (2009). Glutathione S-transferase M1, T1, and P1 polymorphisms as susceptibility factors for noise-induced temporary threshold shift. *Hearing Research*, *257*(1–2), 8–15. https://doi.org/10.1016/j.heares.2009.07.008

Lin, F. R., Maas, P., Chien, W., Carey, J. P., Ferrucci, L., & Thorpe, R. (2012). Association of skin color, race/ethnicity, and hearing loss among adults in the USA. *JARO - Journal of the Association for Research in Otolaryngology*, *13*(1), 109–117. https://doi.org/10.1007/s10162-011-0298-8

Luo, H., Wu, H., Shen, H., Chen, H., Yang, T., Huang, Z., … Li, J. (2016). The European GWAS-identified risk SNP rs457717 within IQGAP2 is not associated with age-related hearing impairment in Han male Chinese population. *European Archives of Oto-Rhino-Laryngology*, *273*(7), 1677–1687. https://doi.org/10.1007/s00405-015-3711-9

Lutman, M. E., & Davis, A. C. (1994). The distribution of hearing threshold levels in the general population aged 18-30 years. *International Journal of Audiology*, *33*(6), 327–350. https://doi.org/10.3109/00206099409071891

Murillo-Cuesta, S., Rodríguez-de la Rosa, L., Contreras, J., Celaya, A. M., Camarero, G., Rivera, T., & Varela-Nieto, I. (2015). Transforming growth factor β1 inhibition protects from noise-induced hearing loss. *Frontiers in Aging Neuroscience*, *7*(FEB). https://doi.org/10.3389/fnagi.2015.00032

Nakanishi, N., Okamoto, M., Nakamura, K., Suzuki, K., & Tatara, K. (2000). Cigarette smoking and risk for hearing impairment: A longitudinal study in Japanese male office workers. *Journal of Occupational and Environmental Medicine*, *42*(11), 1045–1049. https://doi.org/10.1097/00043764-200011000-00001

Nomura, K., Nakao, M., & Morimoto, T. (2005). Effect of smoking on hearing loss: Quality assessment and meta-analysis. *Preventive Medicine*, *40*(2), 138–144. https://doi.org/10.1016/j.ypmed.2004.05.011

Nondahl, D. M., Cruickshanks, K. J., Dalton, D. S., Klein, B. E. K., Klein, R., Tweed, T. S., & Wiley, T. L. (2006). The use of hearing protection devices by older adults during recreational noise exposure. *Noise and Health*, *8*(33), 147–153. https://doi.org/10.4103/1463-1741.34702

Nondahl, D. M., Cruickshanks, K. J., Huang, G.-H., Klein, B. E. K., Klein, R., Tweed, T. S., & Zhan, W. (n.d.). *Generational Differences in the Reporting of Tinnitus*. https://doi.org/10.1097/AUD.0b013e31825069e8

Noorhassim, I., & Rampal, K. G. (1998). Multiplicative effect of smoking and age on hearing impairment. *American Journal of Otolaryngology - Head and Neck Medicine and Surgery*, *19*(4), 240–243. https://doi.org/10.1016/S0196-0709(98)90125-9

Obasi, C. N., Cruickshanks, K. J., Nondahl, D. M., Klein, B. E. K., Klein, R., Nieto, F. J., … R, C. (2013). Dysfunction and Oxidative Stress with Cognitive Impairment. The Epidemiology of Hearing Loss Study (EHLS). *Oxid Antioxid Med Sci*, *60*(8), 1147–1150. https://doi.org/10.231/JIM.0b013e318271fdb9.How

Oishi, N., & Schacht, J. (2011, June). Emerging treatments for noise-induced hearing loss. *Expert Opinion on Emerging Drugs*, Vol. 16, pp. 235–245. https://doi.org/10.1517/14728214.2011.552427

Omichi, R., Shibata, S. B., Morton, C. C., & Smith, R. J. H. (2019). Gene therapy for hearing loss. *Human Molecular Genetics*. https://doi.org/10.1093/hmg/ddz129

Palmer, K. T., Griffin, M. J., Syddall, H. E., & Coggon, D. (2004). Cigarette smoking, occupational exposure to noise, and self reported hearing difficulties. *Occupational and Environmental Medicine*, *61*(4), 340–344. https://doi.org/10.1136/oem.2003.009183

Park, J. A., & Suh, M. J. (2019). Hazardous alcohol consumption and the risk of hearing impairment in adults based on the Korean national health and nutrition survey: A retrospective study. *Journal of Audiology and Otology*, *23*(2), 63–68. https://doi.org/10.7874/jao.2018.00241

Pearson, J. D., Morrell, C. H., Gordon-Salant, S., Brant, L. J., Jeffrey Metter, E., Klein, L. L., & Fozard, J. L. (1995). Gender differences in a longitudinal study of age-associated hearing loss. *Journal of the Acoustical Society of America*, *97*(2), 1196–1205. https://doi.org/10.1121/1.412231

Plomp, R., & Mimpen, A. M. (1979). Speech-reception threshold for sentences as a function of age and noise level. *Journal of the Acoustical Society of America*, *66*(5), 1333–1342. https://doi.org/10.1121/1.383554

Pouryaghoub, G., Mehrdad, R., & Mohammadi, S. (2007). Interaction of smoking and occupational noise exposure on hearing loss: A cross-sectional study. *BMC Public Health*, *7*. https://doi.org/10.1186/1471-2458-7-137

Rosler, G. (1994). Progression of Hearing Loss Caused by Occupational Noise. *Scandinavian Audiology*, *23*(1), 13–37. https://doi.org/10.3109/01050399409047483

Schacht, J., Talaska, A. E., & Rybak, L. P. (2012). Cisplatin and Aminoglycoside Antibiotics: Hearing Loss and Its Prevention. *Anatomical Record*, *295*(11), 1837–1850. https://doi.org/10.1002/ar.22578

Siegelaub, A. B., Friedman, G. D., Adour, K., & Seltzer, C. C. (1974). Hearing Loss in Adults: Relation to Age, Sex, Exposure to Loud Noise, and Cigarette Smoking. *Archives of Environmental Health*, *29*(2), 107–109. https://doi.org/10.1080/00039896.1974.10666542

Sliwinska-Kowalska, M., & Pawelczyk, M. (2013, January). Contribution of genetic factors to noise-induced hearing loss: A human studies review. *Mutation Research - Reviews in Mutation Research*, Vol. 752, pp. 61–65. https://doi.org/10.1016/j.mrrev.2012.11.001

Sun, Y., Xiang, J., Liu, Y., Chen, S., Yu, J., Peng, J., … Peng, Z. (2019). Increased diagnostic yield by reanalysis of data from a hearing loss gene panel. *BMC Medical Genomics*, *12*(1). https://doi.org/10.1186/s12920-019-0531-6

Suter, A. H. (2017). Occupational Hearing Loss from Non-Gaussian Noise. *Seminars in Hearing*, *38*(3), 225–262. https://doi.org/10.1055/s-0037-1603726

Tayoun, A. N. A., Al Turki, S. H., Oza, A. M., Bowser, M. J., Hernandez, A. L., Funke, B. H., … Amr, S. S. (2016, June 1). Improving hearing loss gene testing: A systematic review of gene evidence toward more efficient next-generation sequencing-based diagnostic testing and interpretation. *Genetics in Medicine*, Vol. 18, pp. 545–553. https://doi.org/10.1038/gim.2015.141

Teitz, T., Fang, J., Goktug, A. N., Bonga, J. D., Diao, S., Hazlitt, R. A., … Zuo, J. (2018). CDK2 inhibitors as candidate therapeutics for cisplatin- and noise-induced hearing loss. *Journal of Experimental Medicine*, *215*(4), 1187–1203. https://doi.org/10.1084/jem.20172246

Thonnard, J., Philippe, M., Cao, M. Y., Deggouj, N., Osselaer, J. C., Gersdorff, M., & Tomasi, J. P. (1996). HLA class II-associated genetic susceptibility in idiopathic progressive sensorineural hearing loss. *Annals of Otology, Rhinology and Laryngology*, *105*(8), 628–633. https://doi.org/10.1177/000348949610500808

Uchida, Y., Nakashima, T., Ando, F., Niino, N., & Shimokata, H. (2005). Is there a relevant effect of noise and smoking on hearing? A population-based aging study. *International Journal of Audiology*, *44*(2), 86–91. https://doi.org/10.1080/14992020500031256

Vuckovic, D., Mezzavilla, M., Cocca, M., Morgan, A., Brumat, M., Catamo, E., … Girotto, G. (2018). Whole-genome sequencing reveals new insights into age-related hearing loss: cumulative effects, pleiotropy and the role of selection. *European Journal of Human Genetics*, *26*(8), 1167–1179. https://doi.org/10.1038/s41431-018-0126-2

Wang, J., Li, J., Peng, K., Fu, Z. Y., Tang, J., Yang, M. J., & Chen, Q. C. (2017). Associação do polimorfismo C47T no gene da superóxido dismutase 2 com perda auditiva induzida pelo ruído: metanálise. *Brazilian Journal of Otorhinolaryngology*, *83*(1), 80–87. https://doi.org/10.1016/j.bjorl.2016.01.008

Wells, H. R. R., Freidin, M. B., Abidin, F. N. Z., Payton, A., Dawes, P., Munro, K. J., … Wiliams, F. M. (2019). Genome-wide association study identifies 44 independent genomic loci for self-reported adult hearing difficulty in the UK Biobank cohort. *BioRxiv*, 549071. https://doi.org/10.1101/549071

Wu, Y., Ni, J., Qi, M., Cao, C., Shao, Y., Xu, L., … Yang, L. (2017). Associations of genetic variation in CASP3 gene with noise-induced hearing loss in a Chinese population: A case-control study. *Environmental Health: A Global Access Science Source*, *16*(1). https://doi.org/10.1186/s12940-017-0280-y

Xia, W., Hu, J., Ma, J., Huang, J., Wang, X., Jiang, N., … Ma, D. (2019). Novel TRRAP mutation causes autosomal dominant non-syndromic hearing loss. *Clinical Genetics*. https://doi.org/10.1111/cge.13590

Xu, X., Yang, Q., Jiao, J., He, L., Yu, S., Wang, J., … Zhang, H. (2016). Genetic variation in POu4F3 and GRHL2 associated with noise-induced hearing loss in Chinese population: A case-control study. *International Journal of Environmental Research and Public Health*, *13*(6). https://doi.org/10.3390/ijerph13060561

Yamasoba, T., Lin, F. R., Someya, S., Kashio, A., Sakamoto, T., & Kondo, K. (2013, September). Current concepts in age-related hearing loss: Epidemiology and mechanistic pathways. *Hearing Research*, Vol. 303, pp. 30–38. https://doi.org/10.1016/j.heares.2013.01.021

Yan, D., Zhu, Y., Walsh, T., Xie, D., Yuan, H., Sirmaci, A., … Liu, X. Z. (2013). Mutation of the ATP-gated P2X2 receptor leads to progressive hearing loss and increased susceptibility to noise. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(6), 2228–2233. https://doi.org/10.1073/pnas.1222285110

Yuan, H., Wang, X., Hill, K., Chen, J., Lemasters, J., Yang, S. M., & Sha, S. H. (2015). Autophagy attenuates noise-induced hearing loss by reducing oxidative stress. *Antioxidants and Redox Signaling*, *22*(15), 1308–1324. https://doi.org/10.1089/ars.2014.6004

Zhang, W., Kim, S. M., Wang, W., Cai, C., Feng, Y., Kong, W., & Lin, X. (2018, June 26). Cochlear Gene Therapy for Sensorineural Hearing Loss: Current Status and Major Remaining Hurdles for Translational Success. *Frontiers in Molecular Neuroscience*, Vol. 11. https://doi.org/10.3389/fnmol.2018.00221

Zhang, X., Liu, Y., Zhang, L., Yang, Z., Shao, Y., Jiang, C., … Zhu, Y. (2014). Genetic variations in protocadherin 15 and their interactions with noise exposure associated with noise-induced hearing loss in Chinese population. *Environmental Research*, *135*, 247–252. https://doi.org/10.1016/j.envres.2014.09.021

Zhang, X., Liu, Y., Zhang, L., Yang, Z., Yang, L., Wang, X., … Zhu, Y. (2015). Associations of genetic variations in EYA4, GRHL2 and DFNA5 with noise-induced hearing loss in Chinese population: A case- control study. *Environmental Health: A Global Access Science Source*, *14*(1). https://doi.org/10.1186/s12940-015-0063-2

Zhang, X., Ni, Y., Liu, Y., Zhang, L., Zhang, M., Fang, X., … Zhu, Y. (2019). Screening of noise-induced hearing loss (NIHL)-associated SNPs and the assessment of its genetic

susceptibility. *Environmental Health: A Global Access Science Source*, *18*(1). https://doi.org/10.1186/s12940-019-0471-9

APPENDIX C. LIST OF HUMAN OR MOUSE GENES WITH POLYMORPHISMS ASSOCIATED WITH NON-SYNDROMIC HEARING LOSS

	A730017C20Rik	CLPP	GJB6/DFNB1	LOXHD1	**OTOGL**	TBC1D24
	Aak1	Clrn1	GMST1	LRTOMT	**P2RX2**	TECTA
	ABCC1	CLRN1	Gpr152	MARVELD2	**PAX3**	TIMM8A
	Acsl4	COCH	Gpr50	Med28	PAX3	TMC1
	ACTG1	COL11A2	GPR98	MIR96	**PCDH15**	Tmem30b
	Acvr2a	*Col9a2	**GPSM2**	**MITF**	Phf6	TMIE
	Adgrb1	Cyb5r2	GPX	Mpdz	POU3F4	TMPRSS3
	Adgrv1	DFNA5	**GRHL2**	MSRB3	POU4F3	Tmtc4
	AGL	DFNA5	GRXCR1	mtDNA 12S RNA	Ppm1a	Tox
	Ahsg	DFNB59	GSR	MTRNR1	PRPS1	Tprn
	Ankrd11	DIABLO	GSTP1	MTTS1	PU4F3	TPRN
	Ap3m2	DIAPH1	GSTT1	Myh1	**RDX**	TRAAP
	Ap3s1	Dnase1	HARS	MYH14	RDX	Tram2
	Atp2b1	Duoxa2	HARS2	MYH9	Sema3f	TRIOBP
	ATP6V1B1	EDN3	HGF	**MYO15A**	SERPINB6	TSPEAR
	B020004J07Rik	EDNRB	HLA class II genes	*MYO1A	SIX1	Ube2b
	Baiap2l2	Elmod1	HSD17B4	MYO3A	SLC26A4	Ube2g1
	BCS1L	Emb	HSP70	MYO6	SLC26A4	Ush1c
	BSND	Eps8l1	Ikzf5	**MYO7A**	Slc4a10 [b]	**USH1C**
	CACNA1D	ESPN	Il1r2	Nedd4l	Slc5a5	USH1G
	CASP3	**ESRRB**	Ildr1	Nfatc3	SLITRK6	USH2A
	CAT	Ewsr1	ILDR1	Nin	SMPX	USH38
	CCDC50	EYA1	KARS	Nisch	SMPX	Vti1a
	Ccdc88c	**EYA4**	KCNE1	NOX3	SNAI2	Wdtc1 [a]
	Ccdc92	FOXO3	KCNM1A	Nptn	SOD	WFS1
	CDH23	Gata2	KCNQ1	Ocm	SOD2	**WHRN**
	CEACAM16	Gga1	**KCNQ4**	Odf3l2	SOX10	Zcchc14
	Cib2	Gipc3	Klc2	Otoa	Spns2 [b]	Zfp719
	CIB2	GIPC3	Klhl18	OTOA	**Srrm4**	
	CISD2	**GJB2**	LARS2	**OTOF**	STRC	
	CLDN14	GJB3	LHFPL5	OTOG	SYNE4	

Bold font indicates that more than one reference found for the gene. Asterisk indicates conflicting evidence. Genes in all capitals were identified in human subjects. Genes in sentence case have only been identified in mice.

LIST OF ACRONYMS & ABBREVIATIONS

711 HPW – 711th Human Performance Wing
AFMS – Air Force Medical Service
BOSS – Beaver Dam Offspring Study
CDSS – Clinical Decision Support System
CI – confidence interval
DNA – deoxyribonucleic acid
dB HL – decibels hearing level
DOEHRS-IH – Defense Occupational and Environmental Health Readiness System – Industrial Hygiene
EHLS – Epidemiology of Hearing Loss
EHR – Electronic health record
GWAS – genome-wide association studies
HL – hearing level
HPW – Human Performance Wing
IEHRP – Individual Exposure Health Risk Profile
ILER – Individual Longitudinal Exposure Record
kHz - kilohertz
LD – Linkage disequilibrium
MHS – Medical Health System
NDAA – National Defense Appropriations Act
NGS – Next generation sequencing
NHANES – National Health and Nutrition Examination Survey
NIHL – noise-induced hearing loss
OHR&RP – Operation Health Risk and Readiness Profile
OR – odds ratio
PERS-STAT – Air Force Personnel Center Personnel Statistics Analytics Site
REGENT – Risk Estimation for Genetic and Environmental Traits
RNA – ribonucleic acid
RR – risk ratio or relative risk
SNHL – sensorineural hearing loss
SNP – single nucleotide polymorphism
TEH – Total Exposure Health
TRR - Total relative risk
UKBB – United Kingdom (UK) Biobank
WRCM – word recognition in quiet and in competing message