THE UNIVERSITY OF QUEENSLAND

AUSTRALIA

Diagnosing conductive dysfunction in infants using wideband acoustic immittance: Development and validation of predictive models

Joshua Myers

Master of Audiology Studies

A thesis submitted for the degree of Doctor of Philosophy at

The University of Queensland in 2020

School of Health and Rehabilitation Sciences

Hearing Research Unit for Children

Abstract

Background. Infants with early onset of otitis media have greater risk of recurrent and chronic infections that can affect language and development. Diagnostic tools able to quickly and accurately identify middle ear pathology early in infancy could help to facilitate timely intervention for these children. Wideband acoustic immittance (WAI) is an emerging technology for assessing middle ear function with significant advantages over established clinical tests such as tympanometry. WAI does not require pressurization of the ear canal, and is a high-resolution test, measuring middle ear function over a wider frequency range than is possible with tympanometry. Preliminary studies in infants have shown promising results, but further research assessing the diagnostic performance of WAI is needed. Also, the large amount of data generated by WAI can make results difficult to interpret. Research into suitable quantitative techniques to analyse results is still in its infancy. Prediction models are an attractive method for analysis of multivariate data as they can provide individualized probabilities that an infant has middle ear dysfunction. A clinically useful prediction model must be able to accurately discriminate between normal ears and those with middle ear dysfunction, and be well calibrated (i.e., give accurate predictions). However, the number of variables generated by WAI can cause issues with overfitting when developing multivariate models. An overfitted model will accurately describe the data it was developed on, but is likely to perform poorly when applied to new samples. Some form of data reduction or penalization is therefore necessary when modelling WAI. Another issue with developing WAI models is that there are substantial maturational effects on WAI through infancy that need to be controlled for. This can be achieved by either developing models for specific age groups (e.g., a model specifically for neonates), or by controlling for the effect of age by including interactions between age and WAI variables in a model, or by only using developmentally stable regions as predictor variables.

Objective. The aim of this work was to investigate the diagnostic performance of WAI in infants by developing predictive models. Data reduction strategies such as selecting predictors based on prior research and principal component analysis were used to increase the likelihood that models would generalize to new samples. The effect of age was initially accounted for by developing age-specific models for neonates, 6-month infants, and 12-month infants (Chapters 2, 3 and 4, respectively). Longitudinal developmental effects on WAI through infancy were then investigated in Chapter 5, and this knowledge was used to develop a model controlling for the effect of age through infancy (6 to 18 months) (Chapter 6, Study 1). The neonate model was assessed for generalizability by applying the model to results from a new sample of infants (Chapter 6, Study 2)

Methods. Tympanometry, distortion product otoacoustic emissions (DPOAEs) and WAI were measured in 753 neonates, and longitudinally in 357 infants who attended follow up appointments at around 6, 12 and 18 months of age. High-frequency (1000-Hz) tympanometry was measured in neonates and at 6 months, and 226-Hz tympanometry at 12 and 18 months. Tympanometry and DPOAEs were used to assess middle ear function of infants at each test session. Predictive models were developed for specific age groups through infancy: neonates, 6-, and 12-months (Chapters 2, 3 and 4, respectively), and longitudinal developmental effects on WAI through infancy were investigated (Chapter 5). A model controlling for the effect of age was developed for use in infants aged 6- to 18-months (Chapter 6, Study 1). The neonate model was applied to a new sample of 124 neonates to assess how well it generalized to new infants (Chapter 6, Study 2). Performance of the models was assessed with the *c*-index and calibration curves. Models were internally validated using bootstrap resampling to correct for bias (overfitting) and/or data from the opposite ears of subjects.

Results. The bias-corrected *c*-index results of the neonate, 6-month, 12-month, and 6- to 18-month models were 0.85, 0.87, 0.91 and 0.87, respectively. The *c*-index of the neonate model when applied to the new sample was 0.84. Calibration was satisfactory for all models.

Discussion. The developed models accurately identified middle ear dysfunction in infants. The models were carefully fitted and internally validated, to increase the likelihood that they will generalize to new samples. The neonate model did effectively generalize to a new sample, which indicates that the strategies employed to minimize overfitting were effective. There were large developmental effects on WAI measurements, and this knowledge was used to develop a model that controlled for maturational effects through infancy. The models have potential applications in both screening and diagnostic settings. In a screening context, predictions could be used to set a referral threshold sensitive to the costs associated with true, and false positive referrals, that is intuitive and easy to apply. In a diagnostic setting, predicted probabilities could be used in conjunction with graphical depictions of WAI for individualized diagnoses of conductive dysfunction. Further research validating, updating, and assessing the clinical impact of the models is warranted.

Declaration by Author

This thesis is composed of my original work, and contains no material previously published or written by another person except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my thesis.

I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have carried out since the commencement of my research higher degree candidature and does not include a substantial part of work that has been submitted to qualify for the award of any other degree or diploma in any university or other tertiary institution. I have clearly stated which parts of my thesis, if any, have been submitted to qualify for another award.

I acknowledge that an electronic copy of my thesis must be lodged with the University Library and, subject to the policy and procedures of The University of Queensland, the thesis be made available for research and study in accordance with the Copyright Act 1968 unless a period of embargo has been approved by the Dean of the Graduate School.

I acknowledge that copyright of all material contained in my thesis resides with the copyright holder(s) of that material. Where appropriate I have obtained copyright permission from the copyright holder to reproduce material in this thesis and have sought permission from co-authors for any jointly authored works included in the thesis.

Publications Included in this Thesis

Myers, **J**., Kei, J., Aithal, S., Aithal, V., Driscoll, C., Khan, A., Manuel, A., Joseph, A., Malicka, A. N. (2018a). Development of a diagnostic prediction model for conductive conditions in neonates using wideband acoustic immittance. *Ear and Hearing, 39*(6), 1116-1135.

Myers, **J**., Kei, J., Aithal, S., Aithal, V., Driscoll, C., Khan, A., Manuel, A., Joseph, A., Malicka, A. N. (2018b). Diagnosing middle ear pathology in 6- to 9-month-old infants using wideband absorbance: A risk prediction model. *Journal of Speech Language and Hearing Research, 61*(9), 2386-2404.

Myers, **J**., Kei, J., Aithal, S., Aithal, V., Driscoll, C., Khan, A., Manuel, A., Joseph, A., Malicka, A. N. (2019a). Diagnosing conductive dysfunction in infants using wideband acoustic immittance: Validation and development of predictive models. *Journal of Speech Language and Hearing Research, 62*(9), 3607-3619.

Myers, **J**., Kei, J., Aithal, S., Aithal, V., Driscoll, C., Khan, A., Manuel, A., Joseph, A., Malicka, A. N. (2019b). Diagnosing middle ear dysfunction in 10- to 16-month old infants using wideband absorbance: An ordinal prediction model. *Journal of Speech Language and Hearing Research, 62*(8), 2906-2917.

Myers, **J**., Kei, J., Aithal, S., Aithal, V., Driscoll, C., Khan, A., Manuel, A., Joseph, A., Malicka, A. N. (2019c). Longitudinal development of wideband absorbance and admittance through infancy. *Journal of Speech Language and Hearing Research, 62*(7), 2535-2552.

Submitted Manuscripts Included in this Thesis

No manuscripts submitted for publication.

Other Publications During Candidature

Peer-reviewed articles

Aithal, S., Kei, J., Aithal, V., Manuel, A., **Myers**, **J**., Driscoll, C., & Khan, A. (2017). Normative study of wideband acoustic immittance measures in newborn infants. *Journal of Speech Language and Hearing Research*, *60*(5), 1417-1426.

Conference presentations

Myers, **J**., Kei, J., Aithal, S., Aithal, V., Driscoll, C., Khan, A., Manuel, A., Joseph, A., Malicka, A. N. (2015). Innovative diagnosis of conductive conditions in neonates: Preliminary results. Paper pre-

sented at University of Queensland School of Health and Rehabilitation Sciences Postgraduate Research Conference, Brisbane, Australia.

Myers, **J**., Kei, J., Aithal, S., Aithal, V., Driscoll, C., Khan, A., Manuel, A., Joseph, A., Malicka, A. N. (2016a). Identification of conductive conditions in neonates using wideband absorbance. Paper presented at Audiology Australia National Conference, Melbourne, Australia.

Myers, **J**., Kei, J., Aithal, S., Aithal, V., Driscoll, C., Khan, A., Manuel, A., Joseph, A., Malicka, A. N. (2016b). Identifying conductive conditions in neonates with wideband absorbance. Paper presented at Townsville Hospital and Health Service Research Week Symposium, Townsville, Australia.

Myers, **J**., Kei, J., Aithal, S., Aithal, V., Driscoll, C., Khan, A., Manuel, A., Joseph, A., Malicka, A. N. (2016c). Identifying conductive conditions in neonates using wideband acoustic immittance. *The Journal of the Acoustical Society of America, 140*(4), 3264-3264. Poster presented at 5th Joint Meeting of the Acoustical Society of America and Acoustical Society of Japan, Honolulu, Hawaii.

Myers, **J**., Kei, J., Aithal, S., Aithal, V., Driscoll, C., Khan, A., Manuel, A., Joseph, A., Malicka, A. N. (2017a). A prediction model for diagnosis of middle ear pathology in 6- to 9-month-old infants using wideband absorbance. Paper presented at Townsville Health Research Showcase, Townsville, Australia.

Myers, **J**., Kei, J., Aithal, S., Aithal, V., Driscoll, C., Khan, A., Manuel, A., Joseph, A., Malicka, A. N. (2017b). Identification of outer/middle ear pathology in neonates using wideband acoustic immittance. Poster presented at Townsville Health Research Showcase, Townsville, Australia.

Myers, **J**., Kei, J., Aithal, S., Aithal, V., Driscoll, C., Khan, A., Manuel, A., Joseph, A., Malicka, A. N. (2018c). Diagnosing middle ear pathology in 6- to 9-month-old infants using wideband absorbance. Poster presented at Townsville Health Research Showcase, Townsville, Australia.

Myers, **J**., Kei, J., Aithal, S., Aithal, V., Driscoll, C., Khan, A., Manuel, A., Joseph, A., Malicka, A. N. (2018d). External validation of a wideband acoustic immittance model for identifying conductive conditions in neonates. Paper presented at Audiology Australia National Conference, Sydney, Australia.

Myers, **J**., Kei, J., Aithal, S., Aithal, V., Driscoll, C., Khan, A., Manuel, A., Joseph, A., Malicka, A. N. (2018e). Predicting middle ear dysfunction in 6- to 9-month-old infants with wideband absorbance. Paper presented at Audiology Australia National Conference, Sydney, Australia.

Myers, **J**., Kei, J., Aithal, S., Aithal, V., Driscoll, C., Khan, A., Manuel, A., Joseph, A., Malicka, A. N. (2018f). Validating a wideband acoustic immittance diagnostic model for conductive conditions in neonates. Paper presented at Townsville Health Research Showcase, Townsville, Australia.

Contributions by Others to the Thesis

This work would not have been possible without the contributions of many others. Joseph Kei conceived the idea for the project, and along with Carlie Driscoll, Asad Khan, Venkatesh Aithal and Sreedevi Aithal obtained a grant from the National Health and Medical Research Council that funded the research (APP1046477). Sreedevi Aithal, Venkatesh Aithal and Joseph Kei helped to obtain ethical approval for the study from the Townsville Hospital and University of Queensland, and coordinated an agreement between the Townsville Hospital and the University of Queensland that made the research possible. Alehandrea Manuel and Anjali Joseph assisted with recruiting subjects, data collection and data entry, and Karen Nielsen from the Audiology Department at the Townsville Hospital assisted with administrative duties such as booking appointments and posting letters.

Statement of Parts of the Thesis Submitted to Qualify for the Award of Another Degree

No works submitted towards another degree have been included in this thesis.

Research Involving Human or Animal Subjects

This project was approved by the Townsville Health Service District Institutional Ethics Committee (reference number: HREC/09/QTHS/30) and the University of Queensland Behavioural and Social Science Ethical Review Committee (reference number: 2010000842). Copy of ethical approval letters are included in Appendices I and II.

Acknowledgements

I would like to thank Joseph Kei for giving me the opportunity be involved in such an incredible project. It has been a very rewarding experience, and I am grateful to you for allowing me the freedom to follow my interests. I am grateful to Alicja N. Malicka for introducing me to research, and for her mentorship and support throughout this project. I also thank the Venkatesh Aithal and Sreedevi Aithal, from Audiology Department at the Townsville Hospital for supporting me in undertaking this project whilst working at the hospital. I acknowledge that this work directly builds on their pioneering research at the Townsville Hospital using innovative technologies for diagnosing middle ear conditions in infants.

I thank the Surgical Services and Health and Wellbeing Service Groups of the Townsville Hospital and Health Service for supporting the study, the Healthy Hearing Team at Townsville Hospital for help with recruiting participants, Karen Nielsen from the Audiology Department at Townsville Hospital for administrative support, and Interacoustics A/S (Denmark) for loaning equipment for the study. I also thank Douglas H. Keefe for helpful discussions about the finer points of wideband acoustic immittance when I was doing the data analysis for Chapter 5. I am especially grateful to all the parents who participated in the study, it was uplifting to see so many people willing to participate in scientific research.

Financial Support

This research was funded through a grant from the Australian National Health and Medical Research Council (NHMRC, APP1046477), a scholarship from the Nursing and Allied Health Scholarship and Support Scheme (NAHSSS) funded by the Commonwealth Department of Health, and an Australian Government Research Training Program Scholarship. The views expressed in this thesis do not necessarily represent those of the NHMRC; the NAHSSS or its Administrator, Services for Australian Rural and Remote Allied Health; or the Commonwealth Department of Health.

Keywords

Infants, middle ear, neonates, otitis media, wideband absorbance, wideband acoustic immittance, wideband reflectance

Australian and New Zealand Standard Research Classifications (ANZSRC)

ANZRC code: 920107, Hearing, Vision, Speech and Their Disorders, 40%
ANZRC code: 920203 Diagnostic Methods, 40%
ANZSRC code: 111403, Paediatrics, 20%

Fields of Research (FoR) Classification

FoR code: 9201, Clinical Health (Organs, Diseases and Abnormal Conditions), 40%
FoR code: 9202, Health and Support Services, 40%
FoR code: 1114, Paediatrics and Reproductive Medicine, 20%

Dedication

For my beautiful wife Stacey, thank you for your love, support and patience, and my boys, Aaron and Oliver, I enrolled in this research program just three days after you were born — and now you are in school! I love you all so much and dedicate this work to you.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

A	energy absorbance
ABR	auditory brainstem response
AIC	Akaike's information criterion
ASR	acoustic stapedial reflex
AUC	area under the receiver operating characteristic curve
B	susceptance
DPOAEs	distortion product otoacoustic emissions
EOAEs	evoked otoacoustic emissions
EM means	estimated marginal means
df	degrees of freedom
G	conductance
γ	shrinkage coefficient
HFT	high-frequency (1000 Hz) tympanometry
IQR	interquartile range
LFT	low-frequency (226 Hz) tympanometry
PCA	principal component analysis
φ_Y	acoustic admittance phase angle
PR	pressure reflectance
R	energy reflectance
SNR	signal-to-noise ratio
TEOAEs	transient evoked otoacoustic emissions
TM	tympanic membrane
WAI	wideband acoustic immittance
Y	acoustic admittance
$\|Y\|$	acoustic admittance magnitude
Z	acoustic impedance

Chapter 1. Introduction

1.1 Overview

Otitis media, an inflammation of the middle ear, is one of the most common conditions of infancy (Thomas & Brook, 2014). Early intervention is vital, as early onset of otitis media is associated with recurrent and chronic conditions that can affect language and development (Corbeel, 2007; Damoiseaux, Rovers, Van Balen, Hoes, & de Melker, 2006; Karma, Perälä, & Kuusela, 1989b). Early intervention requires timely diagnosis, but identifying this often asymptomatic condition in infants remains a challenge. Tests that have been developed and validated for use in older children have significant limitations in infants (Aithal, Kei, Driscoll, & Khan, 2013). Pneumatic otoscopy is difficult and unreliable in infants with accuracy varying widely between otoscopists (Chianese et al., 2007; Smith et al., 2006). Tympanometry using a low-frequency (226 Hz) probe tone (LFT) is inaccurate in young infants, and some large-scale studies using stringent reference standards such as otomicroscopy and myringotomy have found poor sensitivity in older infants (Hoffmann et al., 2013; Palmu & Syrjänen, 2005). High-frequency tympanometry (HFT) is more accurate than LFT in young infants but has poor sensitivity in neonates (Margolis, Bass-Ringdahl, Hanks, Holte, & Zapala, 2003; Swanepoel et al., 2007). Other available clinical tests such as acoustic stapedial reflexes (ASR) and auditory brain-stem response (ABR) also have significant limitations for diagnosing otitis media in infants. ASR requires the subject to remain completely still which is rarely possible when testing infants, and ABR testing is very time consuming and requires significant expertise. There remains the need for a quick and accurate diagnostic tool for assessing middle ear function in infants.

Wideband acoustic immittance (WAI) is an innovative, high-resolution test of middle ear function that is quick and easy to use. Initial studies investigating the accuracy of WAI in infants have shown promising results, but large-scale studies have only been undertaken in neonates, using evoked otoa-coustic emissions (EOAEs) as the reference standard (Hunter, Feeney, Miller, Jeng, & Bohning, 2010; Keefe, Gorga, Neely, Zhao, & Vohr, 2003a; Keefe, Zhao, Neely, Gorga, & Vohr, 2003b; Sanford et al., 2009). Used in isolation as a reference test, EOAEs may not accurately reflect the status of the conduc-tive pathway as they are a test of inner, rather than middle ear function. Passing an EOAE test does not necessarily rule out conductive problems, as they have been recorded in ears of neonates, infants and children with known middle ear dysfunction (Aithal, Kei, Driscoll, Khan, & Swanston, 2015; Amedee, 1995; Doyle, Burggraaff, Fujikawa, Kim, & Macarthur, 1997; Driscoll, Kei, & McPherson, 2001; Margolis et al., 2003). Further research is needed in a large sample of neonates using a more stringent reference standard to determine the diagnostic performance of WAI. There is a dearth of evidence

about the diagnostic performance of WAI in infants outside of the neonatal period. Although preliminary studies in infants have reported strong performance (Ellison et al., 2012; Prieve, Vander Werff, Preston, & Georgantas, 2013b), more research is needed in this age group, as WAI could be a valuable tool for identification of middle ear pathology in infancy.

1.2 Otitis media in infants

1.2.1 Definitions

Lack of consensus regarding definitions and diagnostic markers is a major issue in otitis media research. A study that surveyed 165 physicians found 147 different sets of diagnostic criteria for acute otitis media (Hayden, 1981). Lack of standardization has caused difficulties in accurately estimating the incidence and prevalence of otitis media in Australia (Kong & Coates, 2009). Making comparisons between studies is often difficult due to differences in definitions, diagnostic methods and diagnostic criteria (Brennan-Jones et al., 2014; Kværner, Nafstad, Hagen, Mair, & Jaakkola, 1997). For example, in various studies, recurrent otitis media is defined as at least three episodes of otitis media in six or twelve months, at least three episodes in two to three years, or at least six in five years (Harsten, Prellner, Heldrup, Kalm, & Kornfält, 1989).

Acute otitis media is an acute infection of the middle ear accompanied by middle ear effusion. Acute otitis media can occur with fever, ear pain and irritability although it is often asymptomatic in infants (Casselbrant & Mandel, 2014). Recurrent acute otitis media refers to either three episodes of acute otitis media in six months or four in twelve months (Qureishi, Lee, Belfield, Birchall, & Daniel, 2014). Otitis media with effusion is a condition where middle ear effusion is present without acute infection (Rovers, 2008). Chronic otitis media with effusion refers to otitis media with effusion that persists longer than three months. Chronic suppurative otitis media is a discharging perforation of the eardrum that persists for longer than six weeks (Mahadevan et al., 2012). Recurrent otitis media is at least three episodes of otitis media in six months or four in twelve months with the type of otitis media unspecified. Chronic otitis media is otitis media that persists for greater than three months with the type unspecified.

In this thesis “otitis media” is used as a general term encompassing both acute otitis media and otitis media with effusion, and “conductive condition”, refers generally to conditions affecting the outer/middle ear. “Neonates” refers to children under 4 weeks old and “infants” to children under 2 years old. “Young infants” denotes children under 6 months old and “older children” refers to children over 2 years old.

1.2.2 Pathology of otitis media in infants

Otitis media is a multifactorial disease resulting from the interaction between the Eustachian tube, infection, the immune system and inflammatory responses (Allen, Manichaikul, & Sale, 2014; Lee, Kim, & Nguyen, 2013). Maturational factors play an important role in the pathogenesis of otitis media in infants. The ear is not fully mature at birth with development continuing until around 9 years of age. Significant development occurs over the first year of life with most rapid change in the first few months (Calandruccio, Fitzgerald, & Prieve, 2006; Kei, Sanford, Prieve, & Hunter, 2013). The bony portion surrounding the inner two-thirds of the ear canal is not fully formed until approximately 12 months of age which causes the ear canal of young infants to be highly compliant as it is completely surrounded by soft tissue (Baldwin, 2006; Wilson, 2012). Ear canal diameter and length increase, the size of the middle ear increases, ossicular bone density changes, and orientation of the tympanic membrane changes as the ear matures (Baldwin, 2006; Calandruccio et al., 2006; Sanford & Feeney, 2008). The infant Eustachian tube functions less efficiently than that of adults. It is shorter, wider and more flexible. It closes more slowly than that of adults and is almost horizontal, with an inclination of approximately 10 degrees. The Eustachian tube continues to develop over the first seven years of life, becoming longer and increasing in inclination to around 45 degrees (Casselbrant & Mandel, 2014; Kei et al., 2013).

The Eustachian tube regulates pressure, protects and clears secretions from the middle ear (Corbeel, 2007). It equilibrates the middle ear to atmospheric pressure by intermittently opening during movements such as swallowing, chewing and yawning (Casselbrant & Mandel, 2014). The immature, inefficient infant Eustachian tube leaves infants susceptible to Eustachian tube dysfunction, an inflammatory response that obstructs the Eustachian tube causing negative pressure to develop in the middle ear (Corbeel, 2007). Eustachian tube dysfunction often precedes otitis media. The negative middle ear pressure results in effusion and aspiration of pathogens into the middle ear. These secretions are then unable to be discharged because of the impaired function of the Eustachian tube (Cunningham, Guardiani, Kim, & Brook, 2012; Gould & Matz, 2010).

Viral and bacterial infection play a central role in the pathology of otitis media. Upper respiratory tract infection often precedes otitis media as it leaves the middle ear vulnerable to inflammation and infection (Pichichero, 2013). Otitis media has a cyclical peak coinciding with peak upper respiratory tract infection season (Gould & Matz, 2010). Chonmaitree et al. (2008) found that over 60% of upper respiratory tract infections in 6-month- to 3-year-old children were complicated by otitis media. Upper respiratory tract infection causes Eustachian tube dysfunction and creates an environment susceptible to

bacterial infection (Gould & Matz, 2010; Rovers, 2008). The most common viruses associated with otitis media are respiratory syncytial viruses, parainfluenza, influenza, enteroviruses and adenoviruses (Corbeel, 2007; Gould & Matz, 2010).

Bacterial colonization of the nasopharynx is the most significant factor leading to otitis media (Pelton & Leibovitz, 2009). Bacteria can be found in 50% to 90% of middle ear effusion samples (Harmes et al., 2013; Taylor et al., 2012). Infants are especially susceptible to colonization due to immaturity of the Eustachian tube and immune system (Pelton & Leibovitz, 2009; Pukander, Luotonen, Sipilau, Timonen, & Karma, 1982). Otitis media is caused when bacteria from the nasopharynx invade the middle ear (Lee et al., 2013). Traditionally, the dominant pathogen has been Streptococcus pneumoniae followed by nontypeable Haemophilus influenzae, and Moraxella catarrhalis. However, the bacterial landscape has changed since the introduction of the pneumococcal conjugate vaccine circa 2000 resulting in nontypeable Haemophilus influenzae becoming the dominant pathogen and an increase in Moraxella catarrhalis, Staphylococcus aureus and strains of Streptococcus pneumoniae not covered by the pneumococcal conjugate vaccine. The 13-valent pneumococcal conjugate vaccine introduced in 2010 may further change the bacterial landscape of otitis media (Harmes et al., 2013; Thomas & Brook, 2014).

Amniotic fluid plays an important role in the development of neonatal otitis media. Many neonates are born with a conductive condition due to vernix in the outer ear and residual mesenchyme and amniotic fluid in the middle ear. Balkany, Berman, Simmons, and Jafek (1978) found that the ear canal of neonates younger than 24 hours old were at least partially occluded by vernix caseosa, a waxy substance that coats the skin of newborns. Cavanaugh (1987) reported that vernix obscured view of the tympanic membrane in 56% of 1-day-old neonates which reduced to 24% by day 2 and 19% by day 3. Doyle et al. (1997) found that vernix occluded 13% of ears in neonates aged 5 to 120 hours (mean = 24 hours). The middle ear is filled with mesenchyme during foetal development and redistribution is complete between 8 foetal months and 13 postnatal months (Guggenheim, Clements, & Schlesinger, 1956; Piza, Northrop, & Eavey, 1998; Wolff, 1934). Amniotic fluid fills the middle ear during pregnancy which clears postnatally (De Sa, 1973; Eavey, 1993; Priner, Freeman, Perez, & Sohmer, 2003). Roberts et al. (1995) found effusion in 100% of ears at 3 hours of age which resolved in 73% to 92% of neonates by the third day of life. Jaffe, Hurtado, and Hurtado (1970) found that 83% of ears were aerated by the second day of life. Stuart, Yang, and Green (1994) concluded that conductive hearing loss in the first 2 days of life was due to fluid in the middle ear and Kok, Van Zanten, and Brocuar (1992) attributed improvement in transient evoked otoacoustic emissions (TEOAEs) over the first days of life to the clearing of amniotic fluid. The process of vernix, mesenchyme and amniotic fluid clearing

from the outer and middle ear are a part of normal development, but persistent amniotic fluid can cause neonatal otitis media, a foreign-body type inflammation which creates granulation tissue, obstructing the pathways of attic aeration and middle ear clearance (Palva, Northrop, & Ramsay, 2001; Syggelou, Fanos, & Iacovidou, 2011). Neonatal otitis media can predispose children to recurrent otitis media throughout infancy (Palva, Johnsson, & Ramsay, 2000; Syggelou et al., 2011).

1.2.3 Disease burden

1.2.3.1 Epidemiology

Otitis media is one of the most common childhood conditions, affecting almost all children at least once by the time they reach school age (Casselbrant & Mandel, 2014; Harmes et al., 2013). It is a disease of infancy most prevalent in children under 2 years, with peak incidence from 6 to 18 months of age (Alho, Koivu, Sorri, & Rantakallio, 1991; Casselbrant, Mandel, Kurs-Lasky, Rockette, & Bluestone, 1995; Engel, Anteunis, Volovics, Hendriks, & Marres, 1999; MacIntyre et al., 2010; Paradise et al., 1997; Teele, Klein, Rosner, & the Greater Boston Otitis Media Study Group, 1989; Todberg et al., 2014; Wright, McConnell, Thompson, Vaughn, & Sell, 1985). Otitis media incidence declines after infancy with a secondary peak at 5 to 6 years of age when children enter school (Teele et al., 1989; Zielhuis, Rach, Bosch, & Broek, 1990). Ten to 17 percent of infants suffer from recurrent otitis media in the first year of life (Kero & Piekkala, 1987; Teele et al., 1989). These are a population of interest because they are at risk of speech and language delay as they spend a greater period of time with hearing loss.

Recent research has found a decline in otitis media diagnoses in developed countries. It is estimated that in the United States otitis media prevalence in infants has declined by 20%, and in Canada, prevalence in 2- to 3-year-old children has decreased from 26% to 13% (Fortanier et al., 2014; Hoffman et al., 2013; Marom et al., 2014; Taylor et al., 2012). Factors contributing to decrease in prevalence estimates include the introduction of the pneumococcal conjugate vaccine, increased uptake of a "watchful waiting" approach to otitis media management by parents resulting in less General Practitioner (GP) visits (and therefore less diagnoses), stricter diagnostic criteria being used by GPs, children being less exposed to tobacco smoke, and introduction of the influenza vaccine (Taylor et al., 2012).

In Australia, it is estimated that 73% of infants will have at least one episode of otitis media in the first year of life (Mahadevan et al., 2012). Ear problems are the fourth most common issue managed by GPs, and otitis media accounts for 23% of all antibiotics provided to children (Gunasekera et al., 2007). Insertion of tympanostomy tubes for management of recurrent or chronic otitis media with effusion is

the second most common procedure performed in hospital on children (Kong & Coates, 2009). In 2003, there were an estimated 1,174,267 cases of otitis media in Australia with 68% of these occurring in children under 14 years old. Australian Indigenous children suffer a greater burden of otitis media with prevalence rates among the highest in the world. Morris et al. (2005) found otitis media was almost universal in 6- to 18-month-old Indigenous infants. Otitis media in Indigenous children is also less likely to resolve spontaneously and more likely to progress to chronic disease (Daly et al., 2010; Leach, 1999; Mahadevan et al., 2012; Yiengprugsawan, Hogan, & Strazdins, 2013).

Because of the extraordinarily high rate of otitis media in Indigenous children, research efforts in Australia have focused on this population. There is a lack of evidence about otitis media prevalence in the general population with only three studies investigating otitis media prevalence in the general population of Australia since 1980. Lehmann et al. (2008) followed 180 non-Indigenous children from birth to 2 years of age from 1999 to 2005 in Western Australia. Otitis media was diagnosed by LFT and at least one otology consultation (a limitation being that LFT was used in infants from three months of age). They found that prevalence peaked at 40% in 10- to 14-month-old infants and reduced to 28% by 20 to 24 months of age. The authors concluded that in the non-Indigenous population, prevalence was comparable to studies from other developed counties. Brennan-Jones et al. (2014) reported on a cohort of 2280 children born in Western Australia between 1989 and 1991. Otitis media was diagnosed by parental report supplemented by otoscopy. Prevalence of recurrent otitis media (≥3 episodes of otitis media in the first 3 years of life) was 26.8%, and severe recurrent otitis media (≥8 episodes) was 5.5%. Yiengprugsawan et al. (2013) analysed data collected as a part of the Longitudinal Study of Australian Children, a large Australia-wide cohort of 4242 children born in 2003 to 2004, following them from birth until 7 years of age (as well as a cohort of older children). They found a peak prevalence of "ongoing" ear infections of 5.4% in 2- to 3-year-olds, which could include recurrent or chronic otitis media. This figure is lower than the recurrent otitis media prevalence at 3 years of age reported by Brennan-Jones et al. (2014), but very close to the severe recurrent otitis media figure they reported of 5.5%. However, the question in the Longitudinal Study of Australian Children questionnaire used by Yiengprugsawan et al. (2013) was too general to make meaningful comparisons between the studies. An important consideration is that Yiengprugsawan et al. (2013) collected data after the introduction of the pneumococcal conjugate vaccine which is known to have lowered otitis media prevalence. Also, the Longitudinal Study of Australian Children only used parent-reported measures, which may have underestimated prevalence of an often asymptomatic condition (Zielhuis et al., 1990).

1.2.3.2 Complications and sequelae

Although usually self-resolving, otitis media can progress to more severe forms of disease that can cause disability and even death. It is estimated that approximately 21,000 people die each year world-wide due to complications from otitis media, mostly due to brain abscess and meningitis (Monasta et al., 2012). While these conditions do not often lead to deaths in developed countries, complications from otitis media can still be significant, and include serious conditions such as chronic suppurative otitis media, cholesteatoma and mastoiditis. Conductive hearing loss is the most common complication of otitis media in infants and can go on to affect language, auditory processing, learning, and behaviour.

Tympanic membrane perforation is one of the most common complications of otitis media. The perforation may be either acute or chronic and can be accompanied by otorrhea (Bluestone & Klein, 2007). A large perforation can cause significant conductive hearing loss (Klein, 2000). Over 87,000 children experienced tympanic membrane perforation as a result of otitis media in Australia in 2008 (Mahadevan et al., 2012). This condition is much more common in Indigenous children occurring in 14% and 40% by 6 and 18 months of age, respectively (Morris et al., 2005).

Otitis media can progress to chronic suppurative otitis media, a more serious form of ear disease that affects over 60,000 children in Australia each year (Kong & Coates, 2009). It is much more common in Indigenous children affecting 15% by 30 months of age (Morris et al., 2005). Chronic suppurative otitis media has associated conductive hearing loss in 60% of cases, usually more severe than the degree of loss typically caused by otitis media (Mahadevan et al., 2012). Without treatment, chronic suppurative otitis media can progress to more severe forms of disease such as cholesteatoma or mastoiditis (Bluestone & Klein, 2007). Cholesteatoma is a growth in the middle ear or mastoid sinus that occurs as a result of chronic otitis media that usually needs surgical management (Bluestone & Klein, 2007; Lee et al., 2013). It can cause erosion of the ossicles leading to permanent conductive hearing loss. Mastoiditis is an infection of the mastoid periosteum and air cells, caused by infection spreading from the middle ear to the mastoid structures (Lee et al., 2013; Qureishi et al., 2014). Over 200 children under 14 years of age in Australia suffer from this disease each year (Mahadevan et al., 2012). It can occur as a complication of either acute or chronic otitis media, and is a serious condition that can lead to meningitis or brain abscess without treatment (Bluestone, 1998; Mahadevan et al., 2012).

Otitis media is the most common cause of balance problems in children and can have lasting effects even after middle ear effusion has cleared (Bluestone & Klein, 2007; Casselbrant et al., 2000).

Labyrinthitis can occur if infection spreads to the inner ear through the round window and can result in severe to profound sensorineural hearing loss if untreated (Bluestone & Klein, 2007). Facial paralysis can be caused by infection spreading to the facial nerve which comes in close proximity to the middle ear. It can be a complication of acute otitis media, chronic suppurative otitis media, or mastoiditis (Bluestone & Klein, 2007). Intracranial complications can occur as a result of spread of infection from the middle ear and mastoid air cells to the brain meninges (Qureishi et al., 2014). This can lead to meningitis or brain abscess which are the main causes of otitis-media-related deaths worldwide (Bluestone & Klein, 2007; Monasta et al., 2012). Although rare in developed countries, intracranial complications do occur and can be life-threatening (Jung et al., 2013; Penido et al., 2005). Over 200 children had intracranial complications as a result of otitis media in Australia in 2008 (Mahadevan et al., 2012).

Conductive hearing loss caused by middle ear effusion is the most common complication of otitis media in the developed world (Klein, 2000). In Australia, over 350,000 children are affected by transient conductive hearing loss each year and 15 children will develop a permanent hearing loss as a result of otitis media (Hoffman et al., 2013; Kong & Coates, 2009; Mahadevan et al., 2012). Around 50% of children with otitis media experience mild conductive hearing loss and 5% to 10% develop a moderate loss. The degree of loss is determined by the amount of middle ear effusion (Fria, Cantekin, & Eichler, 1985; Roberts et al., 2004a). Even a mild conductive hearing loss can cause difficulty hearing soft speech sounds (Klein, 2000). Olmsted, Alvarez, Moroney, and Eversden (1964) found that hearing loss associated with otitis media persisted for one to six months in 55% of children and for longer than six months in 12%. Otitis media is also associated with temporary and permanent sensorineural hearing loss (da Costa, Rosito, & Dornelles, 2009; Joglekar et al., 2010). Temporary sensorineural hearing loss can be caused by middle ear effusion increasing pressure on the round window of the inner ear (Bluestone & Klein, 2007). Permanent sensorineural hearing loss can result from the spread of toxins or infection into the inner ear through the round window or from complications such as labyrinthitis, chronic suppurative otitis media, and cholesteatoma (Bluestone & Klein, 2007; Klein, 2000). The extended high frequencies (10,000 to 20,000 Hz) are most commonly affected because the basal end of the cochlea is closest to the round window (Margolis, Rykken, Hunter, & Giebink, 1993; Margolis, Saly, & Hunter, 2000). Otitis media has also been linked to hearing problems later in life. An Australian study found that hearing problems in 8- to 9-year-olds were predicted by ear infections at 4 to 5 years of age (Yiengprugsawan et al., 2013). Childhood otitis media has also been linked to adult hearing loss and tinnitus (Aarhus, Tambs, Kvestad, & Engdahl, 2014; Dawes & Welch, 2010).

The effect of otitis media on speech and language development has been a controversial area of research (Roberts, Rosenfeld, & Zeisel, 2004b). There are critical periods of language development during infancy, a critical period of phonology development occurs from 6 to 12 months of age, and complex language develops in the first two years of life (Ruben, 1997). It is thought that infants who suffer auditory deprivation due to otitis-media-related hearing loss during a critical period are at risk of speech and language delay (Klein, 2000; Ruben, 1997). However, there have been conflicting results in the literature with some studies finding an association between otitis media and language development, but others not. Keogh et al. (2005) showed that some children are more affected than others which is important because even if only a small proportion of children with otitis media are affected, it would still be a significant number because otitis media is such a common condition (Bluestone & Klein, 2007; Klein, 2000). There are also special populations who are more at risk such as Indigenous children who typically have more severe otitis media earlier in life than non-Indigenous children (Aithal, Yonovitz, & Aithal, 2008; Morris et al., 2009; Williams & Jacobs, 2009).

A review and meta-analysis in 2004 concluded that typically developing children with otitis media may not be at risk of speech and language delays. However, the authors cautioned that most studies in the review had used otitis media rather than hearing loss as the independent variable, potentially confounding results (Roberts et al., 2004a). They recommended that future research study the impact of duration and degree of hearing loss on speech development, rather than number of otitis media episodes. Two such studies have subsequently been conducted. Serbetcioglu, Ugurtay, Kirkim, and Mutlu (2008) found no association between results of a developmental screening test (Denver II) and otitis-media-related hearing loss. They used a case-control study design with sixteen 3- to 6-year-old children with persistent bilateral otitis media with effusion. However, a significant limitation of the study was that the children's hearing was tested when they completed the screening test, not during the earlier critical periods of development. In a prospective study that followed children from birth, Zumach, Chenault, Anteunis, and Gerrits (2010) found that hearing loss due to otitis media early in life affected language development. They showed that hearing loss during infancy affected phoneme perception at 7 years of age. This is significant as poor phoneme identification and discrimination can cause problems understanding speech in background noise which can be problematic in educational contexts such as trying to understand the teacher in a noisy classroom.

Otitis media early in life can also cause auditory processing disorder. Children with spatial processing disorder often have history of chronic otitis media, and higher prevalence of spatial processing disorder has been found in Indigenous Australian children, likely due to the high burden of otitis media in that population (Cameron, Dillon, Glyde, Kanthan, & Kania, 2014). Animal models have demon-

strated that conductive hearing loss during a critical developmental period has lasting effects on auditory processing capability (Tucci, Cant, & Durham, 1999; Webster, 1984). However, these results cannot immediately be generalised to humans because the hearing loss in animal experiments has been greater than is typically caused by otitis media (Bluestone & Klein, 2007). Results of studies in human infants have been equivocal, but this may be due to the confounding effects of using otitis media as the independent variable rather than hearing loss. In a prospective study using hearing loss as the independent variable, Zumach, Gerrits, Chenault, and Anteunis (2008) found that hearing loss caused by otitis media early in life affected results of speech-in-noise tests at school age. Recently, a study by Graydon, Rance, Dowell, and Van Dun (2017) confirmed previous research showing that conductive hearing loss early in life has long-term effects on binaural processing (Gravel & Wallace, 1992; Hall, Grose, & Mendoza, 1995; Moore, Hartley, & Hogan, 2003; Pillsbury, Grose, & Hall, 1991; Tomlin & Rance, 2014). In a review, Whitton and Polley (2011) concluded that studies using hearing loss as the independent variable consistently find long-term effects on auditory processing consistent with animal studies.

The effects of otitis media on hearing, language and auditory processing can go on to affect educational outcomes. This has also been controversial, and a review by Roberts et al. (2004a) concluded that the evidence was inconclusive. However, a large cohort study found a relationship between early otitis media and difficulties with reading and writing in 11- to 18-year-old children (Bennett, Haggard, Silva, & Stewart, 2001). An Australian study found poorer reading ability in 6- to 8-year-old children with history of otitis media (Winskel, 2006).

Behavioural issues related to otitis media include restlessness, disobedience, inattention, distractedness and limited social interaction (Bluestone & Klein, 2007; Klein, 2000). Children with socioeconomically disadvantaged background are more at risk (Paradise et al., 1999). Issues can persist throughout childhood. Gouma et al. (2011) found more hyperactivity in 6- to 8-year-old children with a history of otitis media compared to their peers and a longitudinal study by Bennett et al. (2001) found hyperactivity and inattention affected children with history of otitis media even into the teenage years.

Otitis media can also affect psychosocial health. A longitudinal Australian study found that recurrent otitis media had a long term impact on the psychosocial health of children (Hogan, Phillips, Howard, & Yiengprugsawan, 2014). Gouma et al. (2011) reported higher rates of depression in 6- to 8-year-old children with a history of otitis media compared to their peers. Children with recurrent otitis media have poorer quality of life than their healthy peers suffering disturbed sleep, loss of appetite and otalgia (Grindler, Blank, Schulz, Witsell, & Lieu, 2014). Otitis media puts extra stress on the family as

well due to loss of sleep and financial burden (Barber, Ille, Vergison, & Coates, 2014; Bluestone & Klein, 2007).

1.2.4 The importance of early intervention

Infants with early onset otitis media are at risk of recurrent and chronic disease (Daly & Giebink, 2000; Homøe, Christensen, & Bretlau, 1999; Marchant et al., 1984a), delays in language development (Ruben, 1997; Zumach et al., 2010), and auditory processing disorder (Tomlin & Rance, 2014; Villa & Zanchetta, 2014). Consequently, early intervention is of high importance. Watchful waiting is an appropriate management strategy for older children (Lieberthal et al., 2013), but many guidelines recommend immediate antibiotic treatment for acute otitis media in infants. The American Academy of Pediatrics recommends this only for infants with bilateral severe acute otitis media (Rovers et al., 2007), but Hoberman, Ruohola, Shaikh, Tähtinen, and Paradise (2013) suggest this be changed to include all infants under 2 years of age with acute otitis media, including unilateral and apparently less severe cases. There are issues, however, with using antibiotics for otitis media treatment as it contributes to antibiotic resistance and can cause side effects such as vomiting, diarrhoea or rash (Tähtinen et al., 2011; Thomas & Brook, 2014; Venekamp, Sanders, Glasziou, del Mar, & Rovers, 2015). Inappropriate and excessive use of antibiotics for otitis media treatment has been identified as a significant contributing factor to antibiotic resistance, as antibiotics are often prescribed even when the diagnosis is uncertain due to difficulties diagnosing otitis media in infants (Goossens, Ferech, Vander Stichele, Elseviers, & ESAC Project Group, 2005; Klein, 2000; Mahadevan et al., 2012). These issues highlight the need for accurate diagnosis of otitis media for targeted use of antibiotics in order to maximize benefit but limit harm (Hoberman et al., 2011; Thomas & Brook, 2014).

1.3 Diagnosing otitis media in infants

Early detection is essential before intervention can occur but diagnosis of otitis media in infants is difficult (Aithal et al., 2013; Karma, Penttilä, Sipilä, & Kataja, 1989a; Lee et al., 2013; Morris et al., 2009; Syggelou et al., 2011). Signs of ear infection, such as pulling the ear or ear pain are unreliable diagnostic markers, as otitis media in infants is often asymptomatic (Baraibar, 1997; Berkun et al., 2008; Marchant et al., 1986; Marchant et al., 1984a; Zielhuis et al., 1990). Furthermore, the tools that have been developed and validated for diagnosis in older children and adults are often inaccurate and impractical when used in infants (Aithal et al., 2013). Research has showed that GPs are not confident in their otitis media diagnosis 58% of the time in infants under 12 months old (Froom et al., 1990). Another study found that accuracy of paediatricians diagnosing otitis media in infants under 2 months

of age was under 50% (Berkun et al., 2008). Accurate diagnosis of otitis media early in life is vital, but remains a significant challenge.

1.3.1 Current methods of diagnosis

1.3.1.1 Myringotomy and medical imaging

Myringotomy and medical imaging are considered the gold standards for determination of presence or absence of middle ear effusion. Myringotomy is a surgical procedure that involves an incision into the tympanic membrane to sample or remove effusion. It is often performed in conjunction with tympanostomy tube insertion. However, myringotomy may not be a perfect gold standard, as use of nitrous oxide during anaesthesia can increase pressure in the middle ear and cause fluid to drain through the Eustachian tube prior to surgery (Nozza, Bluestone, Kardatzke, & Bachman, 1992; Sassen, Aarem, & Grote, 1994). Use of myringotomy or medical imaging as gold standards in otitis media research is limited to clinical studies, as it is unethical to perform these procedures in healthy infants (Aithal, Aithal, Kei, & Driscoll, 2012; Hoffmann et al., 2013).

1.3.1.2 Otoscopy and otomicroscopy

Otoscopy involves visualisation of the tympanic membrane under magnification. It can be performed under pressurised or non-pressurised conditions. Non-pressurised otoscopy is not recommended as the primary method of otitis media diagnosis as it is unable to accurately identify middle ear effusion (Rosenfeld et al., 2004; Sassen et al., 1994; Takata et al., 2003). Pneumatic otoscopy changes the air pressure in the ear canal with an insufflation bulb while visualising the tympanic membrane. It is more accurate in identifying middle ear effusion than non-pressurised otoscopy, and has been recommended as the primary method for diagnosing otitis media in older children (Chianese et al., 2007). Pneumatic otoscopy diagnostic accuracy studies that have included infants using myringotomy as the gold standard have found sensitivity ranging from 68% to 93% and specificity from 58% to 81% (Cantekin et al., 1979; Finitzo, Friel-Patti, Chinn, & Brown, 1992; Nozza, Bluestone, Kardatzke, & Bachman, 1994; Rogers, Boseley, Adams, Makowski, & Hohman, 2010). However, the age range of participants in these studies was wide and most of the subjects were older children, not infants. There have been no studies evaluating the test performance of pneumatic otoscopy in neonates.

A major limitation of pneumatic otoscopy is that accuracy differs between testers (Marchant et al., 1986). The amount of pressure applied to the ear canal varies widely because it is not standardised, and interpretation is subjective (Cavanaugh, 1989). The main factor in diagnostic accuracy is the skill level of the otoscopist (Rogers et al., 2010). Pneumatic otoscopy is accurate when a highly trained otologist

performs the test, but the accuracy decreases when GPs, paediatricians, registrars or nurses are testing (Froom et al., 1990; Pichichero & Poole, 2001; Rogers et al., 2010; Sorrento & Pichichero, 2001). Also, pneumatic otoscopy relies on clear visualisation of the tympanic membrane. Debris in the ear canal such as wax or vernix needs to be removed prior to testing but this is a difficult task in infants (Aithal et al., 2013; Baraibar, 1997; Chianese et al., 2007; Sakran et al., 2006; Syggelou et al., 2011; Turner et al., 2002). Available research reports pneumatic otoscopy as unreliable and difficult to perform in young infants due to the small compliant ear canals changing shape with insufflation (Marchant et al., 1986; Marchant et al., 1984a). The horizontal orientation of the tympanic membrane makes clear visualisation difficult, and even when visualised clearly, landmarks are difficult to interpret (Aithal et al., 2012; Baldwin, 2006; Iacovidou, Falaena, Alexaki, & Nika, 2010; Syggelou et al., 2011). Even healthy eardrums look opaque and are less compliant than those of older children leading to false positive diagnoses (Baldwin, 2006; Berkun et al., 2008; Cavanaugh, 1987; Marchant et al., 1986; Marchant et al., 1984a; Roberts et al., 1995; Syggelou et al., 2011). Furthermore, infants can find the procedure distressing and become uncooperative, further increasing difficulties in the assessment (Baraibar, 1997).

Otomicroscopy examines the eardrum under a binocular microscope. It is more accurate than pneumatic otoscopy for otitis media diagnosis in older children (Lee, 2010). Diagnostic otomicroscopy studies that have included infants have found sensitivity ranging from 0.88 to 0.94 and specificity from 0.89 to 0.94 (Rogers et al., 2010; Young, Ten Cate, Ahmad, & Morton, 2009). Level of training is an important factor in accuracy with consultant doctors being more accurate than registrars. Limitations of otomicroscopy are that the test requires a highly trained otologist and expensive equipment (Lee, 2010). Even for otologists, the procedure is difficult in young infants because of the tiny structures of the ear and need for cooperation (Hoffmann et al., 2013). No studies have investigated the feasibility or diagnostic accuracy of otomicroscopy in young infants.

1.3.1.3 Tympanometry

Tympanometry is a test of middle ear function that measures acoustic admittance (*Y*) as a function of pressure in the ear canal. The resulting graph is called a tympanogram. *Y* is a complex measurement with a real part, conductance (*G*) and an imaginary part, susceptance (*B*). A previous limitation of tympanometry was that *Y* was measured in arbitrary units in the first generation of instruments. Thus, measurements were not comparable between subjects and interpretation was limited to qualitative pattern matching (Smith et al., 2006). The qualitative method of interpretation classifies a tympanogram as type A (normal) if there is a single peak close to ambient air pressure (~0 daPa), type C if there

is a peak at negative pressure (suggestive of Eustachian tube dysfunction) and type B if there is no peak (indicating middle ear effusion) (Jerger, 1970). The next generation of equipment had automatic gain control allowing measurement in absolute physical units (mmho). This allowed for direct comparison between subjects and studies. As well as qualitative pattern matching, tympanograms could now be characterised by tympanometric peak pressure, peak compensated admittance magnitude (Y_{TM}), ear canal volume, and tympanometric width (Shanks & Shohet, 2009). Initially, LFT using a 220/226 Hz probe tone was found to be clinically useful for detecting middle ear effusion in children and adults (Jerger, 1970; Lidén, 1969). It soon became apparent, however, that LFT was inaccurate in infants under 7 months old, with high rates of false positives and false negatives (Alaerts, Luts, & Wouters, 2007; Baldwin, 2006; Paradise, Smith, & Bluestone, 1976). Research using a 660/678 probe tone reported increased accuracy in detecting middle ear disorders, but often resulted in tympanograms with complex notching patterns that were difficult to interpret. Results of HFT studies using a 1000 Hz probe tone were easier to interpret, and more accurate than studies using 660/678 tympanometry (Alaerts et al., 2007; Baldwin, 2006).

There have been various methods proposed for classifying HFT results. Baldwin (2006), used a qualitative method based on Marchant et al. (1986) where a line is drawn between the Y values at the positive and negative pressure extremes. A tympanogram is classified as normal if a peak extends over the line, otherwise abnormal. Other approaches have recommended a combination of qualitative and quantitative measures to interpret HFT results. The morphology plus magnitude compensated approach classifies a tympanogram as normal if there is a peaked trace and Y_{TM} compensated at +200 daPa (Y_{+200}) lies within a predefined normative range. Alternatively, Y_{TM} can be compensated at −400 daPa (Y_{-400}) but this can be problematic as the ear canals of young infants collapse before reaching this point (Kei et al., 2003; Kei & Mazlan, 2012b; Margolis et al., 2003). The morphology plus component compensated approach is similar but the component parts of Y (G and B) are each compensated separately. It has been suggested that this approach may better separate normal from diseased ears because it results in larger values (Alaerts et al., 2007; Calandruccio et al., 2006; Kei & Mazlan, 2012b; Kei, Mazlan, Hickson, Gavranich, & Linning, 2007; Mazlan et al., 2009b). Which method is most accurate in identifying middle ear effusion is an area of ongoing research (Kei & Mazlan, 2012b).

Table 1.1 summarises studies investigating the diagnostic accuracy of tympanometry that have included infants. HFT studies that have included neonates have all reported a high specificity but sensitivity has been variable. HFT studies where the sample consisted entirely of neonates have all found poor sensitivity. However, only DPOAEs were used as the reference standard in these studies (Margolis et al., 2003; Swanepoel et al., 2007). Research using a stronger reference standard (otomi-

croscopy or ABR) have shown much better sensitivity but these studies have had a broader age range, including young infants as well as neonates (Baldwin, 2006; Hoffmann et al., 2013). HFT studies including young infants 2 weeks to 6 months old have all reported high sensitivity and specificity (Baldwin, 2006; Hoffmann et al., 2013; Zhiqi, Kun, & Zhiwu, 2010). Notably, Baldwin (2006), Prieve et al. (2013b) and Zhiqi et al. (2010) reported high accuracy using a strong reference standard (ABR or computerised tomography scan). LFT studies in infants older than 6 months have mostly shown satisfactory test performance with the exception of a study by Hoffmann et al. (2013) who found poor sensitivity in the 6- to 9-month age group and Palmu and Syrjänen (2005) who reported sensitivity of 0.61 in a group of 7- to 10-month-old infants. Most studies of LFT in infants over 9 months old have reported good test performance.

Table 1.1. Diagnostic accuracy studies of tympanometry that have included infants in the sample

Study	Age	n	*f*	Reference standard	Se	Sp	AUC
Margolis et al. (2003)	15 to 76 h	87	1000	DPOAE	0.50	0.91	
Sanford et al. (2009)	9 to 58 h	230	1000	DPOAE	0.36	0.91	0.75
Swanepoel et al. (2007)	1 to 28 d	143	1000	DPOAE	0.57	0.95	
Hoffmann et al. (2013)	<3 m	464	1000	Otomicroscopy	0.70	0.89	
Baldwin (2006)	2 to 21 w	211	1000	TEOAE and ABR (AC & BC)	0.99	0.89	
Hoffmann et al. (2013)	3 to 6 m	313	1000	Otomicroscopy	0.85	0.89	
Zhiqi et al. (2010)	42 d to 6 m	52	1000	CT	0.98	0.98	
Prieve et al. (2013b)	3 to 36 w	60	1000	Tone-burst ABR (AC & BC)			
Hoffmann et al. (2013)	6 to 9 m	99	226	Otomicroscopy	0.43	0.75	
Hoffmann et al. (2013)	6 to 9 m	99	1000	Otomicroscopy	1.00	0.93	
Palmu, Puhakka, Rahko, and Takala (1999)	7 to 11 m	58	226	Myringotomy or PO	0.79	0.99	
Palmu and Syrjänen (2005)	7 to 10 m	630	226	Myringotomy or PO	0.61	0.99	
Hoffmann et al. (2013)	9 to 12 m	38	226	Otomicroscopy	0.88	0.75	
Chianese et al. (2007)	6 to 24 m	786	226	PO			0.83
Smith et al. (2006)	6 m to 3 y	3686	226	PO			0.84
Sassen et al. (1994)	5 m to 11 y	266	226	Myringotomy	0.83	0.63	
Finitzo et al. (1992)	6 m to 9 y	86	226	Myringotomy	0.90	0.86	
Nozza et al. (1992)	1 to 8 y	61	226	Myringotomy	0.90	0.86	
Nozza et al. (1994)	1 to 12 y	171	226	Myringotomy	0.80	0.82	

Prieve et al. (2013b) reported likelihood ratios (LR) rather than sensitivity, specificity or AUC: LR+ 32.03, LR− 0.073. ABR, auditory brainstem response; AC, air conduction; AUC, area under the receiver operating characteristic curve; BC, bone conduction; CT, computerised tomography; DPOAE, distortion product otoacoustic emissions; PO, pneumatic otoscopy; *f*, probe-tone frequency; se, sensitivity; sp, specificity; TEOAE, transient evoked otoacoustic emissions.

However, there are limitations to using tympanometry as a test of middle ear function. First, tympanometry is poor at detecting partial fluid in the middle ear (Palmu & Syrjänen, 2005; Shanks & Shohet, 2009). Also, the test requires pressurisation of the ear canal which some infants find distressing. Pressurisation also causes the ear-canal wall to move in young infants. Significant change in ear canal volume in response to pressurisation has been recorded in neonates which violates the assumption that the ear canal behaves like a rigid-walled cavity (Holte, Cavanaugh, & Margolis, 1990; Prieve et al., 2013b). Moreover, tympanometry utilizes only simple signal processing strategies which render results susceptible to artefact. This can be an issue when testing infants who are prone to move frequently during testing (Liu et al., 2008).

1.3.1.4 Acoustic stapedial reflexes

The ASR test uses a probe tone to measure changes in Y in response to an activation stimulus (pure-tone or noise) presented to either the ipsilateral or contralateral ear. ASRs are usually performed at tympanic peak pressure, and therefore done in conjunction with tympanometry. In normal ears, the stimulus causes the stapedius muscle to contract, stiffening the middle ear system. This typically causes Y to decrease, but can also cause an increase through decoupling of the stapes from the inner ear. The ASR threshold is the lowest intensity level that the stimulus elicits a reflex. ASRs can be used to assist in differential diagnosis of auditory conditions including diagnosis of otitis media (Freyss, Narcy, Manac'h, & Toupet, 1979; Kei & Mazlan, 2012a; Marchant et al., 1986; Nozza et al., 1992, 1994). Traditionally, a 226 Hz probe tone has been used but researchers have found that using a high-frequency probe tone (>800 Hz) to be more effective in neonates because the ASR is present more often in normal ears at lower activation levels (De Lyra-Silva, Sanches, Neves-Lobo, Ibidi, & Carvallo, 2015; Jacob-Corteletti et al., 2015; Mazlan, Kei, & Hickson, 2009a; Weatherby & Bennett, 1980). An activating tone of 2000 Hz or BBN has been recommended for testing neonates (De Lyra-Silva et al., 2015).

Diagnostic accuracy studies using myringotomy as the gold standard that have included infants have found sensitivity from 0.86 to 0.92 and specificity from 0.52 to 0.85 (Cantekin et al., 1979; Nozza et al., 1992, 1994). Nozza et al. (1994) reported that using tympanometry in conjunction with ASRs helped to improve specificity from 0.65 to 0.78. As the youngest infants in these studies were 7 to 12 months old, the test performance for infants younger than 7 months remains unknown.

There are significant limitations to using ASR for otitis media diagnosis in infants. Results can be difficult to interpret, meaning it is not always possible to make a diagnosis (Nozza et al., 1992). Also, results are easily contaminated by artefact, testing requires the subject to remain absolutely still and

quiet which is often not possible when testing infants. Finally, the test requires loud sounds to be presented to the ear which infants can find distressing.

1.3.1.5 Evoked otoacoustic emissions

Otoacoustic emissions are low-level sounds produced by non-linear processes in the cochlea that can be recorded in the ear canal. Spontaneous otoacoustic emissions only occur in approximately 50% of normal-hearing subjects, and as such they are not useful clinically (Prieve & Fitzgerald, 2015). EOAEs, however, can be measured in almost all normal-hearing ears and are regularly used clinically in both diagnostic and screening contexts. EOAEs can be a useful aid in middle ear assessment as the stimulus and resulting emissions need to travel through the middle ear before being measured in the ear canal. EOAEs assess cochlear function, but can only be recorded when the middle ear is functioning normally. Consequently, the absence of EOAEs may be suggestive of middle ear pathology but could also be due to sensory dysfunction. The most commonly used EOAEs are TEOAEs and distortion product otoacoustic emissions (DPOAEs). TEOAEs are elicited with a click stimulus and DPOAEs are elicited through the interaction of two primary tones f_1 and f_2 ($f_1 < f_2$) with corresponding levels L_1 and L_2. The most commonly recorded emission clinically is $2f_1 - f_2$ due to the large emission produced (Prieve & Fitzgerald, 2015).

EOAEs are sensitive to a number of external and middle ear conditions affecting infants. Vernix occludes the ear canal in 13% to 28% of neonates in the first two days of life causing a temporary conductive hearing loss and TEOAEs to be attenuated (Chang, Vohr, Norton, & Lekas, 1993; Doyle et al., 1997; Doyle, Rodgers, Fujikawa, & Newman, 2000). TEAOEs are also reduced in newborn ears that have amniotic fluid in the middle ear, a condition that affects up to 50% of newborns and usually clears over the first two days of life (Doyle et al., 1997; Doyle et al., 2000). Both TEOAEs and DPOAEs are clinically useful in identifying middle ear effusion in infants and children (Choi, Pafitis, Herer, Zalzal, & Patel, 1999; Kei, Brazel, Crebbin, Richards, & Willeston, 2007; Yeo, Park, Park, & Suh, 2002). Two studies using older children (over 5 years of age) concluded that DPOAEs better differentiated between normal and diseased ears (Akdogan & Özkan, 2006; Thakur et al., 2013). The quantity of effusion in the middle ear has been shown to be a significant factor in emissions being present or absent (Koivunen, Uhari, Laitakari, Alho, & Luotonen, 2000). Animal studies have shown that emissions are not affected when fluid only half fills the middle ear, but are eliminated when the middle ear is full (Akinpelu, Funnell, & Daniel, 2015; Ueda, Nakata, & Hoshino, 1998). The quality of effusion is also an important factor as dense mucoid effusion is more likely to eliminate EOAEs than non-mucoid effusion (Amedee, 1995; Tas et al., 2004; Topolska, Hassman, & Baczek, 2000). Ear canal

air pressure also affects TEOAEs and DPOAEs, especially at frequencies below 2000 Hz (Naeve, Margolis, Levine, & Fournier, 1992; Osterhammel, Nielsen, & Rasmussen, 1993; Plinkert, Bootz, & Vossieck, 1994; Plinkert & Plok, 1994; Trine, Hirsch, & Margolis, 1993). A study including children 3 to 39 months old found that although reduced, TEOAEs were still measureable in ears with negative tympanic peak pressure (Prieve, Calandruccio, Fitzgerald, Mazevski, & Georgantas, 2008).

There is a dearth of evidence regarding the diagnostic accuracy of EOAEs in identification of middle ear effusion in infants and children. Available research report emissions being present in 0% to 50% of infants and children with middle ear effusion/chronic otitis media (Amedee, 1995; Koivunen et al., 2000; Zhao et al., 2003). A study using 63 children aged 4 to 17 years found TEOAEs to have sensitivity of 0.83 and specificity of 0.94 (Koike & Wetmore, 1999). Driscoll et al. (2001) with a sample of 940 children aged 5 to 8 years old concluded TEOAEs had a sensitivity of 0.68 and specificity of 0.90.

The limitations of using EOAEs in diagnosis of otitis media in infants are that they are sensitive to environmental and physiological noise and they are actually a test of inner ear function, they do not test the middle ear per se. Although EOAEs are often used as a reference standard for diagnostic accuracy studies in neonates, the test performance in this population is unknown.

1.3.1.6 Tests of hearing acuity

Tests of hearing acuity such as ABR and visual reinforcement audiometry can be used to determine the degree of conductive hearing loss caused by conductive conditions in infants. The ABR is an evoked potential useful for estimating hearing sensitivity in subjects either too young, incapable or unwilling to be tested by behavioural methods. Automated ABR is commonly used to screen for congenital sensorineural hearing loss in newborn hearing screening programs. A neonate passes the test if a response is detected in response to an air-conducted click stimulus of ~35 dB nHL (Northern & Downs, 2014). Automated ABR is not sensitive to middle ear effusion, and is therefore not recommended for diagnosis of conductive conditions in neonates (Aithal, Kei, & Driscoll, 2014a). Doyle et al. (2000) found that 95% of neonatal ears with effusion passed an automated ABR test. Diagnostic ABR uses click or tone-burst stimuli to measure air- and bone-conduction thresholds to determine the type and degree of hearing loss (Vander Werff, Prieve, & Georgantas, 2009). Diagnostic ABR is the gold standard hearing assessment for follow up of neonates who refer from newborn hearing screening (Joint Committee on Infant Hearing, 2007). Diagnostic ABR can be used to determine the degree of hearing loss caused by conductive conditions in young infants but the test is very time consuming and requires considerable

expertise (Aithal et al., 2015). Also, ABR cannot provide information about middle ear status as normal ABR results can be obtained in ears with otitis media (Aithal et al., 2012).

Visual reinforcement audiometry is the gold standard hearing assessment for infants aged 6 months onward (Diefendorf, 2015). The test conditions infants to turn their head towards a visual reward (typically a puppet or picture) in response to auditory stimuli (usually puretone or narrowband noise). Stimuli can be presented either in the sound field as a global test of hearing, or under headphones to obtain ear-specific hearing information (Madell, 2014). Bone-conduction testing can be performed if the infant tolerates wearing the headband of the transducer. However, ear-specific bone-conduction thresholds cannot routinely be obtained as this requires masking noise to be presented to the contralateral ear while the subject listens for the stimuli, a task that is cognitively too complex for infants. Limitations of visual reinforcement audiometry are that the test requires two highly-trained testers, infants do not always condition to the task, ear-specific air-conduction thresholds are often unable to be obtained and ear-specific bone-conduction thresholds cannot usually be obtained. Consequently, the use of this test to diagnose middle ear conditions is limited.

1.3.1.7 Combining diagnostic tests in order to improve diagnostic accuracy

No single test diagnoses otitis media perfectly, and as noted, all tests have limitations. Therefore, some researchers recommend using a battery of tests to improve diagnostic accuracy (Aithal et al., 2015; Kei & Zhao, 2012). Cantekin et al. (1979) suggested using a combination of otoscopy and tympanometry for diagnosis of otitis media in infants. Nozza et al. (1992) found improved sensitivity when combining tympanometry with ASR results, and Nozza et al. (1994) reported improved specificity when combining tympanometry with either otoscopy or ASR. The limitations of using a battery of tests are that it is time consuming, and requires clinicians with significant expertise able to perform and interpret all of the tests in the test battery.

In conclusion, all currently available clinical tests for otitis media diagnosis in infants have significant limitations. Pneumatic otoscopy and otomicroscopy have demonstrated high diagnostic accuracy in studies that have included at least some infants but these tests require highly skilled clinicians and are difficult, unreliable and have not been validated in young infants. The test battery approach may improve diagnostic accuracy but this is time consuming and requires significant training and skill which limits usefulness in contexts such as screening for otitis media. There remains the need for a quick test that is easy to administer and able to accurately diagnose middle ear dysfunction due to otitis media in infants.

1.3.2 Wideband acoustic immittance: An innovative test of middle ear function

WAI is an innovative, high-resolution test able to quickly assess middle ear function over a wide frequency bandwidth (Feeney et al., 2013; Keefe & Feeney, 2009). WAI is an umbrella term encompassing a family of wideband measurements such as pressure reflectance (*PR*), energy reflectance (*R*), energy absorbance (*A*), acoustic impedance (*Z*) and *Y*. *R* was introduced by Stinson, Shaw, and Lawton (1982) and *PR* by Hudde (1983). Keefe, Ling, and Bulen (1992) developed a quick, non-invasive method of measuring WAI in humans based on a technique for measuring *Z* in cats described by Allen (1986).

WAI has significant advantages over traditional middle ear measures that may prove important for assessment of middle ear function in infants. First, WAI is able to assess middle ear function over a greater frequency range than traditional immittance measures. Because *R* is the ratio of reflected to incident energy, it is theoretically insensitive to the location of the probe in the ear canal, and is accurate up to frequencies as high as 8000 to 10000 Hz (Ellison et al., 2012; Liu et al., 2008; Stinson et al., 1982). In comparison, traditional *Y* tympanometry cannot test at frequencies above 2000 Hz because standing waves become an issue when the wavelength of the probe tone is shorter than the length of the ear canal (Keefe & Feeney, 2009; Lilly & Margolis, 2013; Liu et al., 2008; Stinson et al., 1982). Secondly, the test time for WAI is very quick with detailed information about the middle ear obtained in just a few seconds (Vander Werff, Prieve, & Georgantas, 2007). Thirdly, WAI can be performed at both ambient air pressure and with the ear canal pressurised. Figure 1.1 shows an example of *A* measured under ambient (top panel) and pressurised (A_p; bottom panel) conditions in the same subject. Being able to test at ambient air pressure is ideal for testing infants, as some infants find pressurisation of the ear canal distressing. Also, tympanometric pressurisation of the ear canal causes significant movement in the canal wall of young infants and being able to test at ambient pressure removes this potential source of error (Aithal et al., 2015; Ellison et al., 2012; Sanford & Brockett, 2014; Vander Werff et al., 2007). However, WAI is not limited to testing at ambient pressure, and A_p, which measures *A* as a function of pressure, can be performed in a cooperative infant. The additional information obtained may be diagnostically helpful (Liu et al., 2008).

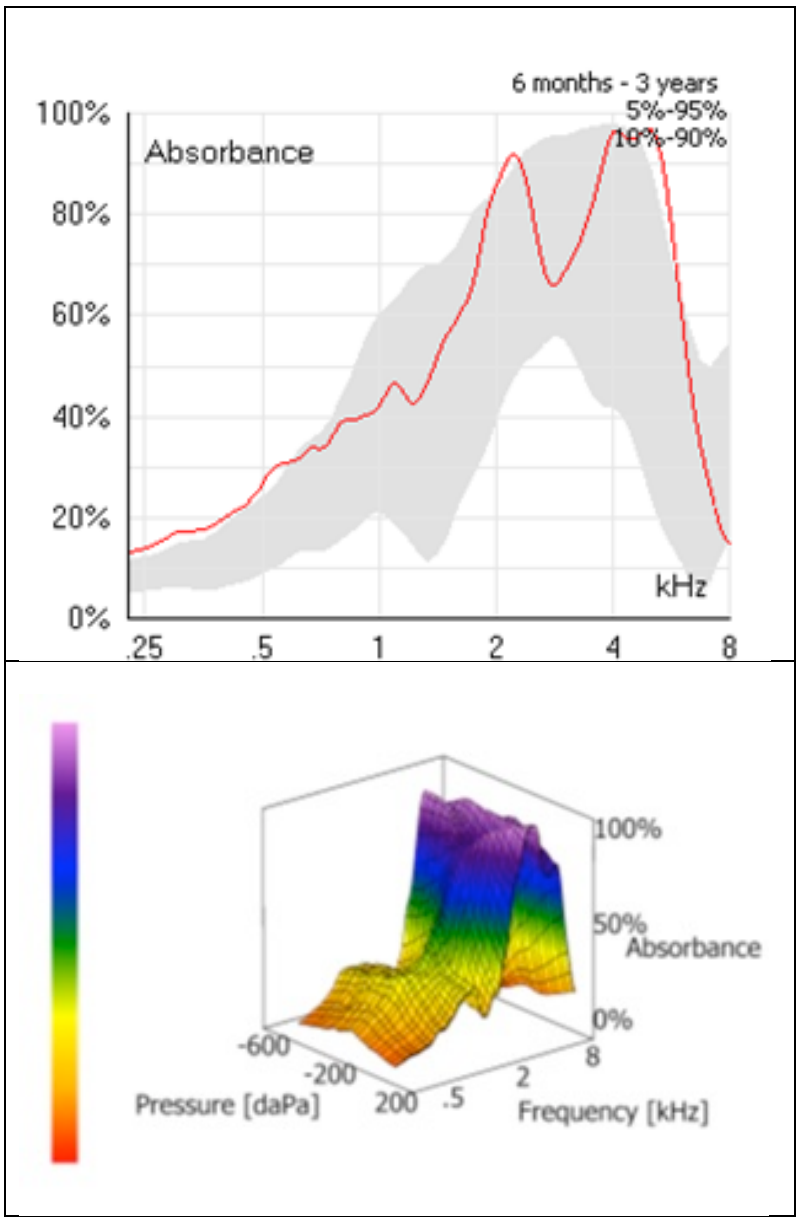


Figure 1-1. Results of ambient and pressurized energy absorbance.
Ambient results are shown in the top panel, and pressurized in the bottom. These measurements were made in the right ear of a 7-month-old male with normal middle ear function using the Interacoustics Titan system.

1.3.2.1 Principles of wideband acoustic immittance

The concept of Z is central to understanding the principles of WAI and assessment of middle ear function in general. Z is the ratio of sound pressure p (the output) to volume velocity u (the input) as a function of frequency. It is a complex measurement with magnitude $|Z|$ and phase φ_Z (Hudde, 1983; Van Camp, Margolis, Wilson, Creten, & Shanks, 1986):

$$Z = \frac{p}{u}. \tag{1.1}$$

Y is the reciprocal of Z and is also a complex measurement with a magnitude $|Y|$ and phase φ_Y:

$$Y = \frac{1}{Z}. \tag{1.2}$$

Complex measurements can also be described in rectangular form as having a real and imaginary part. The real part of Z is resistance (Re) and the imaginary part reactance (iX) such that:

$$Z = Re + iX. \tag{1.3}$$

Alternatively, the real part of Y is G and the imaginary part iB:

$$Y = G + iB. \tag{1.4}$$

The first step in determining *PR* is to calculate the characteristic impedance (Z_c) of a tube with cross-section area S:

$$Z_c = \frac{\rho c}{S}, \tag{1.5}$$

where ρ is the equilibrium density of air and c is the speed of sound (Keefe & Feeney, 2009). Using Z_c in the calculation of *PR* normalises the results in terms of ear-canal diameter. This reduces problematic inter-subject variability that was found in earlier studies, making WAI measurements clinically useful (Allen, Jeng, & Levitt, 2005; Keefe, Bulen, Arehart, & Burns, 1993). *PR* is the ratio of the incident acoustic pressure wave to the reflected acoustic pressure wave (Allen et al., 2005) and is calculated as follows:

$$PR = \frac{Z - Z_c}{Z + Z_c}. \tag{1.6}$$

R equals the square of the absolute value of *PR* (Keefe et al., 1993). It represents the proportion of incident energy that is reflected back from the middle ear (Shahnaz, Cai, & Qi, 2014):

$$R = |PR|^2. \tag{1.7}$$

A is closely related to *R* and represents the amount of energy that is absorbed by the middle ear:

$$A = 1 - R. \tag{1.8}$$

R and *A* are proportions falling between 0 and 1.

The assumptions of WAI are that: (1) no energy is lost through the ear canal wall during measurement; (2) there are no abrupt changes in cross-sectional area along the length of the ear canal; and (3) there is a leak-free seal of the probe in the ear canal (Voss, Stenfelt, Neely, & Rosowski, 2013). Assumptions one and two were tested by Voss, Horton, Woodbury, and Sheffield (2008) who measured *R* in cadaver ears. Although results revealed minor violations of assumptions, the variability introduced amounted to only a small fraction of the total reflectance. The investigators concluded that this insignificant amount of error did not undermine the integrity of the test. Violations of assumption three, however, can lead to significant error, so it is vital to obtain an air-tight seal when testing (Keefe et al., 2000; Voss et al., 2013). Acoustic leaks can be identified by visual inspection of the *R* or *A* data subsequent to measurement. *A* of ≥0.3 in the low frequencies (or ≥0.7 for neonates) indicates an acoustic leak and the probe should be reinserted in attempt to obtain a leak-free seal (Aithal et al., 2015; Groon, Rasetshwane, Kopun, Gorga, & Neely, 2015).

1.3.2.2 Instrumentation and calibration

Apart from custom-made devices, there are currently three available systems capable of making WAI measurements: the Reflwin Interacoustics system, the Mimosa Acoustics MEPA (middle ear power analyser), and the Interacoustics Titan WBT (wideband tympanometry). The Reflwin Interacoustics system is a research instrument, whereas the Mimosa MEPA and Interacoustics WBT are designed for clinical use. MEPA measures WAI at ambient pressure only, whereas the Reflwin Interacoustics and Interacoustics WBT systems can test under both ambient and pressurised conditions. The Reflwin Interacoustics system is the predecessor to Interacoustics WBT, and the calibration and measurement procedures are very similar for both units. Variation has been observed between WAI measurements made with the MEPA and Reflwin Interacoustics systems. These have been attributed to differences in calibration procedure, calculation of ear-canal cross-sectional area, and differences in probe tips used (Shahnaz et al., 2014; Shahnaz, Feeney, & Schairer, 2013). However, the small variability in measurements between systems is clinically insignificant, and Shahnaz et al. (2013) found that using system-specific norms did not improve accuracy in detecting otosclerosis.

Calibration of the Interacoustics WBT system utilises two sets of two tubes. Each tube is closed at one end and each set of tubes consists of one long tube (295 cm) and one short tube (8.4 cm). One set has a large diameter (0.794 cm) and is for calibrating measurements made in older children and adults. The other set has a small diameter (0.476 cm) and is designed for calibrating measurements in young children. The long tube is long enough that the incident signal can be separated from subsequent reflections, and the short tube is short enough that the response consists mainly of reflections (Keefe &

Simmons, 2003). The calibration procedure is the same for both sets of tubes. First, the probe is placed in the open end of the long tube. A graphical display of the waveform shows three spikes: the initial click, and the first two reflections (Figure 1.2, left panel). The probe is then placed in the short tube and the incident waveform (Q) and source reflectance (R_0) curves are shown on a display as a function of frequency (Figure 1.2, right panel). Q is the pressure wave generated by the stimulus and R_0 is the sound pressure at the probe tip (Shaw, 2009).

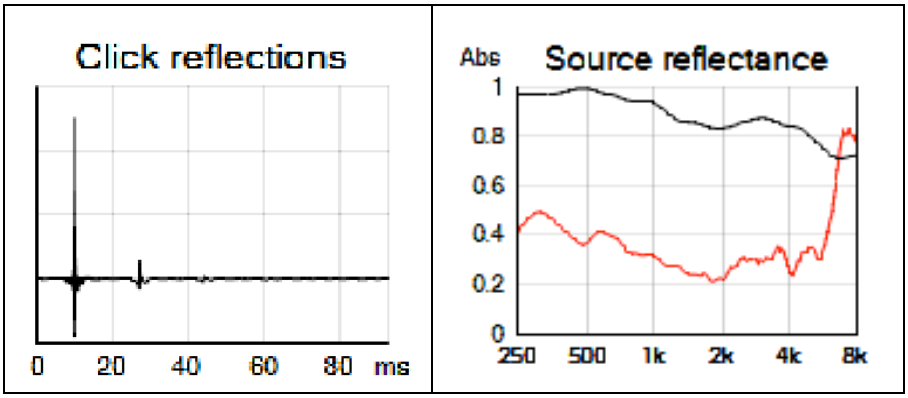


Figure 1-2. Results of the calibration procedure from the Interacoustics Titan WBT system.
These results are from the large diameter tubes. The left panel shows the click reflections. The large spike around 10 ms is the initial click and the subsequent smaller spikes are the first and second reflections. The right panel shows the incident waveform (Q, red curve) and the source reflectance (R_0, black curve).

When measuring WAI in an ear, an appropriately-sized ear tip is chosen and the probe is placed into the ear canal. The stimulus is presented and the parameters calculated during calibration (Q and R_0) are used to calculate the nominal pressure reflectance (R_n), which is PR that has not been normalised for ear canal size:

$$R_n = \frac{P - Q}{R_0 P + Q}, \tag{1.9}$$

where P is the mean of the sound-pressure spectrum. Y is then calculated from R_n and Z_c (Equation 1.5). As discussed, using Z_c controls for differences in ear canal size (Keefe & Simmons, 2003):

$$Y = Z_c^{-1} \frac{1 - R_n}{1 + R_n}. \tag{1.10}$$

PR is then calculated from Y and Z_c (this equation is the same as equation 1.6 but admittance has been substituted for impedance and then simplified, see Section 1.5: Appendix for working):

$$PR = \frac{1 - Z_c Y}{1 + Z_c Y}. \tag{1.11}$$

A is then calculated as per Equations 1.7 and 1.8. The Interacoustics WBT system gives three variables as results of from an ambient WAI test: A, $|Y|$ and φ_Y from 226 to 8000 Hz at one twenty-fourth octave frequency resolution.

1.3.2.3 Test-retest reliability

The reliability of a test is how consistent it is over repeated measurements. It is an important factor in the clinical utility of a test because too much variability may limit the ability of the test to distinguish between normal and diseased subjects (Vander Werff et al., 2007). The test-retest reliability of WAI has been investigated in neonates, infants and children and studies have concluded that measurements are adequately reliable for clinical use (Hunter, Tubaugh, Jackson, & Propes, 2008b; Vander Werff et al., 2007; Werner, Levi, & Keefe, 2010). Infants have greater variability on retest than adults as they are prone to move around more during testing (Vander Werff et al., 2007). Variability is greater when the probe is reinserted between trials, and is highest in the low frequencies (<500 Hz). However, the mid-frequency range is stable and variability between test-retest is smaller than variability between normal and diseased subjects in this frequency range (Vander Werff et al., 2007; Werner et al., 2010). The most important factors in obtaining reliable measurements in infants are adequate probe fit, proper positioning of the probe (e.g., not obstructed by the ear canal wall) and subject state (Vander Werff et al., 2007; Voss et al., 2013).

1.3.2.4 Effects of subject demographics on WAI measurements

Details of studies investigating ethnicity, gender, ear and age effects on WAI measures are presented in Table 1.2. Ethnicity effects have been studied in older children (Beers, Shahnaz, Westerberg, & Kozak, 2010) and neonates (Aithal et al., 2014a). Beers et al. (2010) investigated differences in R between Chinese and Caucasian school-aged children. Results showed a significant effect for ethnicity (with R higher in the Caucasian group at 2000 and 6000 Hz), but using ethnic-specific norms did not improve test performance. This was because differences between ethnic groups were small in comparison to differences between the normal and diseased groups. Aithal et al. (2014a) investigated differences in A between Indigenous and Caucasian neonates and found that Indigenous neonates had significantly lower A from 400 to 2000 Hz. Ethnic differences are thought to arise from body size or anatomical/functional differences in the middle ear system (Aithal et al., 2014a; Shahnaz & Bork, 2006; Shahnaz et al., 2013).

Gender and ear side differences in WAI have been investigated in neonates, infants and older children. Neonates have been studied by Aithal et al. (2014a), Aithal et al. (2013), Hunter et al. (2010),

Keefe et al. (2000) and Merchant, Horton, and Voss (2010) (see Table 1.2 for details). None of these studies reported significant differences between male and female ears except for Keefe et al. (2000), who found lower *R* in male ears at frequencies below 2000 Hz. The evidence concerning differences between neonatal right and left ears is equivocal with Keefe et al. (2000) and Merchant et al. (2010) finding a significant difference but results from Aithal et al. (2014a) and Aithal et al. (2013) did not show any difference.

In infants older than neonates, effects of ear and gender have been investigated by Aithal, Kei, and Driscoll (2014b), Hunter et al. (2008b), Shahnaz et al. (2014) and Werner et al. (2010) (see Table 1.2 for details). Of these studies, only Werner et al. (2010) found any significant ear or gender effects. In their sample, females had greater *Re* and *X* than males but there was no statistically significant gender effect for *R*. They also found a significant ear effect with *R* being lower in left than right ears. However, although significant, ear and gender effects were small. A limitation of that study was that the investigators used LFT as a reference standard for infants under 6 months old. Studies measuring ambient WAI in older children have found no significant ear or gender effects (Beers et al., 2010; Hunter et al., 2008b). In a review, Kei et al. (2013) reported that data on differences in ethnicity, ear and gender are limited and more normative research is needed. Although there are conflicting results in the literature differences that have been reported are small enough to be clinically unimportant.

Table 1.2. Studies investigating developmental or normative aspects of WAI in infants and children.

Study	*Age*	*n*	*Sample*	*Study design and purpose*	*Sys*	*WAI*	*WAI characteristics and main findings*
Keefe et al. (2000)	Neonates	2081	Healthy neonates and NICU graduates	Characteristics and validity of WAI in newborns; no reference standard used. First study to test WAI in newborns.	CM	*R* & *G*	Median *R* (250 to 8000 Hz) fluctuated from 0.1 to 0.25 and was lowest at 2000 and 6000 Hz and highest at 1000, 4000 and 8000 Hz.
Shahnaz (2008)	Neonates (32 to 51 w GA)	31	NICU neonates	Characteristics of WAI for neonates who passed an AABR and TEOAE test.	MA	*R* & *Y*	Mean *R* had a minima at 2000 Hz.
Sanford et al. (2009)	Neonates (1 to 2 d)	230	Healthy neonates	Test performance of WAI and 1-kHz tympanometry against a DPOAE reference standard.	RI	A, A_p, $\lvert Y \rvert$, & φ_Y	*A* of neonates who passed DPOAEs had two peaks, one around 1800 Hz and another around 7000 Hz.
Hunter et al. (2010)	Neonates (3 to 102 h)	324	Healthy neonates	Normative data and test performance of *R* and 1-kHz tympanometry using a DPOAE reference standard.	MA	*R*	Regions involving 2000 Hz had most predictive power. Median *R* peaked around 4000 Hz for the pass group.
Aithal et al. (2013)	Neonates (13 to 116 h)	66	Healthy neonates	Normative data using a test battery reference standard (HFT, ASRs, TEOAEs and DPOAEs).	RI	*R*	*R* was low at 1250 to 2000 Hz and high from 300 to 800 Hz and 3000 to 4000 Hz.
Merchant et al. (2010)	Neonates (3 to 5 d) and 1-m-old infants	18	Healthy infants	Comparison of normative regions for newborn and 1-mth-old infants who passed newborn hearing screening and DPOAE.	MA	*R* & *A*	*R* was not significantly different between the two age groups, although there was a possible difference around 2000 Hz.
Aithal et al. (2014b)	Neonates and infants (1, 2, 4, 6 m)	96	Healthy neonates and infants	Prospective cross-sectional study using a HFT + DPOAE reference standard.	RI	*A*	*A* was greatest between 1500 to 5000 Hz and lowest <1500 and >5000 Hz for all age groups. There was a multi-peaked pattern for 0- to 2-month-old and a single peak for the 4- to 6-month-old groups. A developmental effect was observed with newborns and 6-month-olds different from the other groups. The 1- and 2-month-old groups were similar.
Hunter et al. (2016)	Neonates and infants (1, 6, 9, 12 m)	184	Healthy and NICU neonates	Longitudinal study; reference standards of TEOAE/AABR at birth and DPOAE + TBABR + VRA in infancy	CM	A, A_p, & *GD*	There were large developmental effects on WAI variables over the first 6 months of life, especially at low frequencies for both ambient and pressurized responses.

Keefe et al. (1993)	Infants (1, 3, 6, 12, 24 m)	78	Healthy infants	WAI measured in infants of different age groups to investigate developmental effects. No reference standard used.	CM	*R* & *Z*	*R* was high at low frequencies for all ages and was lower in 1-mth-old babies at frequencies <2000 Hz. *R* was lowest at all ages between 1000 to 4000 Hz. A systematic maturational effect was not complete by 2 years of age.
Sanford & Feeney (2008)	Infants (4, 12, 27 w)	60	Healthy full-term infants	Cross-sectional study investigating developmental effects on pressurised WAI measures.	CM	*R*, |*Y*|, φ_Y, & *G*	Not a lot of change developmentally from 750 to 2000 Hz. All age groups were different from adults with the largest difference in the youngest group (1-mth-olds).
Werner et al. (2010)	Infants (2 to 3 and 5 to 9 m)	458	Healthy full-term infants	Normative data for 2- to 9-month-old infants using a LFT reference standard.	CM	*R*, *Re* & *X*	An age effect was evident. *R* was lowest from 1000 to 4000 Hz for both age groups and results compared well with Keefe et al. (1993). Separate norms for different age groups are necessary.
Shahnaz et al. (2014)	Infants (1 to 6 m)	31	Normal hearing infants	Longitudinal study (tested monthly) using a reference standard of HFT, ASRs and TEOAEs.	MA	*R*	*R* systematically increased at low frequencies (<400 Hz) and decreased at high frequencies (>2000 Hz). Younger babies had lower *R* at low frequencies and higher *R* at high frequencies. The 600 to 1600 Hz frequency region was relatively insensitive to developmental effects. Age-specific norms are necessary.
Hunter et al. (2008b)	Infants and children 3 d to 4 y	97	Well infants and children	Normative and reliability study using a DPOAE and tympanometry (226 and 1000 Hz probe tone) reference standard.	MA	*R*	No differences were found across age groups except at 6000 Hz. This is the only study to not find significant developmental effects.
Beers et al. (2010)	Children (normal group, 5 to 6 y)	142	Elementary school-aged children	Study comparing normal ears to ears with middle ear dysfunction using PTA (AC and BC), LFT and TEOAEs as reference standard.	MA	*R*	Normative data for 5- to 7-year-old children.

Columns show age, number and characteristics of subjects, study design, the system used and WAI measured and a summary of main findings. *A,* ambient energy absorbance; A_p, pressurised energy absorbance; AC, air conduction; ASR, acoustic stapedial reflexes; AABR, automated auditory brainstem response; BC, bone conduction; CM, custom-made equipment; DPOAE, distortion product otoacoustic emissions; *G*, conductance; GA, gestational age; GD, group delay; HFT, high-frequency tympanometry; LFT, low-frequency tympanometry; MA, Mimosa Acoustics MEPA system; φ_Y, acoustic admittance phase angle; PTA, puretone audiometry; *R*, ambient energy reflectance; *Re*, resistance; RI, Reflwin Interacoustics system; Sys, system; TBABR, toneburst auditory brainstem response; TEOAE, transient evoked otoacoustic emissions; VRA, visual reinforcement audiometry; *X*, reactance; *Y*, acoustic admittance; *Z*, acoustic impedance.

WAI measurements show a systematic developmental trend throughout infancy and into childhood (Beers et al., 2010; Keefe et al., 1993; Kei et al., 2013). Most studies have found significant developmental trends with the exception of Hunter et al. (2008b) who reported no differences except at 6000 Hz in a sample of infants and children ranging from 3 days to 4 years old. This finding could be attributed to subjects being grouped into wider age ranges than other studies or from differences in probe design and calibration (Hunter et al., 2008b; Shahnaz et al., 2014). Variability between studies can occur for a multitude of reasons including different study methodologies, sample size and characteristics, equipment, calibration procedure, test environment, age and reference standards (Aithal et al., 2014b; Aithal et al., 2015; Kei et al., 2013).

Changes in WAI have been reported as early as over the first few days of life (Hunter et al., 2010). Sanford et al. (2009) described changes in WAI measurements over the first two days of life and Keefe et al. (2000) found that 1-day-old neonates had higher *R* compared to neonates 2 to 4 days old. These changes were attributed to the clearing up of fluid and debris in the outer and middle ear after birth.

Aithal et al. (2014b), Shahnaz et al. (2014), and Keefe et al. (1993) have investigated changes in WAI over the first six months of life (see Table 1.2 for details). Shahnaz et al. (2014) tested subjects monthly from 1 to 6 months of age using a longitudinal study design. Aithal et al. (2014b), using a cross-sectional design reported data from neonates and 1-, 2-, 4- and 6-month-old infants. Keefe et al. (1993) tested 1-, 3- and 6-month-old infants (a limitation being that they did not use a reference standard). Figures 1.3A–C show average *A* for each age group of these studies. All studies show that *A* decreases in the low frequencies and increases in the mid frequencies over the first six months of life. Aithal et al. (2014b) and Keefe et al. (1993) found a trend of *A* decreasing in the high frequencies also. This effect was not apparent in the Shahnaz et al. (2014) study but this could be because the highest frequency tested was 6300 Hz.

Figures 1.4A–C show average *A* from normative studies for neonates, 6-, 12- and 24-month-old infants (see Table 1.2 for details of these studies). Although there is variability between studies, especially in the six-month-old group, developmental trends are clear, and show a pattern of *A* systematically decreasing in the low frequencies, increasing in the mid frequencies and decreasing in the high frequencies throughout infancy. *A* in the neonatal period tends to have two maxima, one in the 1000 to 2000 Hz region and one around 4000 to 6000 Hz, with two minima around 500 and 6000 Hz. This progressively changes to a single peak in the 2000 to 4000 Hz region that can be seen in the average *A* from 6- to 24-month-old infants in Figures 1.4B and C. Aithal et al. (2014b) found this change occurs as early as 4 months of age (Figure 1.3A). Developmental changes in WAI are attributed to a range of factors. There is a decrease in mass and an increase in stiffness in the outer and middle ear as a result of

growth in length and diameter of the ear canal. The bony portion of the ear canal wall increases and the tympanic membrane decreases in thickness and changes in orientation. There is a loss of fluid and mesenchyme from the middle ear and aeration of the mastoid cavity. The volume of the middle ear and size of the temporal bone increases and there are changes in the bone density of the ossicles (Aithal et al., 2014b).

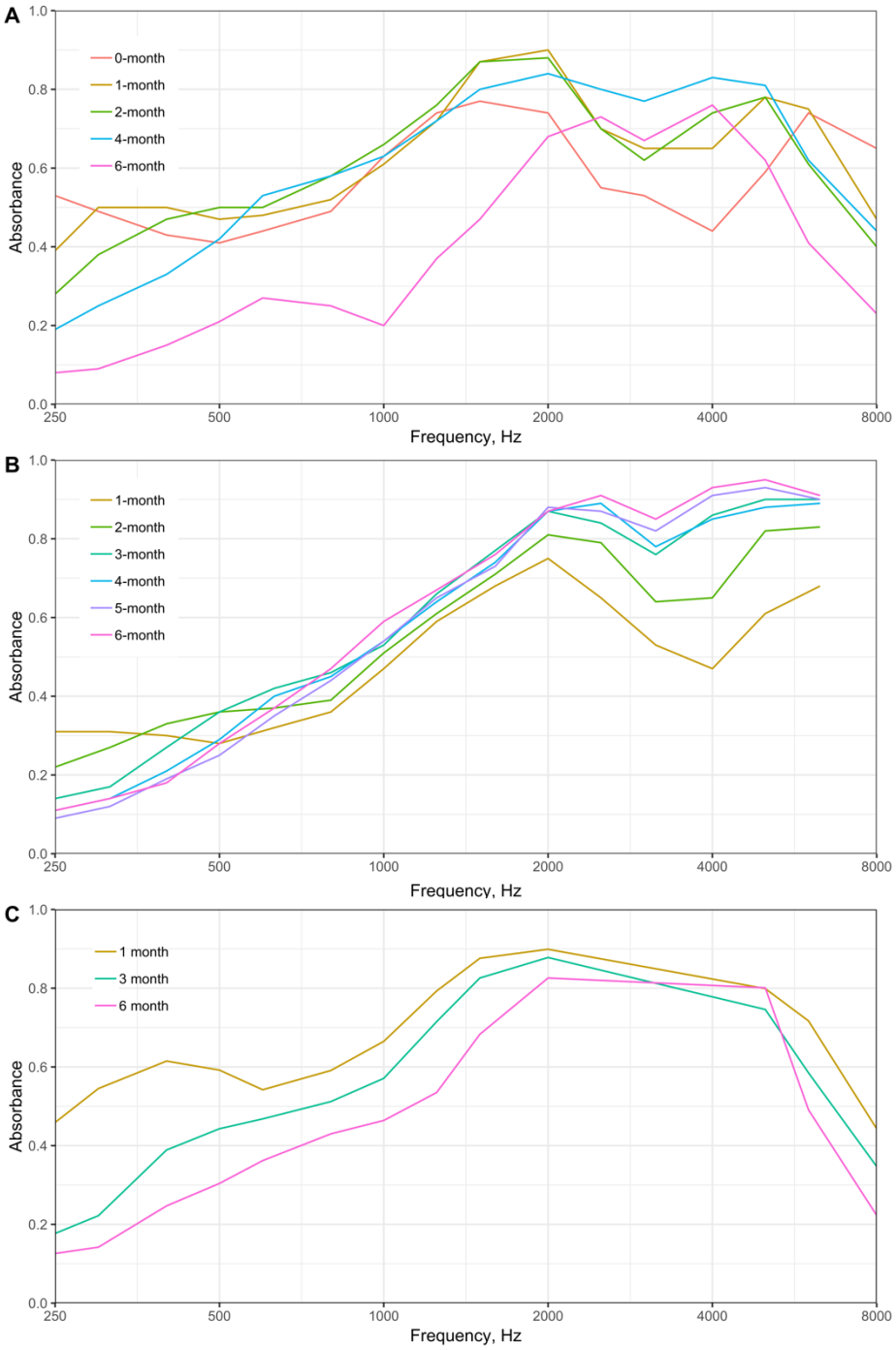


Figure 1-3. Comparison of average ambient absorbance from developmental studies over the first six months of life. **Panel A shows median absorbance reported in a cross-sectional study of infants 0 to 6 months old (Aithal et al., 2014b). Panel B, Median absorbance from a longitudinal study of 1- to 6-month-old infants (Shahnaz et al., 2014). C, Mean absorbance of 1- 3- and 6-month-old infants reported by Keefe et al. (1993). Shahnaz et al. and Keefe et al. measured reflectance which has been converted to absorbance for ease of comparison.**

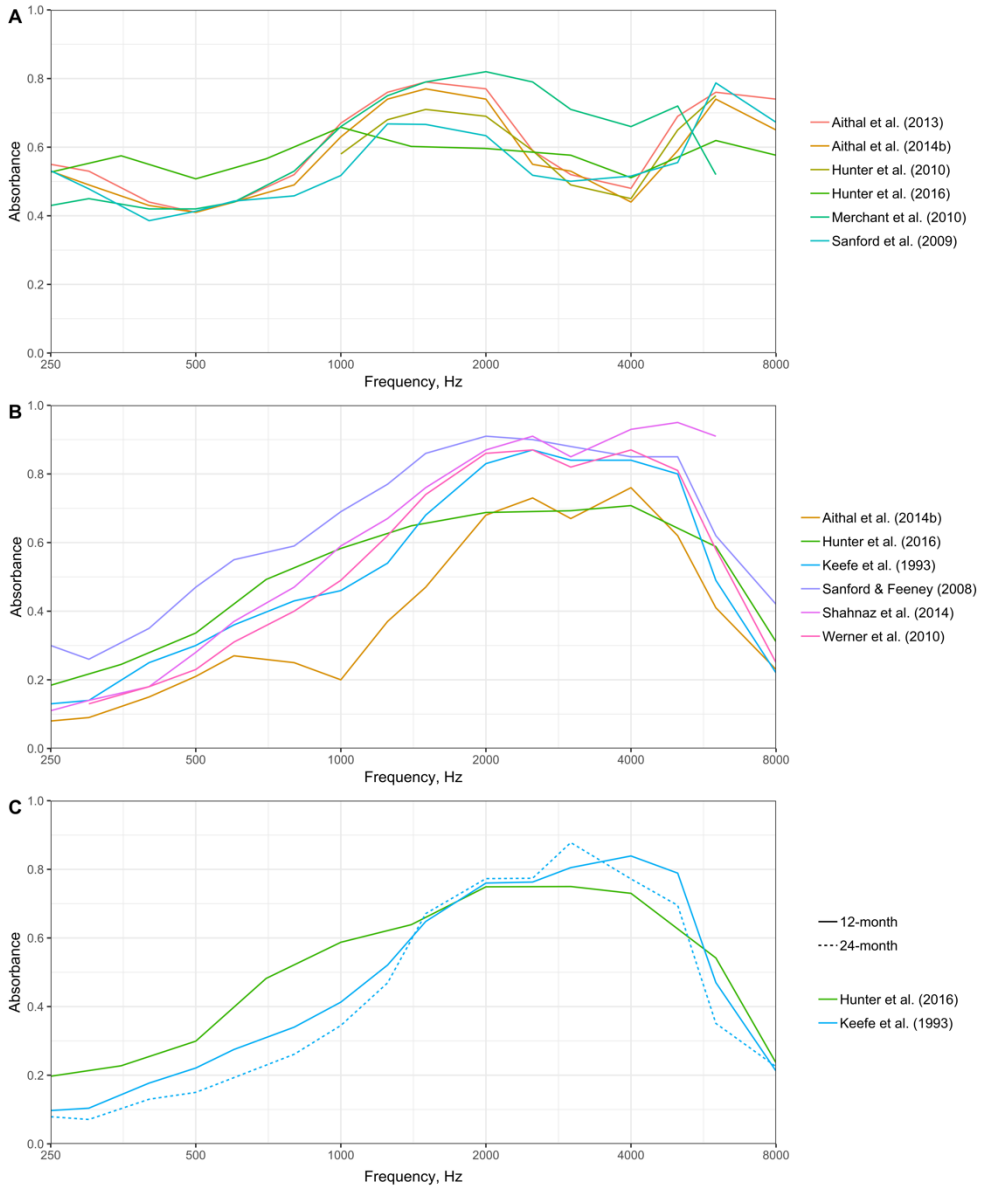


Figure 1-4. Comparison of average ambient absorbance from normative infant studies.
Panel A shows studies done in neonates; Panel B, 6-month-old infants; Panel C, 12- and 24-month-old infants. Aithal et al. (2014b), Aithal et al. (2013), Sanford et al. (2009) and Hunter et al. (2010) reported the median, all other studies the mean. Sanford et al. (2009) shows results from the first day of life. Studies that measured reflectance have been converted to absorbance for ease of comparison.

Because developmental changes significantly affect WAI measurements, it is important to use age-specific normative data and this is an area of ongoing investigation (Feeney et al., 2013; Kei et al., 2013; Shahnaz et al., 2013). There are currently normative data available for A for neonates and infants up to 6 months of age but normative data from other WAI measures such as $|Y|$ and φ_Y are yet to be established. Normative data for infants older than 9 months are limited and further research is needed to investigate the normative aspects of WAI in this population.

1.3.2.5 Diagnostic performance

The diagnostic accuracy of a test is how precisely it identifies the condition being tested for. Studies have consistently shown that *A* is reduced in ears with otitis media (Aithal et al., 2014a; Hunter, Bagger-Sjöbäck, & Lundberg, 2008a; Hunter et al., 2008b). Table 1.3 summarises diagnostic accuracy studies of WAI in infants and children. There have been four studies investigating the test performance of WAI for identifying otitis media or conductive hearing loss in children over two years old. Three used puretone audiometry as the reference standard (Keefe, Sanford, Ellison, Fitzpatrick, & Gorga, 2012; Keefe & Simmons, 2003; Piskorski, Keefe, Simmons, & Gorga, 1999), and one used a reference standard consisting of a combination of pneumatic otoscopy, otomicroscopy and puretone audiometry (Beers et al., 2010). Results of all studies demonstrated that *A* is reduced (or *R* increased) in diseased ears. Keefe et al. (2012) found *A* was reduced from 700 to 8000 Hz and Piskorski et al. (1999) identified 2000 to 4000 Hz as a diagnostically important region. Beers et al. (2010) found a systematic increase in *R* from normal, to Eustachian tube dysfunction, to middle ear effusion. WAI accurately identified middle ear pathology in all studies with an area under the receiver operating characteristic curve (AUC) of 0.87 to 0.99. Beers et al. (2010) reported that 1250 Hz most accurately identified diseased ears with a sensitivity of 0.96 and specificity 0.95 (measuring *R*). With sensitivity fixed at 0.90, Piskorski et al. (1999) reported a specificity of 0.94 (measuring *A* and *Y*). Keefe and Simmons (2003) found a sensitivity of 0.72 for *A* and 0.94 for A_p with specificity fixed at 0.9. The authors concluded that A_p outperformed *A* (AUC = 0.95 and 0.87 respectively), but Keefe et al. (2012) found that both *A* and A_p and performed with equally high accuracy (AUC = 0.98 and 0.99 respectively). Results of these studies have shown that WAI is more accurate than LFT in detecting otitis media with effusion and conductive hearing loss in children (Beers et al., 2010; Keefe et al., 2012). There is significant overlap of *A* between normal and diseased ears at low frequencies which could explain why LFT performs poorly in young infants (Ellison et al., 2012). WAI is at least as accurate as pneumatic otoscopy in diagnosing otitis media in children with the advantage that interpretation of results is less subjective than pneumatic otoscopy (Ellison et al., 2012).

Two studies investigating diagnostic accuracy of WAI have been carried out in infants older than three weeks (Ellison et al., 2012; Prieve et al., 2013b). Both have used strong reference standards but were limited by small sample size. Ellison et al. (2012) used myringotomy or pneumatic otoscopy as the gold standard for diagnosing otitis media and included 88 children 6 months to 7 years old (median = 1 year old). Prieve et al. (2013b) used an air-bone gap demonstrated by ABR to diagnose conductive hearing loss in 60 infants aged 3 to 36 weeks. Ellison et al. (2012) concluded that *A* in the 1500 to 3000

Hz region is important diagnostically and Prieve et al. (2013b) found *R* was higher in the 800 to 2500 Hz region in diseased ears (and also 6300 Hz). Both studies concluded that WAI accurately identifies conductive conditions in infants and children. Ellison et al. (2012) calculated AUC of 0.94 (95% CI, 0.85–0.96). Prieve et al. (2013b) reported positive likelihood ratio (LR) of >10 and negative LR of <0.2 at 1600 and 2000 Hz meaning that *R* at these frequencies can strongly predict the presence or absence CHL.

Five studies have investigated the accuracy of WAI in identifying conductive conditions in neonates (Aithal et al., 2015; Hunter et al., 2010; Keefe et al., 2003a; Keefe et al., 2003b; Sanford et al., 2009). The Keefe et al. (2003a) and Keefe et al. (2003b) studies were a part of the Identification of Neonatal Hearing Impairment study and used subjects drawn from the same database. Choice of reference standard is an issue when studying middle ear dysfunction in neonates. Myringotomy and medical imaging are the gold standards for diagnosis of middle ear effusion, but these tests are not ethical for studies of healthy neonates. Air- and bone-conduction ABR is the gold standard for diagnosis of conductive hearing loss but not middle ear effusion per se, which only causes a measurable air-bone gap approximately 50% of the time in infants and children (Fria et al., 1985). Pneumatic otoscopy is not feasible in neonates and LFT is inaccurate in infants under 6 months old. HFT is accurate in young infants but studies in neonates have shown poor sensitivity (Margolis et al., 2003; Swanepoel et al., 2007). EOAEs have been used as the reference standard in most diagnostic accuracy studies of neonates, however, the diagnostic accuracy of EOAEs in identifying middle ear effusion in neonates is unknown. Use of multiple tests in combination has been suggested to create a stronger reference standard for diagnostic accuracy studies in neonates. Aithal et al. (2015) reported that diagnostic accuracy of WAI improved using a battery of tests rather than a single test only.

Results of studies exploring the test performance of WAI in neonates have found AUC of 0.78 to 0.90. Aithal et al. (2015) concluded that the diseased group had lower *A* with best separation between groups in the 1000 to 2500 Hz region. Sanford et al. (2009) had similar findings with the best separation of *A* from 1400 to 2500 Hz, best separation of $|Y|$ from 1000 to 2000 Hz (with normal and diseased groups overlapping above 4000 Hz) and best separation of φ_Y from 750 to 1000 Hz and 2500 to 4500 Hz. Hunter et al. (2010) reported that ears who failed a DPOAE test had higher *R* than ears that passed, but there was significant overlap between the groups.

1.3.2.6 Predictive modelling

Some studies have used univariate analysis of WAI variables and others have used multivariate techniques (see Table 1.3, Statistical Analysis column). There are benefits and drawbacks to both

approaches. Univariate analysis predicts how well a single WAI frequency (or average across a range of frequencies) identifies middle ear dysfunction. Aithal et al. (2015), Beers et al. (2010), Hunter et al. (2010) and Prieve et al. (2013b) have all taken this approach. These studies have compared univariate results for different frequencies and concluded that WAI measured at frequencies between 1000 and 2000 Hz most accurately identify conductive conditions. Aithal et al. (2015) and Beers et al. (2010) reported that 1250 Hz is the most accurate frequency. Prieve et al. (2013b) found that WAI measured at 1600 and 2000 Hz were most accurate and Hunter et al. (2010) found that 2000 Hz performed the best. The advantage of using univariate analysis in research is that results can be immediately implemented clinically using current-generation equipment (Prieve et al., 2013b). Also, clinicians may more readily adapt univariate measures as they are used to performing similar analysis with clinical tests such as tympanometry and EOAEs. The limitation of univariate analysis is that it does not use all available data in predictions.

Multivariate analysis that combines the predictive power of multiple frequencies and multiple WAI variables is potentially more accurate (Prieve, Feeney, Stenfelt, & Shahnaz, 2013a). The challenge is in identifying the multivariate techniques that can best make use of the large amount of data produced by a single WAI test (Hunter, Prieve, Kei, & Sanford, 2013; Sanford & Brockett, 2014). Studies that have utilised multivariate techniques are outlined in Table 1.3 (Statistical Analysis column). Multivariate techniques have included log likelihood ratios, logistic regression and linear discriminant analysis. However, with the exception of the Identification of Neonatal Hearing Impairment studies (Keefe et al., 2003a; Keefe et al., 2003b), most of these studies have used many predictors in the model with only a small sample size (Ellison et al., 2012; Keefe et al., 2012; Keefe & Simmons, 2003; Piskorski et al., 1999; Sanford et al., 2009). They are therefore at risk of overfitting, which means that although the multivariate models accurately describe the data they were developed on, they may perform poorly in new samples (Steyerberg, 2008). Also, with the exception of the Identification of Neonatal Hearing Impairment studies, no studies have used any form of validation (Keefe et al., 2003a; Keefe et al., 2003b). Validation is a crucial aspect of multivariate modelling to verify that results will generalize well to new subjects. There are two types of validation, internal and external. Internal validation should be performed at the time of model development, and is used to assess the degree of "optimism", or overfitting in the model, that is, whether it is likely to generalize well to unseen data (Moons et al., 2012b). The Identification of Neonatal Hearing Impairment studies internally validated their models by developing the model on one ear of subjects and validating using the opposite ears. No WAI studies have externally validated a model. External validation in a new sample is recommended prior to implementing a model clinically, to provide a realistic idea of model

performance, since results are usually poorer when applied to new subjects (Moons et al., 2012a). This can be due to overfitting, or differences in subject characteristics, environment, or equipment used in the new setting (Steyerberg, 2008).

Multivariate analysis techniques that could be summarized into a single parameter would be valuable as this would incorporate the benefits of multivariate accuracy and univariate utility (Sanford & Brockett, 2014). Prediction models have this property and may provide a useful way of analysing WAI data. Prediction models take a multivariable input and provide a probability estimate (between 0 and 1) that the subject has the condition (Steyerberg, 2008). This type of model is useful because it provides individualised predictions and models disease severity. As well as accurately discriminating between subjects that have or do not have the condition, a prediction model needs to be well calibrated, meaning that predicted probabilities align closely with actual probabilities (Harrell, 2015). Predictions from a model are reliable if they align closely with observed frequencies of the condition. For example, for neonates with predicted probability of 0.3, approximately 3 out of 10 should actually have conductive dysfunction (Steyerberg et al., 2010).

Previous diagnostic WAI studies in infants and children have used a binary (pass/fail) outcome to assess test performance (Aithal et al., 2015; Beers et al., 2010; Ellison et al., 2012; Hunter et al., 2010; Keefe et al., 2003a; Keefe et al., 2012; Piskorski et al., 1999; Prieve et al., 2013b; Sanford et al., 2009). However, creating a dichotomous outcome from a disease that lies on a spectrum can lead to loss of information. Furthermore, there is growing evidence that WAI can detect mild pathology such as Eustachian tube dysfunction, as well as more severe conditions such as otitis media with effusion (Aithal, Aithal, Kei, Anderson, & Liebenberg, 2018; Beers et al., 2010; Ellison et al., 2012; Hunter et al., 2008b; Robinson, Thompson, & Allen, 2016; Shaver & Sun, 2013; Voss, Merchant, & Horton, 2012; Werner et al., 2010). Studies of infants and children have found a systematic decrease in A as the severity of middle ear disease increases (Beers et al., 2010; Ellison et al., 2012; Hunter et al., 2008b). This indicates that models using an ordinal outcome may be appropriate for WAI data. Ordinal models, where the outcome is an ordered scale (e.g., normal, mild, severe) can also be easier to interpret and have increased power compared to binary outcome models (Agresti, 2013).

Table 1.3. Comparison of diagnostic performance WAI studies in infants and children

Study	*Age*	*n*	*Population sampled*	*Ref*	*Sys*	*Variables*	f_{res}	*Statistical analysis*	*AUC*	*Conclusions*
Aithal et al. (2015)	Neonates (8 to 152 hr)	192 (BE)	Healthy babies	Test battery	RI	A	1/3	Univariate	0.78	Test battery recommended over a single test for gold standard.
Hunter et al. (2010)	Neonates (3 to 102 hr)	324 (BE)	Healthy babies	DPOAE	MA	R	Not stated	Univariate	0.9	R predicted DPOAE outcomes better than HFT. The most accurate predictors included 2000 Hz.
Keefe et al. (2003a)	Neonates	2638 ears (BE)	Well babies and NICU graduates (drawn from same sample as Keefe et al., 2000)	DPOAE, TEOAE, AABR, VRA	CM	R & G	1/2	Multivariate (logistic regression)	0.86	A WAI test of middle ear function generalised well (AUC = 0.81 to 0.84) to false positives that failed two-stage newborn hearing screen (AABR + TEOAE or DPOAE)
Keefe et al. (2003b)	Neonates	2766 ears (BE)	Well babies and NICU graduates (drawn from same sample as Keefe et al., 2000)	DPOAE & TEOAE	CM	R & G	1/2	Multivariate (logistic regression)	0.79	Results generalized well to new data as validated by training (n=1278) and test set (n=1147; AUC = 0.82).
Sanford et al. (2009)	Neonates (9 to 58 hr)	230 (BE)	Well babies	DPOAE	RI	A, A_t, $\lvert Y\rvert$ & φ_Y	1/12	Multivariate (log likelihood ratio)	0.87 (φ_Y, CI, 0.82-0.91)	WAI predicted DPOAE status better than HFT. Test performance was similar for ambient and pressurised WAI.
Prieve et al. (2013b)	Infants (3 to 36 wk)	60 (one ear)	Mostly infants who referred from NHS	TB ABR (AC & BC)	MA	R	1/3	Univariate	NA	R and HFT both accurately identified CHL.
Ellison et al. (2012)	Infants and children (0.5 to 7 yr)	88 (BE)	Case control study (TT group and normal group)	myringotomy or PO	RI	A, $\lvert Y\rvert$ & φ_Y	1/12	Multivariate (log likelihood ratio)	0.94 (CI, 0.88–0.98)	Multivariate analysis was more accurate. Highest accuracy was achieved by including all WAI variables.
Beers et al. (2010)	Children (3 to 12 yr)	142 (BE)	Normal group and OME group	PO and microscopy or PTA, tympanometry and TEOAE	MA	R	1/3	Univariate	0.97 (CI, 0.94-0.99)	R was more accurate than 226 Hz tympanometry.
Keefe et al. (2012)	Children (3 to 8 yr)	50 (BE)	Children with CHL and control group (age matched)	PTA (AC and BC)	RI	A & A_t	1/12	Multivariate (log likelihood ratio)	0.99	WAI predicted CHL better than tympanometry. Test performance of ambient and pressurised energy absorbance was the similar.
Keefe & Simmons (2003)	Children and adults (10 to 55 yr)	35 (BE)	Normal hearing and CHL	PTA (AC and BC)	CM	A & A_t	Not stated	Multivariate (likelihood moment analysis)	0.87 (A), 0.95 (A_t)	Pressurised absorbance outperformed ambient measurements.
Piskorski et al. (1999)	Children (2 to 10 yr)	92 (BE)	Normal group and CHL group	PTA (AC and BC)	CM	R, G & Eq Vol	1/3	Multivariate (logistic regression and DA)	0.96	Logistic regression outperformed DA, multivariate analysis was more accurate than univariate. Analysis using one-third-octave frequency resolution was not better than one-octave frequency resolution.

AUC is the highest area under the ROC curve reported in a study. BE, both ears; CHL, conductive hearing loss; DA, discriminant analysis; Eq Vol, Equivalent volume; f_{res}, frequency resolution; Ref, reference standard used; TT, tympanostomy tube; Otherwise as Table 1.2.

1.4 Rationale and aims for this thesis

Intervention for otitis media is vital early in infancy, as early onset is associated with recurrent and chronic conditions that can impact language and development. Early intervention is conditional on diagnosis but identifying otitis media in infants remains a significant challenge. The condition is often asymptomatic in this population, and currently available diagnostic tools are inaccurate and difficult to use in infants. WAI is an innovative, high-resolution test of middle ear function that is quick and easy to administer. Early studies in neonates have shown promising results (Aithal et al., 2015; Hunter et al., 2010; Keefe et al., 2003a; Keefe et al., 2003b; Sanford et al., 2009) but further large-scale studies using a stringent reference standard are needed. Little is known about the diagnostic performance outside the neonatal period as most studies have been done using newborns (Aithal et al., 2015; Hunter et al., 2010; Keefe et al., 2003a; Keefe et al., 2003b; Sanford et al., 2009). The only study with a sample consisting exclusively of infants outside of the newborn period has been Prieve et al. (2013b) who investigated diagnostic accuracy of WAI to detect conductive hearing loss in a sample of infants aged 3 weeks to 9 months. Ellison et al. (2012) included infants as young as 6 months of age in their study, but the sample included children up to 7 years of age as well. Further research is needed into diagnostic performance of WAI in detecting middle ear pathology in infants outside the neonatal period.

A single WAI test generates a large amount of data and research investigating the most effective ways of analysing and presenting these data is in its infancy (Hunter et al., 2013). It is thought that multivariate analysis techniques have an advantage over univariate methods, but studies using univariate methodology in infants have found high predictive accuracy with AUC of up to 0.90 which is as high or higher than some multivariate studies (Hunter et al., 2010; Prieve et al., 2013a). Furthermore, many multivariate studies are likely to be overfitting the data as they have used many predictors, small sample sizes, and no validation (Ellison et al., 2012; Keefe et al., 2012; Sanford et al., 2009). Overfitting means that the model has been fit of the idiosyncrasies of the data rather than the relationship that is generalizable to new subjects. It is a serious issue, as an overfitted model is likely to perform poorly in new observations meaning which would result in a high proportion of misdiagnoses if adopted clinically (Steyerberg, 2008). Prediction models may be a useful type of multivariate model for analysis of WAI data as they provide a clinically useful univariate summary in the form of a probability estimate. Models that use an ordinal outcome may be appropriate for WAI data, since previous research has found that *A* systematically decreases as the severity of middle ear disease increases (Beers et al., 2010; Ellison et al., 2012; Hunter et al., 2008b). It is vital that prediction models include internal

validation as part of development to assess for overfitting (Moons et al., 2012b). No predictive WAI models have been externally validated in a new sample, and research is needed in this area also.

An issue with developing predictive models for use with infants, however, is that there are substantial maturational effects on WAI that need to be controlled for (Aithal et al., 2014b; Hunter, Keefe, Feeney, Fitzpatrick, & Lin, 2016; Keefe et al., 1993; Shahnaz et al., 2014; Werner et al., 2010). One solution to this problem has been to create age-specific models, such as models that have been developed specifically for neonates (Keefe et al., 2003a; Sanford et al., 2009). Alternatively, age could be controlled for by using regions of the WAI response that are relatively unaffected by age as predictors in a model, or by including an interaction between age and WAI predictors in a model, which would allow interpretation of results to vary with age (Sanford & Feeney, 2008). Knowledge of maturational effects on WAI is needed to implement such strategies, however. Developmental effects over the first year of life on A and R have been investigated (Aithal et al., 2014b; Aithal et al., 2013; Hunter et al., 2018; Hunter et al., 2010; Merchant et al., 2010; Sanford & Feeney, 2008; Sanford et al., 2009; Shahnaz et al., 2014; Werner et al., 2010), but little is known about age effects over the second year of life. Keefe et al. (1993) measured R in 12- and 24-month-old infants but did not use any reference standard to assess ear status. There is a dearth of evidence about the effect of age for other WAI measures such as $|Y|$ or φ_Y through infancy.

The overall aim of the thesis was to investigate the diagnostic performance of WAI in infancy by developing predictive models. Specifically, the research aimed to:

1. Develop predictive models for specific infant age groups: Neonates, 6 months, and 12 months with appropriate internal validation (Chapters 2, 3 and 4, respectively).
2. Investigate strategies for modelling WAI data (Chapters 2, 3, 4 and 6), including:
    a. Comparing univariate to multivariate modelling approaches (Chapters 2, 3 and 4).
    b. Approaches to reducing the large volume of WAI data, such as frequency averaging, predictor selection, and principal component analysis (Chapters 2, 3 and 4).
    c. Allowing WAI predictors to have a nonlinear association with the outcome (Chapters 2, 4 and 6).
    d. Using an ordinal outcome in predictive WAI models (Chapters 4 and 6).
    e. Whether including demographics such as ethnicity and ear side improves diagnostic performance of models (Chapters 2, 3 and 4).
3. Investigate developmental effects of WAI through infancy, and establish normative data for various age groups (Chapter 5).

4. Development a model for use in a broader age range through infancy (6- to 18-months) controlling for developmental effects on WAI (Chapter 6, Study 1).
5. Externally validate the model developed for neonates (Chapter 2) in a new sample to assess generalizability to new subjects (Chapter 6, Study 2).

1.5 Appendix: Pressure reflectance in terms of admittance

Steps for writing the equation for pressure reflectance (*PR*) in terms of admittance (*Y*). The equation for *PR* in terms of impedance (*Z*) is:

$$PR = \frac{Z - Z_c}{Z + Z_c}. \tag{A.1}$$

First, replace *Z* with $1/Y$ (i.e. rewrite in terms of admittance):

$$PR = \frac{\frac{1}{Y} - Z_c}{\frac{1}{Y} + Z_c}. \tag{A.2}$$

Then, put the numerator and denominator terms over a common denominator:

$$PR = \frac{\frac{1}{Y} - \frac{Z_c Y}{Y}}{\frac{1}{Y} + \frac{Z_c Y}{Y}}. \tag{A.3}$$

And simplify:

$$PR = \frac{\frac{1 - Z_c Y}{Y}}{\frac{1 + Z_c Y}{Y}}. \tag{A.4}$$

Which equals:

$$PR = \frac{Y}{Y} \times \frac{1 - Z_c Y}{1 + Z_c Y} = \frac{1 - Z_c Y}{1 + Z_c Y}. \tag{A.5}$$

Chapter 2. Development of a Diagnostic Prediction Model for Conductive Conditions in Neonates Using Wideband Acoustic Immittance

This chapter develops a prediction model for diagnosing conductive conditions using wideband acoustic immittance in neonates. It has been previously published in the article: Myers, J., Kei, J., Aithal, S., Aithal, V., Driscoll, C., Khan, A., Manuel, A., Joseph, A., Malicka, A. N. (2018a). Development of a diagnostic prediction model for conductive conditions in neonates using wideband acoustic immittance. *Ear and Hearing, 39*(6), 1116-1135.

I made substantive contributions to the article in the areas of study design, data collection, data analysis and drafting of the article, as outlined below:

Contributor	**Statement of contribution**
Joshua Myers (Candidate)	Study design (60%) Recruitment and data collection (60%) Data analysis (100%) Wrote the article (100%)
Joseph Kei	Study design (20%) Edited the article (40%)
Sreedevi Aithal	Study design (5%) Edited the article (15%)
Venkatesh Aithal	Study design (5%)
Carlie Driscoll	Study design (5%) Edited the article (15%)
Asaduzzaman Khan	Study design (5%) Edited the article (15%)
Alehandrea Manuel	Recruitment and data collection (20%)
Anjali Joseph	Recruitment and data collection (20%)
Alicja N. Malicka	Edited the article (15%)

2.1 Abstract

Purpose: Wideband acoustic immittance (WAI) is an emerging test of middle ear function with potential applications for neonates in screening and diagnostic settings. Previous large-scale diagnostic accuracy studies have assessed the performance of WAI against evoked otoacoustic emissions, but further research is needed using a more stringent reference standard. Research into suitable quantitative techniques to analyse the large volume of data produced by WAI is still in its infancy. Prediction models are an attractive method for analysis of multivariate data as they provide individualized probabilities that a subject has the condition. A clinically useful prediction model must accurately discriminate between normal and abnormal cases, and be well calibrated (i.e., give accurate predictions). The present study aimed to develop a diagnostic prediction model for detecting conductive conditions in neonates using WAI. A stringent reference standard was created by combining results of high-frequency tympanometry (HFT) and distortion product otoacoustic emissions (DPOAEs).

Methods: HFT and DPOAEs were performed on both ears of 629 healthy neonates to assess outer and middle ear function. Wideband absorbance and complex admittance (magnitude and phase) were measured at frequencies ranging from 226 to 8000 Hz in each neonate at ambient pressure using a click stimulus. Results from one ear of each neonate were used to develop the prediction model. WAI results were used as logistic regression predictors to model the probability that an ear had outer/middle ear dysfunction. WAI variables were modelled both linearly and nonlinearly, to test whether allowing nonlinearity improved model fit, and thus calibration. The best fitting model was validated using the opposite ears, and with bootstrap resampling.

Results: The best fitting model used absorbance at 1000 and 2000 Hz, admittance magnitude at 1000 and 2000 Hz, and admittance phase at 1000 and 4000 Hz modelled as nonlinear variables. The model accurately discriminated between normal and abnormal ears, with an area under the receiver operating characteristic curve (AUC) of 0.88. It effectively generalized to the opposite ears (AUC = 0.90), and with bootstrap resampling (AUC = 0.85). The model was well calibrated, with predicted probabilities aligning closely to observed results.

Conclusions: The developed prediction model accurately discriminated between normal and dysfunctional ears, and was well-calibrated. The model has potential applications in screening or diagnostic contexts. In a screening context, probabilities could be used to set a referral threshold that is intuitive, easy to apply, and sensitive to the costs associated with true and false positive referrals. In a clinical setting, using predicted probabilities in conjunction with graphical displays of WAI could be used for

individualized diagnoses. Future research investigating the use of the model in diagnostic or screening settings is warranted.

2.2 Introduction

Evaluation of the conductive pathway is an important aspect of neonatal hearing assessment. It is vital in diagnostic settings to aid in diagnosing the type of hearing loss (conductive, sensorineural, or mixed), and would also be valuable in the context of newborn hearing screening (Sanford et al., 2009). The goal of hearing screening at birth is to identify permanent hearing loss, but the majority of referrals (90%) occur due to transient conductive conditions of the outer/middle ear (Merchant et al., 2010; Thompson et al., 2001). Incorporating a test of middle ear function into the screening process could reduce the false positive rate for detecting permanent hearing loss. It could also facilitate improved management of neonates with middle ear dysfunction, who have higher risk of chronic otitis media in the first year of life (Doyle, Kong, Strobel, Dallaire, & Ray, 2004; Marchant et al., 1984b; Pereira, Azevedo, & Testa, 2010).

However, diagnosis of middle ear pathology in neonates is challenging, as conventional instruments are ineffective in this population (Aithal et al., 2013). Traditional 226-Hz tympanometry is inaccurate in infants under 7 months of age (Paradise et al., 1976), and using a higher frequency (1000 Hz) probe tone has been recommended in this age group (Baldwin, 2006; Zhiqi et al., 2010). Prieve et al. (2013a) found that high-frequency tympanometry (HFT) accurately identified conductive hearing loss diagnosed with air- and bone-conduction auditory brainstem response (ABR) in a sample of infants aged 3 to 25 weeks. However, studies investigating the performance of HFT to detect middle ear dysfunction in neonates have shown good specificity (0.91 to 0.95), but poor sensitivity (0.36 to 0.57) (Margolis et al., 2003; Sanford et al., 2009; Swanepoel et al., 2007). These results indicate that HFT may accurately diagnose conductive pathology severe enough to cause conductive hearing loss, but could be insensitive to milder forms of outer and middle ear dysfunction in neonates such as partial occlusion of the ear canal with vernix, or incomplete filling of the middle ear cavity with fluid/debris (Northrop, Piza, & Eavey, 1986; Palmu & Syrjänen, 2005; Pitaro, Al Masaoudi, Motallebzadeh, Funnell, & Daniel, 2016).

Wideband acoustic immittance (WAI) is an emerging tool for middle ear assessment with potential applications in neonates in both screening and diagnostic settings (Aithal et al., 2015; Hunter et al., 2010; Sanford et al., 2009). WAI is an umbrella term encompassing a family of wideband acoustic transfer functions including pressure reflectance (PR), energy reflectance (R), energy absorbance (A), and acoustic admittance (Y). PR is the ratio of an incident to reflected acoustic pressure wave. R is $|PR|^2$ and represents the proportion of incident energy that is reflected back from the middle ear. A is $1 - R$, representing the proportion of energy that is absorbed by the conductive pathway (Rosowski,

Stenfelt, & Lilly, 2013). Y is the ratio of volume velocity to acoustic pressure, and is a complex measurement with magnitude ($|Y|$) and phase (φ_Y) (Keefe & Levi, 1996). Age-specific normative data are needed for newborns and young infants, as the WAI response is affected by developmental changes in the outer and middle ear through infancy (Aithal et al., 2014b; Keefe et al., 1993; Kei et al., 2013; Shahnaz et al., 2014). In newborns, changes in A and R have been reported over the first few days of life, attributed to the clearing up of fluid and debris in the outer and middle ear after birth (Hunter et al., 2010; Keefe et al., 2000; Sanford et al., 2009).

Diagnostic studies in neonates have found that WAI accurately discriminates between normal, and ears with conductive dysfunction, with area under the receiver operating characteristic curves (AUC) of 0.78 to 0.90 (Aithal et al., 2015; Hunter et al., 2010; Keefe et al., 2003a; Keefe et al., 2003b; Prieve et al., 2013b; Sanford et al., 2009). Aithal et al. (2015) used various combinations of evoked otoacoustic emissions (EOAEs), HFT, and automated ABR as reference standards. They found that ears with outer/middle ear dysfunction had reduced A compared to normal ears, with best separation between the groups in the frequency range from 1000 to 2500 Hz. Sanford et al. (2009) drew similar conclusions, using distortion product otoacoustic emissions (DPOAEs) as the reference test in a sample of 1-day-old neonates. They reported that A best separated between normal and abnormal ears from 1400 to 2500 Hz, $|Y|$ between 1000 and 2000 Hz, and φ_Y in two regions: from 750 to 1000 Hz, and 2500 to 4500 Hz. Univariate analyses of A and R have found the frequency of highest accuracy to be between 1000 and 2000 Hz. Aithal et al. (2015) reported 1250 Hz as the best discriminating frequency and Hunter et al. (2010) found that 2000 Hz best predicted DPOAE status. Less is known about the most important predictive frequencies for $|Y|$ and φ_Y. For A and R, the best discriminating predictors occurred in the frequency range from 1000 to 2500 Hz, where separation was greatest between normal and abnormal ears, and this is likely to be true for $|Y|$ and φ_Y as well.

Large-scale diagnostic studies of WAI, to date, have used EOAEs as the reference test to determine outer and middle ear status (Hunter et al., 2010; Keefe et al., 2003a; Keefe et al., 2003b; Sanford et al., 2009). EOAEs are not a test of middle ear function per se, but the stimuli and response need to travel through the outer and middle ear, and significant dysfunction is thought to interfere with this process. A limitation of using EOAEs as a reference test for middle ear function is that they are affected by cochlear status, and may be reduced or absent due to sensory, as well as conductive pathology. Furthermore, passing an EOAE test does not completely rule out conductive problems, as EOAEs have been recorded in ears of neonates, infants and children with known middle ear dysfunction (Aithal et al., 2015; Amedee, 1995; Doyle et al., 1997; Driscoll et al., 2001; Margolis et al., 2003). In a sample of

200 neonates, Doyle et al. (1997) found that 38% of 32 ears judged to have restricted tympanic membrane mobility on pneumatic otoscopy passed a transient evoked otoacoustic emissions (TEOAE) test (pass = 50% reproducibility). In a study of 30 children aged 1 to 7 years (mean = 3 years), Amedee (1995) reported present TEOAEs (pass = 5 dB signal-to-noise ratio [SNR], and 70% reproducibility) in 50% of ears with middle ear effusion diagnosed by myringotomy. Studies investigating the test performance of TEOAEs for detecting middle ear dysfunction in children have reported higher specificity than sensitivity using typical clinical diagnostic criteria. Driscoll et al. (2001) reported on the diagnostic accuracy of TEOAEs in a sample of 940 children aged 5 to 7 years. Test performance was measured against a reference standard consisting of puretone audiometry and 226-Hz tympanometry. Sensitivity increased (from 0.51 to 0.73), and specificity decreased (from 0.93 to 0.74) as the TEOAE diagnostic threshold increased from 3 to 9 dB SNR. A diagnostic criterion of 6 dB SNR resulted in specificity of 0.86 and sensitivity of 0.60. Koike and Wetmore (1999) assessed diagnostic accuracy of TEOAEs in 63 children aged 4 to 17 years using puretone audiometry and 226-Hz tympanometry as the reference standard. A TEOAE diagnostic criterion of 50% reproducibility resulted in specificity of 0.94 and sensitivity of 0.84. These results all indicate that passing an EOAE test using typical clinical diagnostic criteria (where a response is judged to be either present or absent), results in a test of middle ear function with high specificity (most normal ears are classified correctly), but lower sensitivity, some abnormal ears are incorrectly diagnosed as normal (i.e., the test misses some cases).

Therefore, further diagnostic WAI research using more stringent reference standards has been recommended (Hunter et al., 2013). However, there is no gold standard available to determine the presence or absence of middle ear effusion in neonates. Use of myringotomy or medical imaging in a study of otherwise healthy babies would be unethical, and the shortcomings of available clinical tests limit their use as reference tests (Aithal et al., 2013). Some researchers recommend combining results from multiple tests, such as DPOAEs and HFT, to create a more accurate reference standard (Aithal et al., 2014a; Aithal et al., 2015). A two-test reference standard can classify an ear as abnormal if it fails both tests, the “both fail criterion”, or either test, the “either fail criterion” (Naaktgeboren et al., 2013; Pepe, 2003). The characteristics of the component tests can be used to decide which criterion to adopt. Tests with high specificity can be combined using the either fail criterion to retain high specificity while improving sensitivity (Alonzo & Pepe, 1999). This is because tests with high specificity correctly classify most normal ears as normal. Therefore, ears that fail the test are likely to have the condition, as the test is unlikely to mislabel normal ears as abnormal. A diagnostic test with poor sensitivity, however, may often misclassify subjects with the condition as normal. A test that has high specificity but lower sensitivity will have few false positives, but may have many false negatives. Combining tests

with high specificity and lower sensitivity using the either fail criterion can reduce the number of false negatives, as a subject who fails one of the tests is likely to have the condition (i.e., be a true positive). One test may identify some true positives that the other test missed, and vice versa.

Both univariate and multivariate statistical methodologies have been used to analyse the diagnostic performance of WAI in previous research. Univariate analyses aim to find the most accurate diagnostic frequency in the response. The strength of this approach is that results are easy to interpret. Findings may therefore be readily adopted by clinicians, who are familiar with using results from similar research with current clinical tests such as tympanometry and EOAEs (Prieve et al., 2013b).

Multivariate analyses, however, incorporating multiple frequencies and/or WAI measures, could potentially use the additional information to develop a better performing statistical model (Prieve et al., 2013a). Studies using WAI to identify conductive hearing loss in children have reported multivariate models to be more accurate than univariate (Keefe et al., 2012; Piskorski et al., 1999). Similarly, Ellison et al. (2012) found that adding information from $|Y|$ and φ_Y, along with A, improved accuracy in detecting middle ear dysfunction in children. However, because of the number of variables in the response, overfitting can be an issue with multivariate modelling of WAI (Piskorski et al., 1999). A model using many variables may accurately describe the data it was developed on, but is likely to perform poorly when applied to new samples (Steyerberg, 2008). As there are potentially hundreds of predictors, some form of data reduction or penalization is therefore necessary when modelling WAI.

It is generally preferable for data reduction methods to be masked to the outcome. This avoids selecting predictors based on the idiosyncrasies of the data, which could affect the generalizability of a multivariate model (Harrell, 2015). Diagnostic WAI studies from the Identification of Neonatal Hearing Impairment project used principal component analysis to reduce the response to 5–7 factors (Keefe et al., 2003a; Keefe et al., 2003b). An alternative approach is to decrease the frequency resolution. Previous studies in neonates have used one-twelfth, one-third, or one-half octave frequency resolution (e.g., Aithal et al., 2015; Keefe et al., 2003a; Sanford et al., 2009). However, the frequency resolution need not be this fine, as studies have found high accuracy averaging A or R over single (Piskorski et al., 1999), and even multiple octaves (Aithal et al., 2015; Hunter et al., 2010; Keefe et al., 2012). The resolution should be fine enough to capture important diagnostic information, but broad enough that the number of predictors in the model is not excessive (Keefe, Hunter, Feeney, & Fitzpatrick, 2015). Another way of achieving data reduction masked to the outcome is to select predictors based on expert opinion or previous research (Harrell, 2015). However, it can also be beneficial to take an iterative approach to data reduction, by selecting predictors based on relationships with the outcome in the

dataset at hand. This may be a useful in a new field of research where knowledge of the most important predictors is still developing (Bagherzadeh-Khiabani et al., 2016).

Validation of a multivariate model is essential to assess how well it will generalize to new subjects, as performance is generally poorer in new samples (Steyerberg et al., 2001c). This can be due to overfitting (bias) in the model or differences in subject characteristics, environmental factors, or equipment used in the new setting (Steyerberg, 2008). Multivariate models therefore need to be validated both internally, to assess overfitting, and externally, to assess generalizability. Internal validation is performed during model development, using the same subjects that the model was developed on (Moons et al., 2012b). External validation assesses model performance in a new sample, that was not used in model development (Moons et al., 2012a).

The authors of the Identification of Neonatal Hearing Impairment studies internally validated their models using the opposite ears of subjects. They found that the models validated well with a difference in AUC between development and validation samples of 0.03 to 0.04 (Keefe et al., 2003a; Keefe et al., 2003b). An alternative approach to internal model validation is bootstrap resampling, which estimates the degree of overfitting in the model to provide a bias-corrected estimate of model performance (Harrell, Lee, & Mark, 1996).

Research into the most effective multivariate methods for analysis of WAI data is still in its infancy (Hunter et al., 2013). Clinically, a multivariate model that could be summarized into a single parameter would be valuable as it could incorporate the benefits of both multivariate accuracy and univariate clinical utility (Sanford & Brockett, 2014). Prediction models have this property as they convert multiple predictors into the probability that a subject has the condition. Probability of >0.5 indicates that the subject is more likely to have the condition than not. Prediction models are particularly suited to conditions that occur on a spectrum, such as outer/middle ear dysfunction, where the diagnostic threshold is somewhat arbitrary (Northrop et al., 1986; Palmu & Syrjänen, 2005; Vickers, Basch, & Kattan, 2008). Previous multivariate WAI research has focused on classification, correctly labelling a case as normal or abnormal (Keefe et al., 2012; e.g., Sanford et al., 2009). However, probabilistic risk modelling could provide additional information about diagnostic certainty. This may be particularly useful to help clinical decision making in borderline, or difficult-to-diagnose cases (Keefe, Fitzpatrick, Liu, Sanford, & Gorga, 2010; Keefe et al., 2012). Presenting an interpretable numerical parameter such as probability, provides the clinician with extra information, rather than just a pass/fail result. For example, if the diagnostic criterion was probability >0.5, it would be useful for the clinician to know that one subject failed the test with probability of 0.55, where another one failed with probability of 0.95. Both failed the test, but the diagnosis is more certain in the latter case.

A clinically useful prediction model must accurately discriminate between normal and abnormal cases, and have satisfactory calibration, that is, high agreement between predictions and observed results. (Steyerberg et al., 2010). Standard regression assumes that the relationship between continuous predictors and the reference test changes in a linear fashion over the entire range of predictor values. However, truly linear relationships are rare in biological data, and allowing continuous predictors to have a nonlinear association with the outcome often improves model fit, and thus, calibration (Harrell, 2015). Including covariates that are strongly associated with the outcome can also help to improve model fit (Harrell, 2015).

The purpose of the present study was to develop a diagnostic prediction model for detecting conductive conditions in neonates using WAI. A stringent reference standard was created by combining HFT and DPOAE results using the either fail criterion. With the aim of developing a model that would generalize to new subjects, the number of predictors was limited with respect to the sample size. One ear of each neonate was used to create the prediction model. WAI variables were modelled both linearly and nonlinearly, to test whether allowing predictors to have a nonlinear association with the outcome improved model fit. The best fitting model was validated internally using the opposite ears, and with bootstrap resampling.

2.3 Methods

The study was approved by the Townsville Health Service District Institutional Ethics Committee (reference number: HREC/09/QTHS/30) and the University of Queensland Behavioural and Social Science Ethical Review Committee (reference number: 2010000842). Neonates were recruited from the Maternity Ward of the Townsville Hospital from July 2014 to August 2016. Parents were informed of the study by the neonatal hearing screening nurses and written consent was obtained from those who wished to participate. The initial newborn hearing screen (automated ABR) was performed before testing. Not all neonates born at the hospital were eligible to participate as data collection was limited to days of the week that a research audiologist was available to do the testing. Neonates who failed a second newborn hearing screen were followed up at the audiology department within 6 weeks for audiological assessment, including air- and bone-conduction ABR, to determine the type and degree of hearing loss.

2.3.1 Subjects and test environment

Six hundred and twenty-nine healthy neonates were recruited to the study (295 females and 334 males). Six neonates were excluded on the basis of age and hearing status. Five were over one week old, and

were not included in the study, since the normative aspects of WAI are known to change over the first month of life (Aithal et al., 2014b; Shahnaz et al., 2014). These neonates were excluded from the study as they may have belonged to a different population, but were too few in number to quantify this potential source of variability. One neonate with confirmed sensorineural hearing loss was also excluded because one of the component tests of the reference standard was a cochlear response. Therefore, this subject may have failed the DPOAE test, even with normal outer and middle ear function. Table 2.1 shows characteristics of subjects included in the study: age, gender, ethnicity, gestational age at birth, birth type, birth weight, head circumference at birth, and birth length. Most babies were tested as inpatients, but were occasionally tested as outpatients if they were discharged from hospital before testing could be completed.

Table 2.1. Characteristics of neonates included in the study

Characteristic	Value
Age (hours)	3 missing
Median (IQR)	42 (28 to 52)
Range	8 to 163
Gender (count)	Missing 0
Female (%)	291 (47)
Male (%)	332 (53)
Ethnicity (count)	2 missing
Caucasian (%)	537 (86)
Asian (%)	52 (8)
Oceanian (%)	22 (3)
South American (%)	6 (1)
African (%)	4 (1)
Gestational age (weeks)	4 missing
Median (IQR)	39.2 (38.4 to 40.1)
Range	30.4 to 41.6
Birth type (count)	1 missing
Vaginal (%)	375 (60)
C-section (%)	247 (40)
Birth weight (grams)	4 missing
Median (IQR)	3460 (3150 to 3750)
Range	2170 to 5120
Head circumference (cm)	7 missing
Median (IQR)	35 (34 to 36)
Range	24.8 to 39
Birth length (cm)	7 missing
Median (IQR)	50 (49 to 52)
Range	37 to 58.5

The number of subjects with missing data for each characteristic is provided in the Value column. IQR, interquartile range.

All measurements were made by a research audiologist either at bedside, or in a quiet room in the Maternity Ward. Neonates were tested either in their crib or being held by a parent. Ambient noise

levels in the test environment ranged from 26 to 54 dBA (median = 38 dBA; IQR [interquartile range] = 34 to 41 dBA).

2.3.2 Test procedure

Measurements were made using an Interacoustics Titan device controlled by a windows laptop computer running Titan Suite software (version 3.2, Middelfart, Denmark). The equipment was calibrated annually by the manufacturer and probe function was checked daily in a 2 cm^3 cavity and with a biological test. Both ears of each neonate were tested, with the most accessible ear tested first. A suitably sized, soft plastic tip was attached to the probe, and DPOAEs, HFT and WAI were tested, in no particular order. All efforts were made to test neonates in a settled condition. If a baby was unsettled, testing was postponed if time permitted.

DPOAEs were measured using pairs of primary tones f_1 and f_2 at f_2 = 2000, 3000, 4000, and 6000 Hz. The f_2 / f_1 ratio was 1.22 and the intensity levels of f_1 and f_2 were 65 and 55 dB SPL, respectively. Emissions were considered present at a given frequency if the SNR was ≥6 dB with DPOAE level ≥–10 dB (Hunter et al., 2010). An ear was classified as normal if DPOAEs were present at three out of four f_2 frequencies, otherwise as abnormal (Aithal et al., 2015; Sanford et al., 2009). If an abnormal result was obtained with noise levels ≥0 dB SPL at two or more f_2 frequencies, the test was repeated to avoid the impact of physiological or environmental noise on the test outcome (Hunter et al., 2010).

HFT was performed using a 1000 Hz probe tone with an intensity of 85 dB SPL regulated by automatic gain control. Pressure was swept from +200 to –300 daPa at a speed of approximately 300 daPa/s at the tails slowing down to 100 daPa/s around the peak of the tympanogram. Measurements were repeated if the trace was difficult to interpret due to artefact caused by a restless or noisy neonate. Tympanogram traces were classified using a method similar to Baldwin (2006). A baseline was drawn by hand on hard copy between the points at the positive (+200 daPa) and negative (−300 daPa) extremes of the trace. A trace was classified as normal if there was a peak above the baseline, otherwise abnormal. Baldwin (2006) had a third "indeterminate" category, where there was no clear peak above, or trough below the baseline. These traces were classified as abnormal in the present study (Prieve et al., 2013b).

Ambient A, $|Y|$, and φ_Y were measured at 1/24 octave frequency resolution in response to a 226 to 8000 Hz broadband stimulus delivered at 96 dB peSPL. During a test, a series of 32 clicks was presented to the ear and averaged, to minimize effects of artefact caused by physiological noise. The Titan system automatically checked for air leaks according to the criterion detailed by Sanford et al. (2009),

and did not allow testing to proceed if an air leak was determined to be present. As a secondary check for air-leaks, a graphical display of *A* was also manually monitored during testing and the test was aborted and the probe reinserted if *A* was ≥0.7 at low frequencies (<500 Hz, Aithal et al., 2015; Keefe et al., 2000). WAI values were obtained using the Titan's Research Module which saves the data as a text file after each test. The WAI response was averaged into octave bandwidths for statistical modelling, which reduced the number of frequencies for each WAI measure from 107 to six.

All tests were performed with a single insertion of the probe, if possible. Reinsertion of the probe was required if a hermetic seal was not obtained for HFT, WAI results indicated an air leak, or the probe was dislodged.

2.3.3 Reference standard pass/fail criterion

DPOAEs and HFT were used as the component tests of the reference standard to classify the outer/middle ear as normal or abnormal. It was not considered feasible to include otoscopy in the test battery as there was no paediatric otologist available to perform the test. As such, ears that failed the reference standard were diagnosed with a "conductive condition", as it was not possible to differentiate between conditions of the outer and middle ear on the basis of the tests conducted (Sanford et al., 2009).

As discussed in the Introduction, available data indicate that both HFT and EOAEs have higher specificity than sensitivity for detecting middle ear dysfunction. As such, these tests will classify most normal ears correctly (true negatives), and ears that fail the test are likely to have the condition (true positives). There may be ears with conductive dysfunction, however, that pass the test (false negatives), as the test is not sensitive enough to detect these, perhaps, milder forms of dysfunction. Therefore, the either fail criterion was chosen to classify ears as "pass" or "fail", to create a reference standard with improved sensitivity (Alonzo & Pepe, 1999). Ears were classified as pass if they passed both HFT and DPOAEs and fail if they failed either test. This is referred to as the "a priori" reference standard throughout this report.

2.3.4 Missing data and statistical analyses

Of the 623 neonates included in the study, 44 ears were removed from the dataset due to either missing reference standard, or WAI results. Ears that failed either HFT or DPOAEs with missing results for the other reference test were retained as an observation was classified as fail on the basis of failing only one test. Table 2.2 provides the number of ears with missing data for each test and reasons for absence. Ears with missing data were due to the baby being unsettled, inability to obtain a hermetic seal, or

technical issues. These were all considered to be "missing at random", and since they made up only a small fraction of total observations (0.039), it was considered appropriate to remove them from the dataset for subsequent analyses. The dataset after removal of excluded neonates and observations with missing reference standard/WAI results is referred to as the "full sample" throughout this report. There were also some neonates with missing characteristics data (see Table 2.1). In the full sample (after removal of observations with missing reference standard/WAI results) there remained one neonate with missing data for age, one for gestational age at birth, one for birth weight, four for head circumference at birth, four for birth length, and one for ethnicity. How these were dealt with is detailed below, as it differed for univariate and multivariate analyses.

Table 2.2. The number of missing observations for each test

	Number of ears missing	Reasons for absence
DPOAEs	18 (10 right, 8 left)	All due to unsettled neonates
HFT	28 (14 right, 14 left)	All due to unsettled neonates
Reference standard	13 (7 right, 6 left)	
WAI	35 (18 right, 17 left)	• 15 due to unsettled neonates • 11 because unable to obtain a hermetic seal • 9 because of technical issues

DPOAEs, distortion product otoacoustic emissions; HFT, high-frequency tympanometry; WAI, wideband acoustic immittance.

Logistic regression was used to model the probability that an ear had a conductive condition. Since an assumption of logistic regression is independence of observations, one ear of each neonate was randomly selected to develop the model (the development sample), and the opposite ears were used for validation (the validation sample). Neonates with results for one ear only were included in the development sample to maximize sample size. Figure 2.1 shows the flow of participants and observations through the study, including the number of ears that passed and failed the reference standard for the development and validation samples.

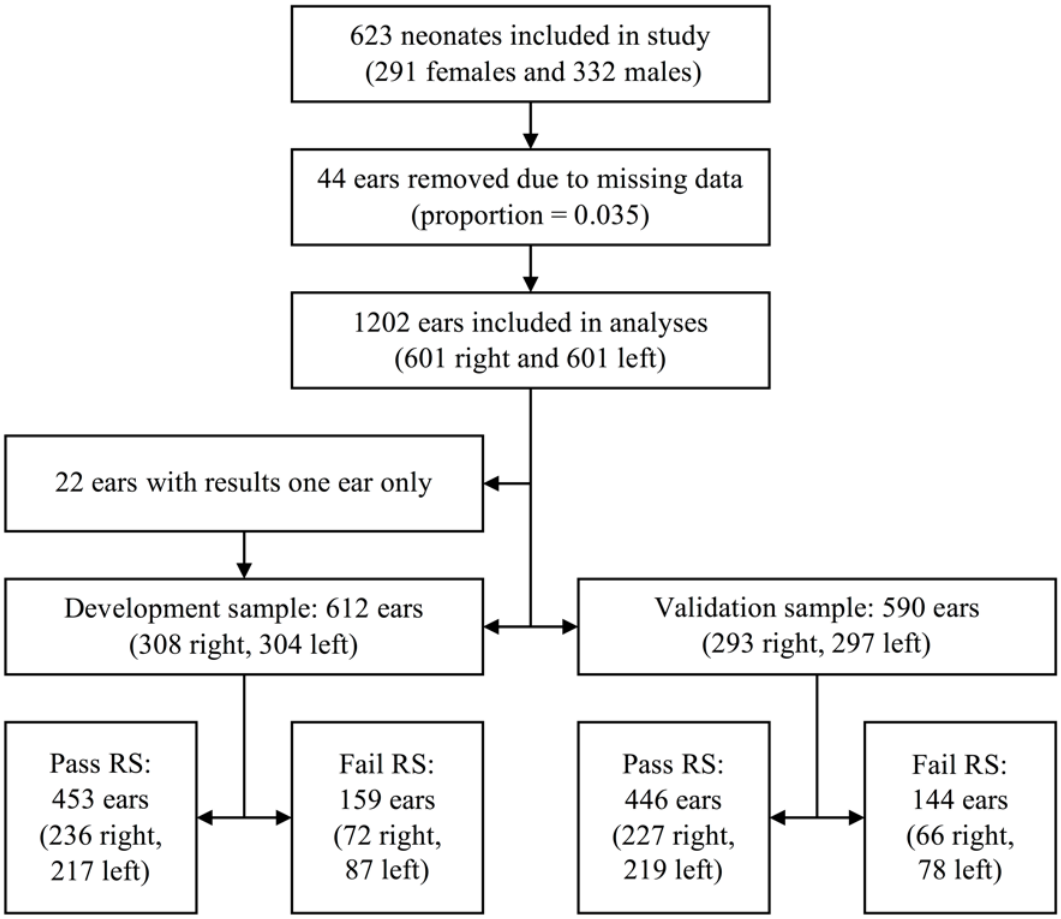


Figure 2-1. Flow of participants and observations (ears) through the study.
The pass and fail reference standard (RS) boxes give the number of ears that passed and failed the a priori RS for the development and validations samples.

2.3.4.1 Univariate analyses

Results of previous research investigating the effects of covariates such as gender, ear side and ethnicity on WAI outcomes have been equivocal (Aithal et al., 2014a; Aithal et al., 2013; Beers et al., 2010; Kei et al., 2013; Merchant et al., 2010; Shahnaz et al., 2013). Therefore, an unadjusted (univariate) analysis was initially performed on the development sample to investigate whether covariates were related to the results of the a priori reference standard (either fail). Ears with missing data for the covariate being modelled were removed from the training sample for that analysis. For example, the development sample contained 612 observations, one with missing data for age. Therefore, the univariate analysis of age was done with 611 observations.

Univariate analyses were also performed on the WAI data to identify the most predictive frequencies. Univariate modelling was done using the a priori reference standard (HFT and DPOAEs combined with the either fail criterion), and also with other reference standards to compare the effect

that the reference standard criterion had on diagnostic performance. Three other reference standards were calculated for comparison with the either fail criterion: DPOAEs and HFT each used as the reference test in isolation, and both tests combined using the both fail criterion (fail = fail both HFT and DPOAEs, otherwise pass).

2.3.4.2 Multivariate modelling

As a general guide, to avoid overfitting, a logistic regression model should have at least 10 observations in the smallest group (usually the abnormal group) for each predictor (Agresti, 2013). However, this can be relaxed if the signal-to-noise ratio is high, and the shrinkage coefficient (γ) can be used to assess whether too many model parameters (degrees of freedom) are being estimated (Harrell, 2015):

$$\gamma = \frac{\text{LR } \chi^2 - df}{\text{LR } \chi^2}, \tag{2.1}$$

where *df* is the total degrees of freedom from the predictors in the model, and LR χ^2 is the likelihood ratio chi-squared statistic of the model (the statistical significance test of the model). If γ indicates overfitting, further reducing the number of predictors may improve model generalizability (Harrell, 2015).

As there were 159 ears that failed the a priori reference standard in the development sample (see Fig. 2.1), the 10:1 observation to predictor rule limited the number of predictor variables to 15. Therefore, data reduction was necessary, as at 1/24 octave frequency resolution there were 107 test frequencies for each WAI measure. As mentioned, averaging the response into octave bandwidths reduced the number of predictors to 18 (six for each WAI measure). This was still over the 10:1 ratio, but was thought to be acceptable, as the powerful discriminatory ability of WAI reported in the literature indicates a high signal-to-noise ratio. The full WAI response was initially modelled at octave-frequency resolution. If γ was under 0.9, overfitting was indicated, as the model was expected to perform over 10% worse on new data (Harrell, 2015). Hence, the number of variables was further reduced, using two methods: iterative and a priori.

The iterative data reduction method included the top two most accurate univariate frequencies for each WAI measure as predictors. The a priori data reduction used predictors encompassing diagnostically important regions reported in the literature: A_{1000}, A_{2000}, $|Y|_{1000}$, $|Y|_{2000}$, φ_{Y1000}, and φ_{Y4000} (subscript numbers denote the centre frequencies of octave bandwidths). The single best univariate predictor was also modelled to investigate whether there was a multivariate advantage.

Also, a model was fitted including statistically significant covariates identified in the univariate analyses, along with WAI variables as predictors. This was done to assess whether including predictive covariates produced a better-fitting model than using WAI alone as model predictors. For these analyses, covariates with missing data in the development sample were assigned a value using multiple imputation (Buuren & Groothuis-Oudshoorn, 2011). This was done so that all multivariate models could be fitted using a single dataset, without having to discard observations with complete reference standard and WAI results, missing data only for covariates. Multiple imputation of statistically significant covariates in the development sample resulted in five imputed values being used: one for age, and four for birth length.

Predictor variables were modelled both linearly and nonlinearly, to test whether relaxing the assumption of linearity improved model fit. The logistic regression equation takes the form:

$$\text{Prob}\{\text{ear} = fail\} = \frac{1}{1 + e^{-X\hat{\beta}}}, \tag{2.2}$$

where Prob is the probability that an ear has a conductive condition (fail), and *e* is Euler's number. $X\hat{\beta}$ represents $\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k$, where the *X*'s ($X_1$ to X_k) are the predictor variables in the model (i.e., WAI frequencies or covariates), β_0 is the intercept, and β_1 to β_k are the regression coefficients for each variable. In models where predictors were modelled linearly, variables in $X\hat{\beta}$ (the *X*'s) were assumed to have a linear relationship with the outcome (the reference standard), but this assumption was relaxed for the models with nonlinear predictor variables using restricted cubic splines (RCS). Cubic splines are piecewise, third-order polynomials that are allowed to differ between joining points called "knots". Knots were placed at the 0.05, 0.275, 0.5, 0.725, and 0.95 percentiles of a predictor (Harrell, 2015). For example, when modelling A_{2000} as nonlinear, knots were placed at *A* of 0.15, 0.48, 0.65, 0.78, and 0.87, which were the 0.05, 0.275, 0.5, 0.725, and 0.95 percentiles, respectively, for *A* at 2000 Hz in the development sample. A five-knot RCS requires estimation of four model parameters (i.e., it "costs" four *df*). Continuous covariates (e.g., age, birth weight) were also modelled using restricted cubic splines for both the unadjusted (univariate) and multivariate analyses. The location of knots for all WAI predictors and covariates are provided in Section 2.6: Appendix A.

Akaike's information criterion (AIC), a measure of model fit that penalizes complexity (overfitting), was used to select the best fitting model (Burnham & Anderson, 2002). AIC was chosen as the metric for model selection because model fit, including both calibration and discrimination, is important for a clinical prediction model. A lower AIC indicates a better fitting model, if the models were

fitted using the same dataset. AUC was not used in the model selection process, as it does not penalize for overfitting, making it a poor metric for choosing between competing multivariate models.

The performance of the final model was evaluated through measures of calibration and discrimination. Calibration was assessed with calibration curves, which plotted a smoothed graph (using locally weighted smoothing) of actual against predicted probabilities. Predictions from a model are reliable if they align closely with observed frequencies of the condition. For example, for neonates with predicted probability of 0.3, approximately 3 out of 10 should actually have conductive dysfunction (Steyerberg et al., 2010). Satisfactory calibration is vital if predictions are used in decision making, because a neonate could be misdiagnosed, or mismanaged if predictions are very different from actual probabilities. However, calibration is not the only necessary condition for a model to be useful, discrimination is also critical. For example, a model would be perfectly calibrated, but uninformative, if it simply predicted the disease prevalence for each subject. A clinically useful model should make accurate predictions across a wide range of predictions, i.e., be well calibrated, and also discriminating (Harrell et al., 1996). The discriminative ability of the final model was assessed with AUC. This calculated the probability that for two neonates selected at random, one normal and one with conductive dysfunction, the model assigned higher probability of conductive dysfunction to the neonate with the condition (Steyerberg, Van Calster, & Pencina, 2011).

The model was validated by applying the model coefficients to the validation sample (opposite ears) and through bootstrap resampling of the development sample. Bootstrapping was employed as a secondary form of validation due to correlations between the development and validation samples. The procedure involved sampling with replacement from the development sample, a sample the size of the original (a training sample). The model was fit on this bootstrapped (training) sample and the coefficients applied to the development sample (the test sample). The difference in performance between the training and test samples gave an estimate of the amount of bias in the model. This process was repeated 500 times and averaged. The estimated bias was subtracted from the model performance of the development sample to provide a bias-corrected estimate, that is, the expected performance of the model on new data (Steyerberg, 2008).

Statistical analyses were performed with R (R Core Team, 2017), using the *rms* library (Harrell, 2016). This report has been written to comply with the guidelines of the TRIPOD (transparent reporting of a multivariable prediction model for individual prognosis or diagnosis) statement for reporting multivariate clinical prediction models (Collins, Reitsma, Altman, & Moons, 2015).

2.4 Results

2.4.1 WAI results and statistical modelling

Median and IQR WAI for the pass and fail groups defined by the a priori reference standard are shown in Figure 2.2 as a function of frequency. Median A for the pass group was higher than the fail group across the entire frequency range and median $|Y|$ of the pass group was greater up to 2800 Hz. Median φ_Y had a more complicated pattern, with the pass group higher than fail from 500 to 1200 Hz; and lower from 226 to 500 Hz, and also 1200 to 8000 Hz. IQRs of the groups were completely separated from 1200 to 3000 Hz for A, 800 to 1800 for $|Y|$, and 1800 to 3500 Hz for φ_Y.

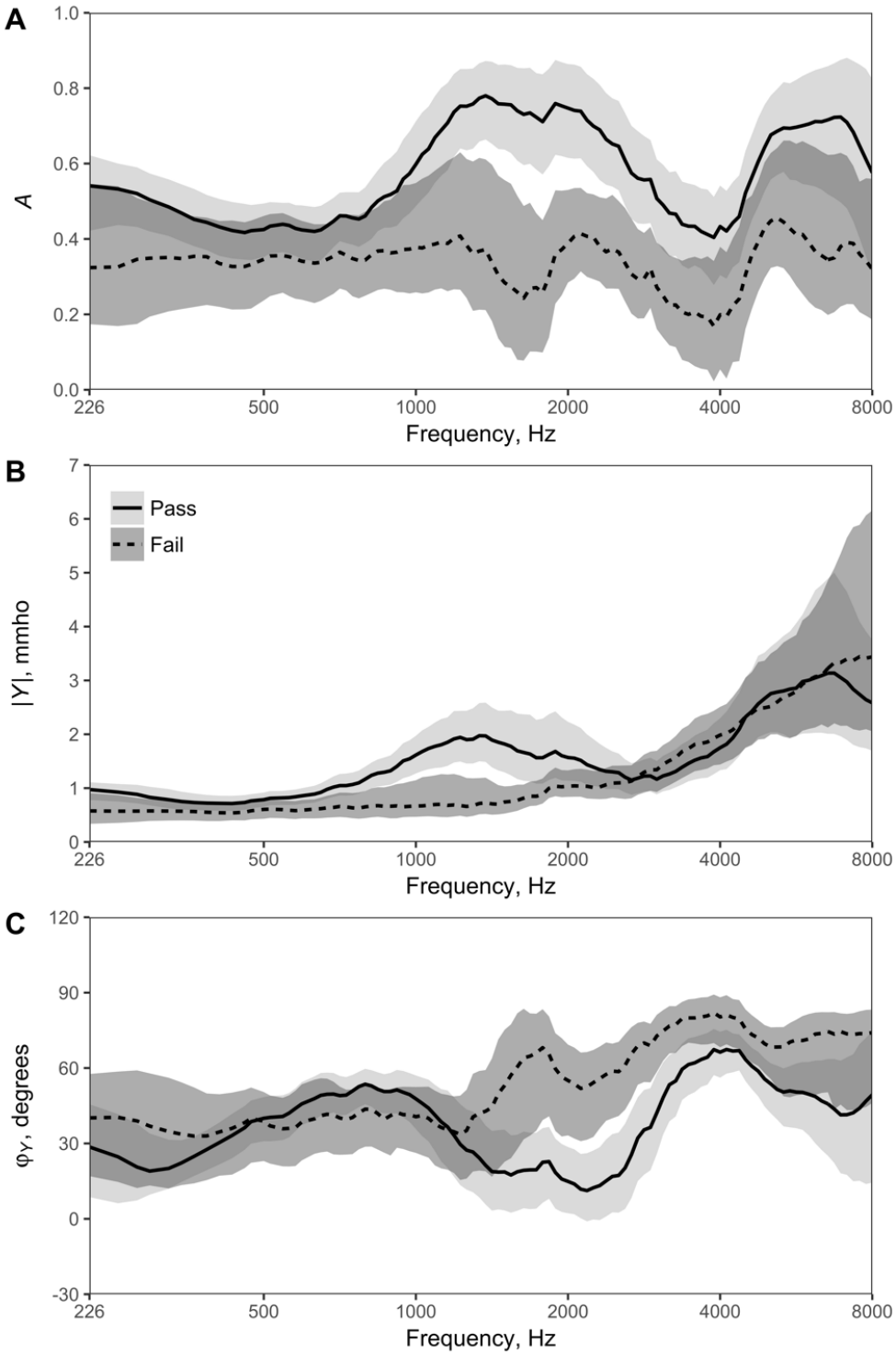


Figure 2-2. Median and interquartile range for the pass and fail groups.
The top panel (A) shows absorbance (*A*), the middle panel (B), admittance magnitude (|*Y*|), and the bottom panel, admittance phase (φ_Y; C). The solid and dashed lines represent the median of the pass and fail groups, respectively. The shaded regions represent the interquartile range (IQR) for the pass (light grey) and fail (dark grey) groups. The darkest shaded areas show where IQRs for the groups overlap. Data are from the full sample plotted at 1/24 octave frequency resolution.

2.4.1.1 Univariate results

Results of the univariate WAI analyses are shown in Figure 2.3 as a function of frequency for the a priori reference standard (either fail HFT or DPOAEs = fail, otherwise normal), the both fail reference standard (both fail HFT and DPOAEs = fail, otherwise normal), and DPOAEs and HFT used in isolation as the reference tests. The top two predictors for the a priori reference standard were A at 1000 and 2000 Hz (AUC = 0.77 and 0.85, respectively), $|Y|$ at 1000 and 2000 Hz (AUC = 0.83 and 0.74, respectively), and φ_Y at 2000 and 4000 Hz (AUC = 0.83 and 0.73, respectively). The reference standard with highest AUC for each WAI measure was the both fail criterion (AUC = 0.91 for A_{2000}, 0.91 for $|Y|_{1000}$, and 0.92 for φ_{Y2000}). The reference test with the second highest AUC for each WAI measure was HFT only (AUC = 0.87 for A_{2000}, 0.89 for $|Y|_{1000}$, and 0.89 for φ_{Y2000}). The DPOAE only reference test results were similar to those of the a priori (either fail) reference standard. DPOAE had higher AUC at some frequencies, a priori at others, and they were almost identical at some frequencies. The frequency with highest AUC for each WAI measure was the same for both reference standards (A_{2000}, $|Y|_{1000}$, and φ_{Y2000}), with a priori fail being slightly higher than DPOAEs for each: 0.853 compared to 0.846 for A_{2000}, 0.83 to 0.82 for $|Y|_{1000}$, and 0.83 to 0.80 for φ_{Y2000}.

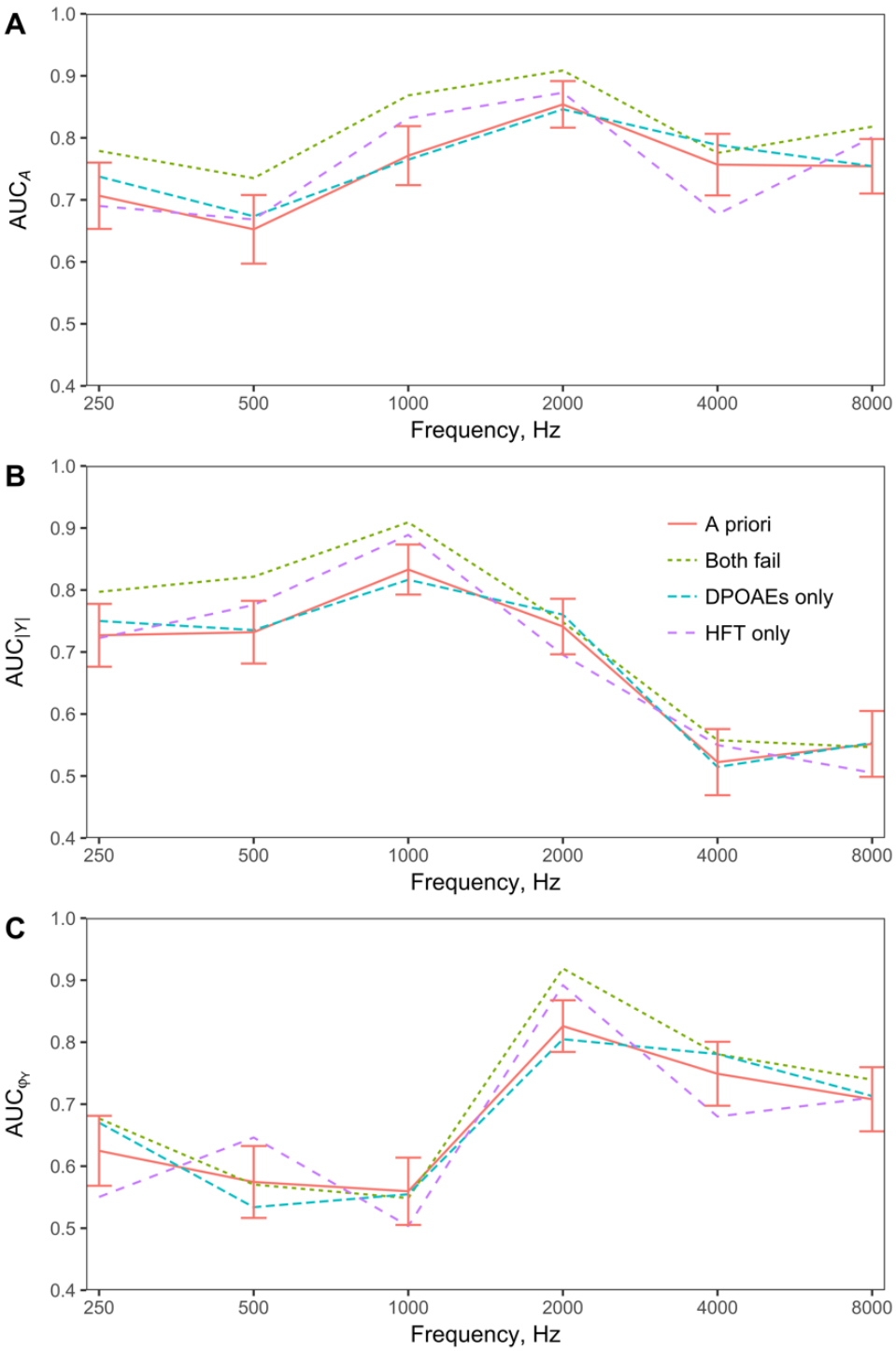


Figure 2-3. Results of univariate WAI analyses.
AUC as a function of frequency is shown for the absorbance (*A*; A, top), admittance magnitude (|*Y*|; B, middle) and admittance phase (φ_Y; C, bottom) WAI measures. AUC for each WAI measure is plotted at each frequency for four different reference standards: the a priori reference standard (either fail, i.e., fail either HFT or DPOAEs = fail, otherwise pass), both fail (i.e., fail both DPOAEs and HFT = fail, otherwise pass), DPOAEs only, and HFT only. The 95% confidence interval bars are provided for the a priori reference standard.

Results of the unadjusted (univariate) covariate analyses are presented in Table 2.3. Age, birth type and birth length were all statistically significantly associated with the outcome (the a priori reference

standard). Figure 2.4 shows the effect that changing each of the significant covariates had on the probability of conductive dysfunction. For age (left panel), the probability of conductive dysfunction increased sharply for neonates less than 40 hours old, and for birth length (middle panel), risk of conductive dysfunction increased under 52 cm.

Table 2.3. Statistics for the unadjusted covariate models

Covariate	LR χ^2	*df*	*p*-value	AUC
Ear side	2.19	1	0.139	0.534
Age	40.59	4	<0.001	0.655
Gender	0.08	1	0.773	0.507
Ethnicity	1.54	5	0.908	0.519
Gestational age	1.22	4	0.875	0.526
Birth type	23.96	1	<0.001	0.608
Birth weight	7.08	4	0.132	0.557
Head circumference	5.59	4	0.232	0.561
Birth length	13.36	4	0.001	0.576
Age × Birth type	47.09	9	<0.001	0.684

LR χ^2 and AUC results for the unadjusted (univariate) covariate models. Results for the model including age, birth type, and the interaction between the two (age × birth type) are also shown. Continuous variables had 4 df because they were modelled as nonlinear with 5-knot restricted cubic splines. Ethnicity had 5 df because there were 6 ethnic groups. The age × birth type model had 9 df: 4 for age, 1 for birth type, and 4 for the interaction. AUC, area under the receiver operating characteristic curve, df, degrees of freedom; LR χ^2, model likelihood ratio chi-squared statistic.

Birth type (Fig. 2.4, right panel), however, had an unanticipated association, showing that neonates delivered vaginally had higher risk of conductive pathology than those delivered via C-section. This was unexpected as it has been reported that babies delivered via C-section are at higher risk of failing their initial newborn screen due to fluid/debris in the outer/middle ear (Smolkin et al., 2012). It was thought that age may be contributing to this finding, as the hearing screening team at Townsville Hospital follow the recommendation of Smolkin et al. (2012) of prolonging screening babies born via C-section until they are at least 48 hours old when possible. In the full dataset, median age of neonates born by C-section was 50 hours compared with median age of 32 hours for babies delivered vaginally.

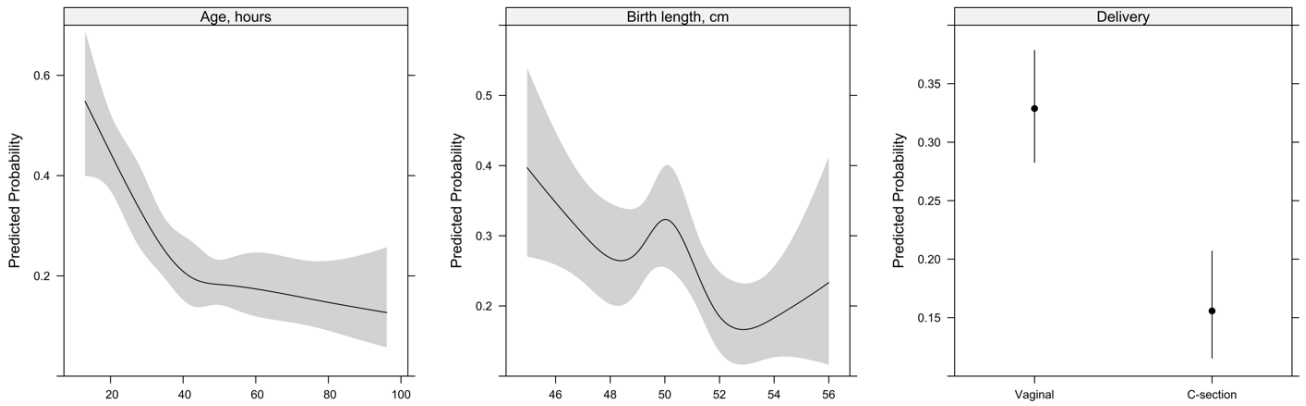


Figure 2-4. The effect of varying statistically significant covariates.
Statistically significant covariates from unadjusted (univariate) analyses showing the effect that varying age (left panel), birth length (middle panel), and birth type (right panel) had on probability of conductive dysfunction (*y* axes). The *x* axes labels and units are provided in the heading above each panel. For example, the *x* axis for the left panel is age in hours. The line for the plots of continuous variables (age and birth length) is the restricted cubic spline function truncated at the 10th lowest and 10th highest values. The shaded area denotes the 95% confidence interval. The 95% confidence interval for birth type is shown by the error bars.

To further investigate this relationship, a model was fitted including age, birth type, and an interaction between the two as variables (with age modelled as nonlinear). The results are shown in Figure 2.5, which depicts the probability of conductive dysfunction as a function of age for each type of birth. This model shows that neonates delivered via C-section did have significantly higher risk of conductive dysfunction over the first day of life, although babies delivered vaginally had high risk over this period as well.

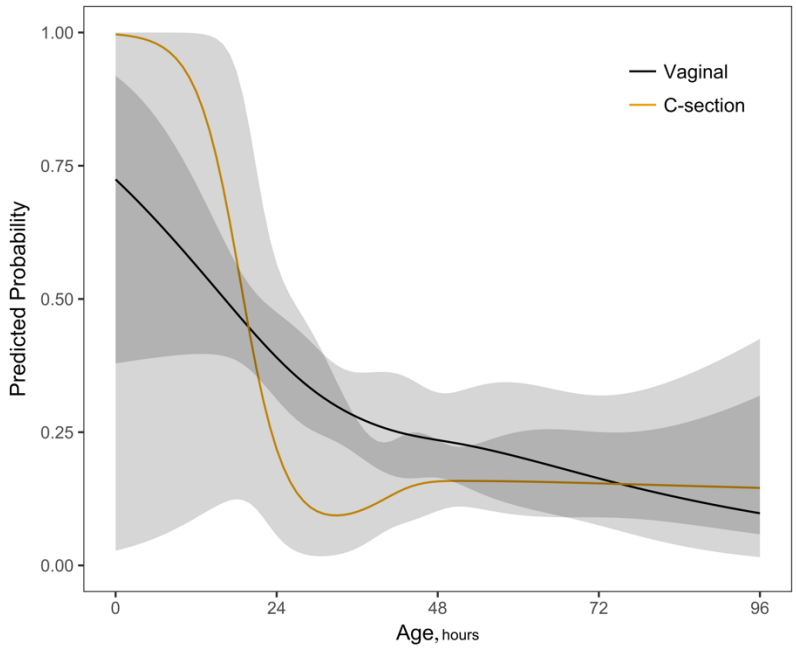


Figure 2-5. Predictions from the model including age and birth type.
The effect that varying age (*x* axis) had on probability of conductive dysfunction (*y* axis) for neonates born vaginally (black line) or via C-section (gold line), from the model including age, bith type, and their interaction as variables.

Age was the most important covariate from the unadjusted analyses (highest AUC and LR χ^2, and lowest *p*-value). To investigate the effect of age on the response, WAI was plotted for 1-, 2- and ≥3-day-old neonates for the pass and fail groups determined by the a priori reference standard (Fig. 2.6). There was a systematic change in A over the three groups with older neonates generally having higher values above 1200 Hz for the pass group, and from 800 to 2200 Hz for the fail group. There was also a systematic increase in $|Y|$ from 600 to 3000 Hz for the pass group, with values decreasing above 2000 Hz for the fail group. Again, φ_Y had a more complex pattern with older ears from the pass group having higher values between 400 Hz and 1500 Hz, and lower values above 1500 Hz. The fail group had higher φ_Y between 400 and 1000 Hz, and lower values from 1000 to 4000 Hz.

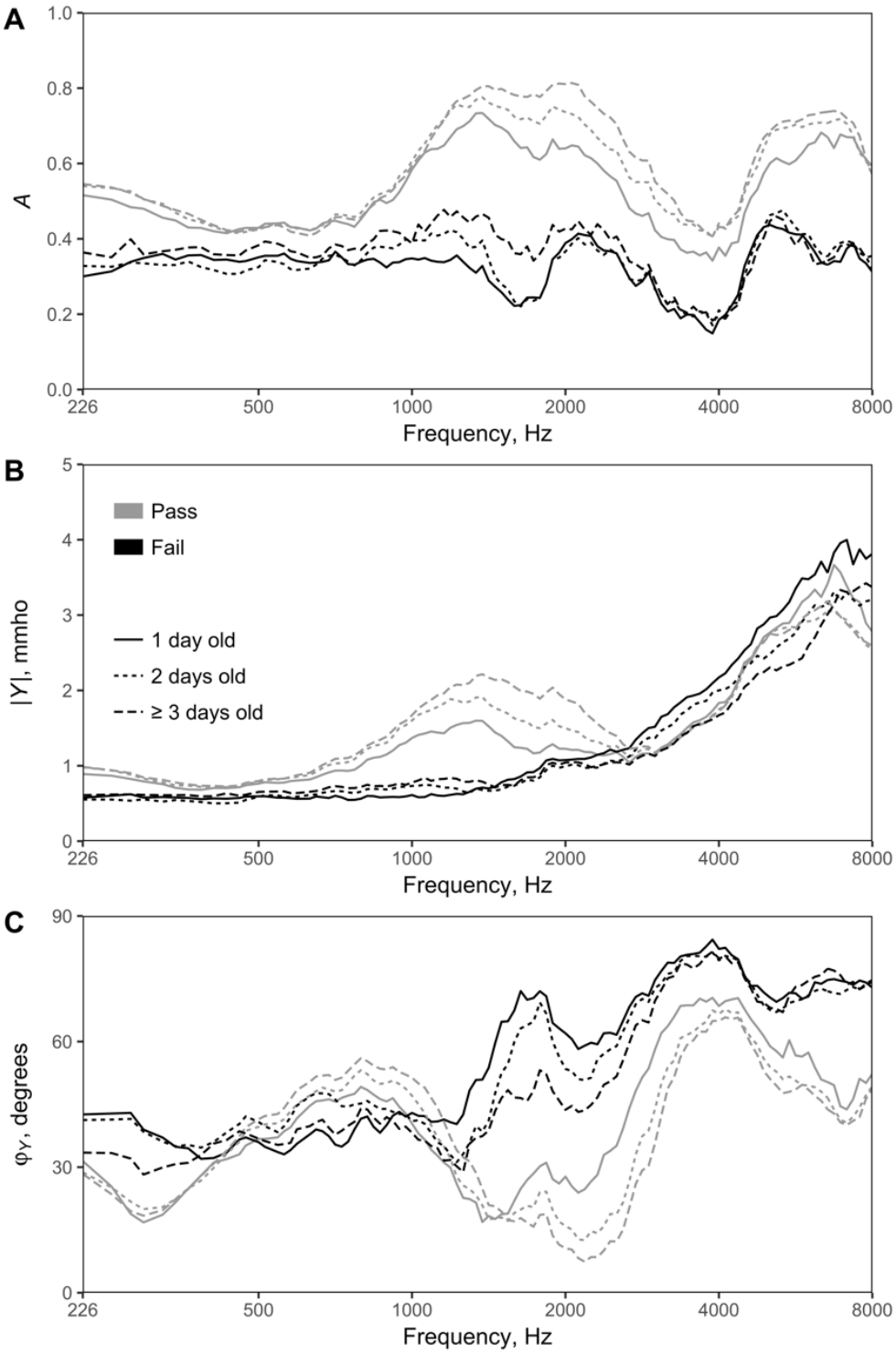


Figure 2-6. Age effects on WAI over the first three days of life.

Median absorbance (*A*; A, top), admittance magnitude (|*Y*|; B, middle) and admittance phase (φ_Y; C, bottom), for the a priori reference standard pass (grey) and fail (black) groups for 1-, 2- and ≥3-day-old neonates (full, dotted and dashed lines, respectively). Data are from the full sample plotted at 1/24 octave frequency resolution.

2.4.1.2 Multivariate modelling

All multivariate modelling utilized the a priori (either fail) reference standard. The predictors included in each model, along with explanations of the number of df are provided in Table 2.4. Statistical results of each model, including LR χ^2, γ, and AIC are presented in Table 2.5.

Table 2.4. Details of the fitted models

Name	Predictors in model	L/NL	*df*	Explanation of *df*
Model A_1	Best UV predictor (A_{2000})	L	1	1 predictor modelled as linear
Model A_2	A_{2000} nonlinear	NL	4	1 predictor nonlinear[1]
Model B_1	All WAI predictors[2]	L	18	18 predictors modelled as linear
Model B_2	All WAI predictors[2] nonlinear	NL	72	18 predictors modelled as nonlinear (4 *df* each)
Model C_1	Top 2 UV predictors nonlinear for each WAI measure[3]	NL	24	6 predictors with assumption of linearity relaxed (4 *df* each)
Model C_2	Model C_1 predictors, along with significant covariates from UV analyses[4]	NL	37	6 nonlinear WAI predictors, plus 4 *df* each for age and body length (nonlinear), 1 *df* for birth type, and 4 *df* for age × birth type interaction
Model C_3	Top UV predictor from each WAI measure[5], along with significant covariates from UV analyses[4]	NL	25	3 nonlinear WAI predictors (4 *df* each) plus 13 *df* for covariates as per Model C_2
Model D_1	WAI predictors encompassing important diagnostic regions identified in prior research[5]	L	6	6 predictors modelled as linear
Model D_2	Model D_1 predictors[5] nonlinear	NL	24	6 predictors with assumption of linearity relaxed (4 *df* each)

[1]1 predictor modelled as nonlinear cost 4 df because a restricted cubic spline requires estimation of 4 model parameters.
[2]A, $|Y|$ and φ_Y at 250, 500, 1000, 2000, 4000, and 8000 Hz.
[3]A_{1000}, A_{2000}, $|Y|_{1000}$, $|Y|_{2000}$, φ_{Y2000}, and φ_{Y4000} (numbers in subscripts denote the frequency).
[4]Significant covariates were age, birth type, and body length, an interaction was included between age and birth type.
[5]A_{2000}, $|Y|_{1000}$, φ_{Y2000}.
[6]A_{1000}, A_{2000}, $|Y|_{1000}$, $|Y|_{2000}$, φ_{Y1000}, and φ_{Y4000}.
Information about the fitted multivariate models, detailing the predictors included in each model, and explanation of the number of df. Models with linear predictors enforced the assumption that the relationship between continuous predictors and the probability of conductive dysfunction was linear. Models with nonlinear predictors relaxed this assumption by allowing predictors to have a flexible relationship with the outcome. However, this "cost" four *df*, rather than one. *A*, absorbance; $|Y|$, admittance magnitude; φ_Y, admittance phase; *df*, degrees of freedom; L, linear; NL, nonlinear; UV, univariate.

The multivariate modelling process began by modelling the top univariate WAI predictor (A_{2000}) by itself, for comparison with later multivariate models to assess whether there was a multivariate advantage. A_{2000} was modelled both as a linear, and nonlinear variable (Models A_1 and A_2, respectively), to investigate whether relaxing the assumption of linearity improved model fit. The model that allowed A_{2000} to have a nonlinear association with the outcome (A_2) was better fitting, with lower AIC (480.83 compared to 491.60).

Table 2.5. Statistics for each of the fitted models

	LR χ^2	*df*	γ	AIC
Model A_1	213.57	1	1.00	491.60
Model A_2	230.34	4	0.98	480.83
Model B_1	257.70	18	0.93	481.47
Model B_2	338.97	72	0.79	508.20
Model C_1	273.57	24	0.91	477.60
Model C_2	287.72	37	0.87	489.45
Model C_3	265.10	25	0.91	488.07
Model D_1	227.56	6	0.96	487.61
Model D_2	276.21	24	0.91	474.96

LR χ^2, *df*, γ, and AIC for each of the fitted models. *p*-values for the LR χ^2 were <0.001 for all models. AIC, Akaike's information criterion; *df*, degrees of freedom; γ, shrinkage coefficient; LR χ^2, likelihood ratio chi-squared statistic of the model.

Next, the full WAI response (250 to 8000 Hz, for each WAI measure) was modelled both linearly and nonlinearly (Models B_1 and B_2, respectively). These models were fitted to check for multivariate advantage compared to the univariate models (A_1 and A_2). The multivariate model with linear predictor variables (B_1) had lower AIC than the univariate linear variable model (A_1) (481.47 compared to 491.60), indicating that including multiple frequencies as predictors improved model fit.

Model B_1 had 18 predictors (*df*) with γ of 0.93 indicating that overfitting was not of concern (see Table 2.5). Model B_2, however, had 72 *df*, because allowing a variable to be nonlinear "cost" four *df* ($18 \times 4 = 72$). The corresponding γ of 0.80 indicated this model may perform poorly on new samples (because it was <0.9). Therefore, to minimize risk of overfitting, it was necessary to reduce the number of variables in the model.

The initial data reduction strategy (the iterative approach) involved including predictors based on the univariate WAI analyses. The first model fitted with this strategy (Model C_1) included the top two

univariate predictors for each WAI measure as nonlinear variables in the model: A_{1000}, A_{2000}, $|Y|_{1000}$, $|Y|_{2000}$, φ_{Y2000}, and φ_{Y4000}. This model had γ of 0.91, indicating that further data reduction was not necessary. Another model was also fitted that included statistically significant covariates identified in univariate analyses. This model (C_2) had the same predictors as Model C_1, and also included age, birth type, and birth length as covariates. Numeric covariates (age and length) were modelled as nonlinear and an interaction was included between age and birth type. This model may have been overfitting, with γ of 0.87, so further data reduction was performed. The model was refitted (Model C_3) using the same covariates, but including only the top univariate predictor for each WAI measure, allowed to be nonlinear (A_{2000}, $|Y|_{1000}$, and φ_{Y2000}). This model had γ of 0.91, indicating that overfitting was no longer of concern. AIC for Model C_3 was not better than Model C_1 (488.07, compared to 477.60), suggesting that including covariates did not produce a better fitting model than using WAI predictors only.

The second data reduction strategy was to use WAI predictors that encompassed diagnostically important frequency regions identified in prior research: A_{1000}, A_{2000}, $|Y|_{1000}$, $|Y|_{2000}$, φ_{Y1000}, and φ_{Y4000}. Models were fitted using both linear and nonlinear predictor variables (Models D_1 and D_2, respectively), to assess whether allowing nonlinearity improved model fit. The nonlinear predictor variable model (D_2) had lower AIC (474.96 compared to 487.61), indicating that it was better fitting. Overfitting was not an issue for this model (γ = 0.91). Since AIC was lower for Model D_2 than Model C_1 (the best fitting model using the iterative data reduction strategy), Model D_2 was chosen as the final model for interpretation and validation.

The relative importance of predictors in Model D_2 was investigated with a Wald analysis, which calculated the χ^2 statistic and associated *p*-value for each variable in the model. (An important variable will have a large χ^2 statistic, and therefore, low *p*-value.) The results are presented in Table 2.6. A_{2000} was the most important variable with χ^2 of 25.92 which was more than double the χ^2 of the next highest predictor ($|Y|_{2000}$, χ^2 = 13.11). The other four predictors, A_{1000}, $|Y|_{1000}$, φ_{Y4000}, and φ_{Y4000}, were not statistically significant in the model. The Wald analysis investigated the total contribution from each variable (linear + nonlinear), as well as the nonlinear components separately for each variable (see Table 2.6). The results show that the nonlinear components for each of the statistically significant predictors (A_{2000}, and $|Y|_{2000}$) were also statistically significant, and made up a significant proportion of the χ^2 statistic. For example, of the 25.92 χ^2 for A_{2000} ($p < 0.0001$), 20.96 ($p = 0.0001$) of this came from nonlinear contributions. Variables that were not statistically significant were still retained in the model, as statistical significance is a poor metric for predictor selection, especially when predictors are correlated, as with WAI data (Harrell, 2015). Variables that were not statistically significant may still

be contributing to the model, and retaining them is not detrimental if there is a reason for including them, and there are enough degrees of freedom to accommodate them (Gelman & Hill, 2007).

Table 2.6. Wald statistics for Model D_2 predictors

Factor	χ^2	df^1	*p*-value
Absorbance 1000 Hz	1.77	4	0.777
Nonlinear	1.50	3	0.683
Absorbance 2000 Hz	25.92	4	<0.001
Nonlinear	20.96	3	<0.001
Admittance magnitude 1000 Hz	3.34	4	0.502
Nonlinear	3.19	3	0.364
Admittance magnitude 2000 Hz	13.11	4	0.011
Nonlinear	9.19	3	0.027
Admittance phase 1000 Hz	5.05	4	0.282
Nonlinear	2.92	3	0.405
Admittance phase 4000 Hz	6.27	4	0.180
Nonlinear	5.47	3	0.140
TOTAL NONLINEAR	38.23	18	0.004
TOTAL	161.47	24	<0.001

[1]Each 5-knot restricted cubic spline required estimation of 4 parameters.
The top line in the table for each predictor (next to the predictor name) summarizes the total (linear + nonlinear) χ^2 statistic and *p*-value. The second line (*Nonlinear*) summarizes the contribution of the nonlinear components only. χ^2, chi-squared statistic; *df*, degrees of freedom.

To provide an intuition of how Model D_2 functions, Figure 2.7 shows the effect that changing each variable in the model has on predicted probability, with other variables held at typical values (the median). For example, take two hypothetical neonates, whose WAI results differed only for A_{2000}: for one neonate A_{2000} was 0.6, and for the other, it was 0.3. Results for all other variables for both neonates were typical (median) values (0.59 for A_{1000}, 1.46 mmho for $|Y|_{1000}$, 1.36 mmho for $|Y|_{2000}$, 40.96 degrees for φ_{Y1000}, and 59.15 degrees for φ_{Y4000}). We can now input the WAI predictor variable values into the equation for Model D_2, to obtain the probability of conductive dysfunction for each neonate. For the neonate with A_{2000} of 0.6, this results in predicted probability of 0.2, and for the neonate with A_{2000} of 0.3, probability of 0.8. Figure 2.7 is a visual representation of this process, repeating it to calculate probability across the entire range of values for each variable (with other variables held to

typical values). Notice that for A_{2000} (top middle panel), risk of conductive dysfunction is low (<0.2) when A is >0.6, but increases sharply from A of 0.6 to 0.4 where the risk is quite high (around 0.8). Interestingly, risk begins to decline when A is <0.45. Relaxing the assumption of linearity for a variable allows the relationship to be non-monotonic in this way. A similar effect is evident for $|Y|_{2000}$, as probability peaks at around 2 mmho, but then declines for $|Y|$ below this. This seems to protect against false positives. There were 16 ears in the development sample with $|Y|_{2000}$ less than 2 mmho and other variables greater or equal to median (typical) values. Of these, 13 passed the a priori reference standard, and 3 failed. Observations with $|Y|_{2000}$ below 2 mmho when other variables were typical were more likely to be normal. However, this does not mean that all ears with $|Y|_{2000}$ under 2 mmho will result in low predicted probability. Selecting observations in the development sample where $|Y|_{2000}$ was under 2 mmho and other variables were lower than the median resulted in 6 observations that all failed the reference standard, and this is reflected in the model. For example, inputting 10th percentile values for all variables into Model D_2 (A_{1000} = 0.29, A_{2000} = 0.25, $|Y|_{1000}$ = 0.52, $|Y|_{2000}$ = 0.76, φ_{Y1000} = 17.43, φ_{Y4000} = 80.16[1]), yields 0.83 probability of conductive dysfunction.

[1] For φ_{Y4000} this is the 90th percentile because at this frequency the fail group had higher median than the pass group (see Fig. 2).

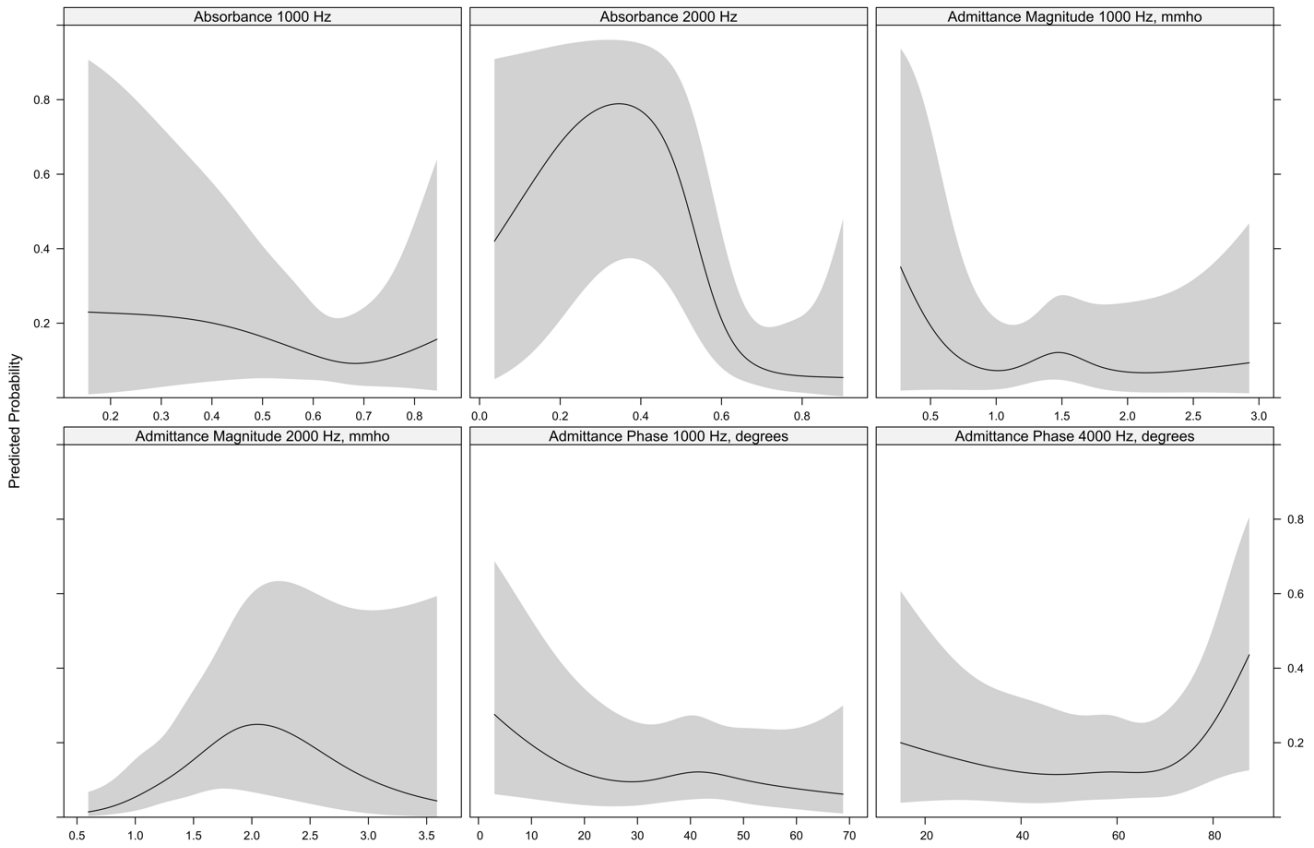


Figure 2-7. The effect of varying Model D_2 predictors

The effect of varying Model D_2 absorbance (A), admittance magnitude ($|Y|$), and admittance phase (φ_Y) predictor variables (x axes) on predicted probability (y axes) with other variables held at median value. The lines are the restricted cubic spline function for a predictor variable truncated at the 10th lowest and 10th highest values for each variable. The shaded areas denote the 95% confidence interval. A_{1000}, A_{2000}, and $|Y|_{1000}$ are shown on the top row (left to right); and $|Y|_{2000}$, φ_{Y1000}, and φ_{Y4000} are presented in the bottom row (left to right; the number in subscripts denotes frequency). The x axes labels and units are provided in the heading above each panel. For example, the x axis for the bottom left plot is $|Y|_{2000}$ with units of mmho. Median values were: 0.59 for A_{1000}, 0.65 for A_{2000}, 1.46 mmho for $|Y|_{1000}$, 1.36 mmho for $|Y|_{2000}$, 40.96 degrees for φ_{Y1000}, and 59.15 degrees for φ_{Y4000}. For example, the top middle plot shows how the predicted probability of conductive dysfunction changes as values for A_{2000} change (with other predictors held to their median). For A_{2000}, risk of conductive dysfunction is low (<0.2) when A_{2000} is >0.6, but increases sharply from A_{2000} of 0.6 to 0.4 where the risk is quite high (around 0.8).

The calibration plots for Model D_2 for the development (apparent), bootstrapped (bias-corrected) and validation samples are shown in Figure 2.8. The calibration plots were fairly close to the ideal calibration line for the development and bootstrapped samples (predictions were slightly high between probabilities of 0.1 to 0.3). The plot for the validation sample showed that predictions were slightly high up to probabilities of 0.6, and slightly low for predictions >0.7. Overall, calibration for the development, bootstrapped and validation samples was satisfactory. The distributions of predictions are depicted by the histograms (the small vertical lines) along the inside of the x-axes of the calibration curves (Fig. 2.8). The distribution is skewed right, with the majority of predictions falling between 0 and 0.1 (the vertical lines are tallest in this region), as most ears were normal. There are tick marks all

the way along the *x*-axes, however, showing that Model D_2 made predictions over almost the whole range of probabilities (0 to 1). This, in conjunction with the calibration lines being close to the ideal, indicates that predictions are reliable across the entire spectrum of probabilities. AUC for Model D_2 for the fitted model (development sample), bias-corrected (bootstrapped), and validation samples are presented in Table 2.7. AUC for the fitted model was 0.88 (95% confidence interval [CI], 0.84–0.91). Bootstrap resampling calculated a bias of 0.03, leaving a bias-corrected AUC of 0.85. Applying the model coefficients to the validation sample (opposite ears) resulted in an AUC of 0.90 (95% CI, 0.87–0.94). This was unexpected, being better than the development sample, but 0.90 was within the 95% CI of the development sample. The equation for Model D_2 to calculate probabilities is provided in Section 2.7: Appendix B. A web application implementing Model D_2 is available (https://joshmyers.shinyapps.io/WAIPredictions/), that can compute probabilities for a neonate using a file exported from the Titan Suite software, or by manually entering values.

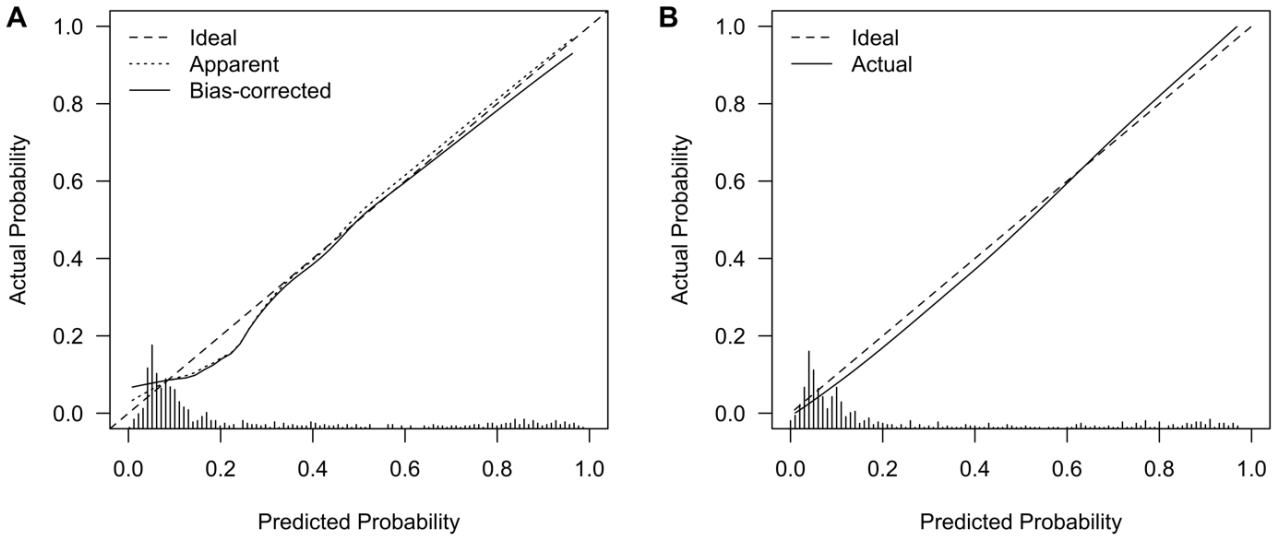


Figure 2-8. Calibration curves for Model D_2.
Calibration curves for Model D_2 plotting predicted (*x* axes) against observed probability (*y* axes). The dashed lines show an ideal model where predicted equals observed frequency (probability). The histograms (vertical lines) along the inside of the *x* axes show the distribution of predicted values. A, calibration for the development sample (apparent, dotted line), and after being corrected for bias using bootstrap resampling (solid line). B, Calibration of Model D_2 applied to the validation sample (actual, solid line).

Test accuracy characteristics including sensitivity, specificity, predictive values, and contingency table results for various diagnostic thresholds for Model D_2 are presented in Table 2.8. Rationale for the thresholds presented is provided in the Discussion section.

Table 2.7. AUC for Model D_2 for the development, bias-corrected, and validation samples

Sample	AUC (95% CI)
Development	0.876 (0.840 to 0.911)
Bootstrap training	0.897
Bootstrap test	0.866
Bias (bootstrap training – test)	0.030
Bias-corrected (development – bias)	0.845
Validation	0.903 (0.870 to 0.936)

AUC, area under the receiver operating characteristic curve; CI, confidence interval.

2.4.2 Reference standard criterion

The rationale for using the a priori (either fail criterion) reference standard, was that ears that failed HFT but passed DPOAEs were likely to have conductive dysfunction. If this was indeed the case (and ears that failed one test only were not false positives), we would expect it to be reflected in the WAI results. Median WAI for ears that passed both tests in the reference standard (n = 899), failed HFT only (n = 39), failed DPOAEs only (n = 172), and failed both tests (n = 90) are plotted in Figure 2.9 as a function of frequency. For A and $|Y|$, ears that passed both tests generally had the highest values across the entire frequency range (apart from 2500 to 4000 Hz for $|Y|$). For ears that failed both tests, A mostly had the lowest values across the range of frequencies, and $|Y|$ had the lowest values up to 1800 Hz. Median A and $|Y|$ for ears that failed HFT only were generally higher than those that failed DPOAEs only (except for 1000 to 1800 Hz, and 6000 to 8000 Hz for A). Median φ_Y had a more complex pattern. The group that passed both tests had highest values for frequencies below 1200 Hz, and lowest values from 1800 to 4000 Hz. The group that failed both tests had lowest φ_Y from 300 to 900 Hz, and highest φ_Y from 1200 to 5000 Hz. Median φ_Y for ears that failed one test only lay between these extremes from 300 to 800 Hz, and 1800 to 4000 Hz. The HFT-fail group had higher φ_Y than the DPOAE fail group from 500 to 3000 Hz.

Table 2.8. Diagnostic accuracy statistics for various probability thresholds

Threshold	Description	Sp	Se	NPV	PPV	TN	FN	TP	FP
0.08	Sensitivity = 0.9	0.46	0.91	0.94	0.37	209	14	145	244
0.18	Point of symmetry	0.81	0.81	0.92	0.61	369	30	129	84
0.37	Specificity = 0.91	0.91	0.75	0.91	0.75	414	39	120	39
0.43	10th percentile pass	0.93	0.68	0.89	0.77	421	51	108	32
0.60	1.5:1 HBR	0.96	0.60	0.87	0.84	435	64	95	18
0.67	2:1 HBR	0.97	0.57	0.86	0.87	439	69	90	14
0.75	3:1 HBR	0.98	0.48	0.84	0.88	442	82	77	11

Specificity, sensitivity, predictive values, and contingency table results for different probability thresholds from Model D_2. FN, false negatives; FP, false positives; HBR, harm-to-benefit ratio; NPV, negative predictive value; PPV, positive predictive value; Se, sensitivity; Sp, specificity; TN, true negatives; TP, true positives.

2.5 Discussion

2.5.1 Clinical application of the prediction model

A well-fitting prediction model gives valid, individualized risk estimates. Prediction models also allow for grey zones (e.g., probability = 0.5), which is not possible when using group statistics such as sensitivity and specificity that are defined at a single cut-off point. Hunter et al. (2010) defined a region of uncertainty for A_{2000} between the 10th percentile of the pass group (0.44) and 90th percentile of the fail group (0.61).[2] Using the developed model (D_2), we can quantify how uncertainty changes over this region. With other predictors at median value, A_{2000} of 0.61 has probability of 0.19, and A_{2000} of 0.44 has a probability of 0.72 (see Fig. 2.7). Probability of conductive dysfunction increases sharply over this region. This additional information could be useful when making a diagnosis for a neonate with results falling inside this region. For example, there is a substantial difference between a predicted probability of 0.3 and 0.7, and management for two neonates with these respective results may be very different, as the former is more likely to be normal, and the latter likely has conductive dysfunction. However, A_{2000} for both of these predictions lie within the region of uncertainty defined by Hunter et al.

[2] Hunter et al. (2010) actually calculated the 90th percentile of the pass group and 10th percentile of the fail group measured in *R*.

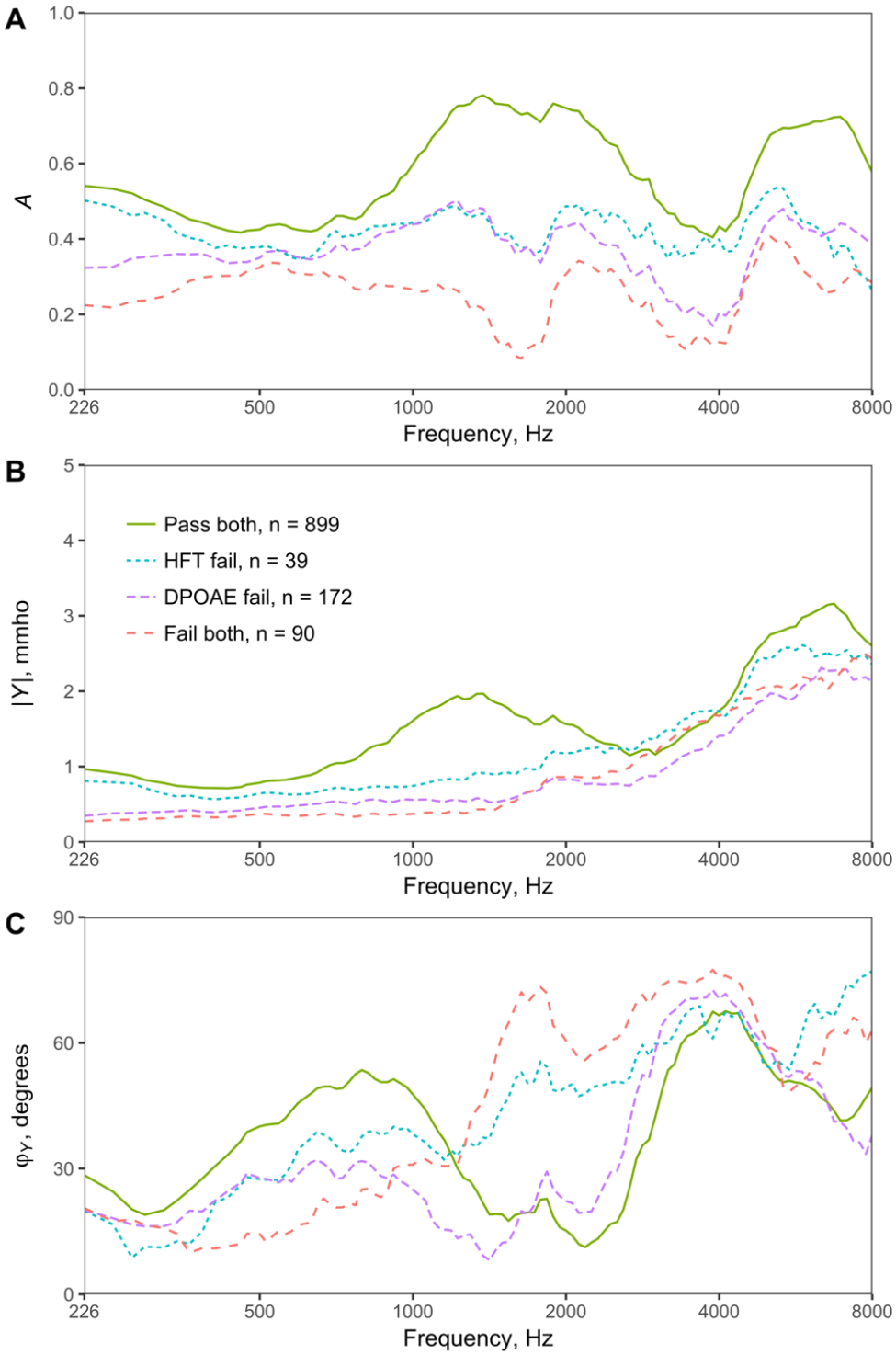


Figure 2-9. Median WAI stratified by reference test results.

Median absorbance (*A*; A, top panel), admittance magnitude (|*Y*|; B, middle panel) and admittance phase (φ_Y; C, bottom panel), for ears that passed both tests in the reference standard (green), failed high-frequency tympanometry (HFT) only (blue), failed distortion product otoacoustic emissions (DPOAEs) only (purple), and failed both tests (red). Data are from the full sample plotted at 1/24 octave frequency resolution.

Different diagnostic thresholds have been suggested in the literature. Hunter et al. (2010) and Aithal et al. (2015) used the 10th percentile of A from the normal group to set the threshold. Sanford et al. (2009) calculated thresholds on the ROC curve where: specificity = 91%, sensitivity = 90%, and sensitivity ~ specificity. Using these thresholds from Model D_2 shows the test characteristics for the various cut-offs (see Table 2.8). Setting all model predictors to 10th percentile of the pass group resulted in a predicted probability threshold of 0.43. At this point specificity = 0.93 and sensitivity = 0.68. The point where sensitivity = 0.90 (probability = 0.08) had specificity of 0.46. This is lower than Sanford et al. who reported specificity of 0.66 for φ_Y at this threshold. Sensitivity for the point on the curve where specificity = 0.91 in the present study was 0.75, which is higher than results of Sanford et al., who reported sensitivity of 0.65 at this point for φ_Y. The point of symmetry (the threshold where sensitivity ~ specificity) for the present study yielded sensitivity and specificity of 0.81 (using a probability cut-off of 0.18). This is slightly higher than the point of symmetry reported by Sanford et al. of 0.78 for A and φ_Y.

A limitation to the above approaches to choosing the cut-off point is that they are data-driven, using thresholds derived from the test characteristics. Ideally, the cost of misclassification should come externally, from the severity associated with different types of misdiagnoses (D. J. Hand in discussion to Briggs & Zaretzki, 2008; Harrell, 2015). Predicted probabilities provide a simple and flexible way to set the referral threshold, taking into account the harm (or cost) associated with false positives. Setting the threshold at probability = 0.5 assumes that the costs of harms (false positives) and benefits (true positives) are equal (Steyerberg, 2008). The diagnostic thresholds used in previous research in Table 2.8 are all <0.5, which implies that the cost of a false negative (missing a case) is higher than a false positive (over-diagnosis). However, in the context of using WAI as an adjunct test when screening for permanent hearing loss, the cost of false positives is arguably higher than false negatives. This indicates that a threshold >0.5 would be more appropriate. For example, setting the threshold at 0.6, means that we consider false positives to "cost" 1.5 times true positives (odds = 1.5:1, or 6:4). Another way to think about where to set the threshold is to consider the maximum number of false positives the screening program would we be willing to bear for every true positive. For example, setting the threshold at 0.6, implies that up to four false positives would be acceptable for every 10 refer results. The threshold could be adjusted depending on considerations such as available resources and the goals of the program.

The effect of increasing the threshold can be seen in Table 2.8. As the cut-off increases, specificity increases and negative predictive value decreases (as the number of false negatives increases), but

sensitivity decreases and positive predictive value increases (as the number of false positives decreases).

In a diagnostic context, probabilities could be used in conjunction with graphical depictions of WAI to aid diagnostic decision making. Figure 2.10 demonstrates an example of applying the model to a case, showing *A*, along with predicted probability from Model D_2, for the right ear of a 29-hour-old male. Diagnosis is challenging, as *A* is within the 90% range at most frequencies, but is consistently toward the lower bound, and the clinician may be unsure if this is significant. Knowing that the probability is 0.61 for this ear may help with decision making, as a conductive condition is more likely than not. In a clinical context, there may still be a preferred diagnostic threshold that the clinic uses as a guideline (based on the relative costs of false positives and negatives). However, rather than only labelling a result as pass or fail, the probability provides the clinician with degree of diagnostic certainty, in the form of an interpretable quantitative parameter. In the above example, a positive diagnosis carries a 39% chance or error (1 − 0.61).

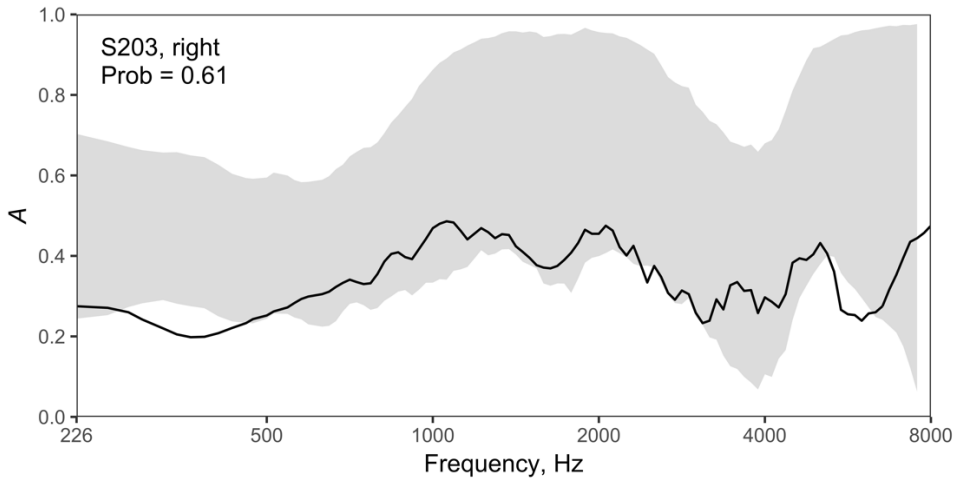


Figure 2-10. Example of applying the model to a case.
Absorbance (*A*, black line) and predicted probability from Model D_2 (Prob), for the right ear of subject 203, a 29-hour-old male. The shaded area is the 90% range (5th to 95th percentile) of *A* for the normal group, derived from the full dataset. Data are plotted at 1/24 octave frequency resolution.

2.5.2 Multivariate model development

The aim of the present study was to develop a diagnostic prediction model for detecting conductive conditions in neonates. A clinically useful prediction model needs to be well fitting, both discriminating and well calibrated. Relaxing the assumption of linearity improved model fit, as shown by lower AIC in the models with nonlinear, compared to linear predictor variables. Model A_2, which used the top univariate predictor (A_{2000}), modelled as a nonlinear variable was better fitting (with lower AIC) than its linear variable counterpart (Model A_1). Also, Model D_2, the multivariate model with nonlinear

predictor variables fitted using a priori variable selection, was better fitting than Model D_1, which had the same predictors assumed to be linearly related to the outcome (the a priori reference standard). Allowing a WAI predictor variable to be nonlinear "cost" more *df* (four rather than one), meaning that fewer frequencies were able to be included as variables in the model. However, including fewer predictors but allowing nonlinearity (Models C_1 and D_2) produced better fitting models than including the full WAI response with linearity imposed (Model B_1). Furthermore, the nonlinear components of significant predictor variables (A_{2000}, and $|Y|_{2000}$) in the final model (D_2) were statistically significant (see Table 2.6). Allowing the most significant variable (A_{2000}) to be nonlinear seemed especially important, as probability increased sharply between A_{2000} of 0.65 and 0.45 (see Fig. 2.7).

Including multiple WAI predictors also improved model fit. The multivariate models using linear predictor variables (B_1 and D_1), were better fitting (with lower AIC) than the equivalent univariate (linear predictor variable) model (A_1). The multivariate WAI models that used nonlinear variables with acceptable γ (>0.9; Models C_1 and D_2) were also better fitting than their univariate (nonlinear) counterpart (Model A_2).

In summary, relaxing the assumption of linearity, and including multiple WAI frequencies as predictors, both improved model fit. The best-fitting model on the development sample (Model D_2) used both of these strategies. Data reduction was necessary to be able to incorporate both nonlinearity and multiple predictors, as the nonlinear predictor variable model including the full WAI response was likely overfitting (γ = 0.79). The a priori data reduction method (Model D_2) had lower AIC than the iterative approach using the best univariate predictors (Model C_1), but only slightly (475 compared to 478). The difference between the models was only one variable. Model D_2 included φ_{Y1000}, but Model C_1 used φ_{Y2000} instead. There was a crossing in the φ_Y of the pass and refer ears between 1000 and 2000 Hz (see Fig. 2.2C), which suggests that including φ_{Y1000} included extra information that was not contained in φ_{Y2000} or φ_{Y4000}. The a priori data reduction method was the preferable method, as it decreases the risk of fitting the model based on idiosyncrasies in the current dataset, which can affect generalizability.

Including statistically significant covariates did not produce a better fitting model than using WAI variables only. The covariates were predictive (the Age × Birth type model had AUC of 0.68, see Table 2.3), but this extra information did not improve model fit when combined with WAI predictors. This was not unexpected, as although including important covariates can improve model fit, it is not uncommon that variables that are predictive in univariate analyses, do not improve performance of a multivariate model (Steyerberg et al., 2011). Not having covariates in the model means that all information needed to make a prediction is contained in the WAI response. This is beneficial, as a clinician

would not need to obtain information such as birth length before using the model to test middle ear function of a neonate.

Internal validation of the final model (D_2) showed limited bias, with AUC for the bias-corrected and validation samples within the 95% CI of the development sample. The calibration plots for the development, bootstrapped, and validation samples all showed satisfactory calibration, which indicates that Model D_2 was not overfitting the data.

2.5.3 Discriminative ability of WAI

AUC for the univariate and multivariate analyses in the present study are comparable with results reported in the literature. The multivariate AUC of 0.88 (95% CI, 0.84−0.91) for Model D_2 for the development sample in the present study was similar to the best AUC of 0.87 for multivariate φ_Y reported by Sanford et al. (2009). AUCs for the development and validation samples for Model D_2 (0.88 and 0.90, respectively) in the present study were comparable to the AUCs of 0.86 and 0.82 reported by Keefe et al. (2003a), being slightly higher, but within the 90% CI for the development for the development sample and higher for the validation sample (opposite ears). Technical and methodological differences between the studies could account for the variation in results. Keefe et al. used EOAEs as the reference standard, different equipment, and different data reduction techniques (principal component analysis). Despite these differences, both studies demonstrated good internal validation (limited bias).

Results of the univariate WAI analyses in the present study found A_{2000} to be the most accurate predictor. This is in agreement with Hunter et al. (2010), who reported R_{2000} as the best performing frequency (AUC = 0.90), using DPOAEs as the reference test. Results for A_{2000} against the DPOAE reference standard in the present study, however, was not as high (AUC = 0.85; 95% CI, 0.81−0.89; see Fig. 2.3). The present study differs from Aithal et al. (2015), who found 1250 Hz to be the most discriminating frequency (AUC = 0.77), using a HFT-DPOAE reference standard. However, results from the present study are not directly comparable with Aithal et al., as they calculated AUC using a single threshold on the curve, after classifying A for a given frequency as pass or fail using the 10th percentile from a normal group.

In summary, there are many similarities in AUC results between the present, and previous studies, but also some differences. Differences between studies could be due to sample size and characteristics, equipment, reference standards, data reduction techniques, and statistical methodology. Importantly, all studies have demonstrated that WAI accurately discriminates between normal, and ears with conductive dysfunction.

2.5.4 Covariates

Statistically significant covariates were birth length, birth type, and age. Research in animals has found body size, including length, to be related to dimensions of the ear canal, middle ear and tympanic membrane (Huang, Rosowski, & Peake, 2000; Shahnaz et al., 2013; Werner & Igic, 2002; Werner, Montgomery, Safford, Igic, & Saunders, 1998). In the present study, having a longer birth length (>51 cm, see Fig. 2.4) was protective. Having a larger ear canal and middle ear may make the system more efficient in removing fluid and debris from the outer and middle ear after birth. Age was the covariate with the highest LR χ^2 statistic and AUC, with younger neonates (<40 hours) more likely to have outer/middle ear dysfunction. This is consistent with previous research demonstrating that conductive dysfunction affects newborn hearing screening results in the first days of life (Chang et al., 1993; Doyle et al., 2000; Sanford et al., 2009; Thornton, Kimm, Kennedy, & Cafarelli-Dees, 1993). The fact that developmental trends were evident over the first three days of life, even in the group that passed both DPOAEs and HFT (see Fig. 2.6), indicates that the effect of age on the WAI response may be due in part to physiological maturation, and not solely due to the clearing of fluid and debris from the conductive pathway.

Type of birth was also expected to be significantly related to conductive dysfunction, as previous research has found that babies delivered by C-section are more likely to fail an initial EOAE hearing screen (Smolkin et al., 2012). Results of the univariate analysis, however, showed that C-section delivery had a protective effect, because these babies were typically older. The model including birth type, delivery type, and their interaction, showed that neonates delivered by C-section did have substantially higher risk of conductive dysfunction over the first day of life (see Fig. 2.5). Since risk for C-section babies was higher only for the first day, future research could investigate whether the recommendation of postponing hearing screening of babies delivered by C-section until 48 hours after birth could be revised to 24 hours (Smolkin et al., 2012).

2.5.5 Choice of reference standard and WAI results

AUC for the univariate WAI results was fairly similar for the either fail (a priori) and DPOAE only reference standards. The either fail reference standard was slightly higher at some frequencies, and DPOAE at others, but AUC using the DPOAE reference test was within the 95% CI for either fail at all frequencies. However, this does not mean that including HFT along with DPOAEs in the a priori reference standard (using the either fail criterion) was not worthwhile. There were 39 ears that failed HFT but passed DPOAEs in the full sample (see Fig. 2.10), making up 12.5% of ears in the fail group

of the either fail reference standard. The a priori reference standard was adopted with the goal of creating a more sensitive reference standard (to identify more true positives), than using DPOAEs as the sole reference test. Although there was not a large effect on univariate AUC results, including the extra test identified an extra 39 ears with conductive dysfunction that would have been otherwise classified as normal using DPOAEs as the sole reference test. Given the high specificity of both HFT and DPOAEs, these 39 ears were more likely to be true positives than false positives. Also, ears that failed only one reference test (HFT or DPOAEs) generally had median WAI values that fell between the median for ears that passed and failed both tests, indicating that failing one test identified milder conditions (see Fig. 2.10). It is noteworthy that this was the case over most frequency regions thought to be diagnostically important for conductive conditions.

Interestingly, univariate results showed that the both fail reference standard had the highest AUC, followed by HFT only, for the most accurate frequency for each WAI measure (A_{2000}, $|Y|_{1000}$, and φ_{Y2000}). This was because these reference standards are stricter, failing only ears with (presumably) more severe conductive dysfunction, making the diagnostic task easier (Schmidt & Factor, 2013). In the development sample 159 ears failed the either fail reference standard, and 134 ears failed DPOAEs only. This is on the order of three times the 52 and 68 ears that failed the both fail and HFT only reference standards, respectively. These results demonstrate the importance of pre-specifying the reference standard based on the target condition (Rutjes, Reitsma, Coomarasamy, Khan, & Bossuyt, 2007). In the present study, using a data-driven approach for choosing the reference standard would have resulted in using the both fail reference standard, as it had the highest AUC. However, it would be difficult to argue that ears that failed DPOAEs (with normal cochlear hearing) should be classified as normal, since they most likely failed the test due to conductive dysfunction (Hunter et al., 2010).

The median and IQR of ears that passed or failed the reference standard were comparable with results from previous studies. The present study found best separation between groups for A from 1200 to 3000 Hz. This compares favourably with results from Sanford et al. (2009) and Aithal et al. (2015). $|Y|$ results from the present study best separated groups between 800 and 1800 Hz which also compares well with results from Sanford et al. In the present study, φ_Y best separated groups between 1800 to 3500 Hz which is comparable with Sanford et al. who found of a region of good separation between 2500 to 4500 Hz. Sanford et al. also found a region of best separation from 750 to 1000 Hz for 1-day-old neonates. In the present study, median φ_Y was greater for the pass group from 500 to 1200 Hz but the separation between the IQRs was not as clear as the 1-day-old results from the Sanford et al. study. However, results from the present study are similar to the 2-day-old results reported by Sanford et al., which is closer to the median age in the present study.

2.5.6 Strengths, limitations, and directions for future research

The modelling strategy was intended to avoid overfitting by limiting the number of predictors in the model. Internal validation with bootstrapping and the validation sample showed limited bias in Model D_2, indicating that it was not substantially overfitting the data. The final model (D_2) included predictors based on results of previous research, which increases the likelihood that the model will generalize to new subjects. However, the degree to which the developed model generalizes to new subjects needs to be assessed through external validation, by assessing performance on a sample separated from the development sample temporally/geographically (Moons et al., 2012a). Future research could investigate external validation of Model D_2, using a sample of neonates that were not used to develop the model.

Recruiting neonates as a cohort enabled direct calculation of predicted probabilities which could potentially be useful in clinical or screening contexts. However, the model may need to be updated if being used only in neonates who failed the newborn hearing screen, as the prevalence of conductive conditions would likely be higher in that population than in the current sample (Thompson et al., 2001). This would not necessitate development of an entirely new model, however, as the prevalence in the new population can be used to update the intercept term of the present model (Moons et al., 2012a). Future research could assess the prediction model in clinical and screening settings, including in the population of neonates who failed newborn hearing screening.

Although including covariates related to the outcome did not improve model fit in the present study, this may be important if developing a prediction model for babies in special care. Future research could investigate whether including risk factors for middle ear dysfunction for neonates in special care, such as low birth weight and history of ventilation, improves model fit (Keefe et al., 2000).

A limitation of the present study was that data collection was not blinded. When testing a neonate, both the reference standard tests and WAI measurements were performed by the same research audiologist. However, interpretation of reference test results was objective, a tympanogram was either peaked, or not, and DPOAEs either satisfied the pass criterion, or did not. Doing all tests with the same insertion of the probe meant that neonates were not unsettled by changing of the probe tip between tests, and the status of the ear was unlikely to have changed from one test to the next (Aithal et al., 2015).

2.5.7 Conclusions

The developed prediction model accurately discriminated between normal and abnormal ears, and had satisfactory calibration, with performance and calibration verified in two ways. The model has potential applications in screening or diagnostic settings. In a screening context, probabilities could be used to set a cost-sensitive referral threshold that is intuitive and easy to apply. In a diagnostic setting, predicted probabilities could be used in conjunction with graphical depictions of WAI for individualized diagnoses. Further research investigating the use of the model in clinical and screening contexts is warranted.

2.6 Appendix A: The location of restricted cubic spline knots

The location of knots for continuous variables fitted with restricted cubic splines for Model D_2 in Chapter 2.

	Percentile				
	0.05	0.275	0.5	0.725	0.95
Absorbance					
250 Hz	0.15	0.39	0.50	0.55	0.67
500 Hz	0.20	0.34	0.42	0.48	0.59
1000 Hz	0.21	0.46	0.59	0.67	0.80
2000 Hz	0.15	0.48	0.65	0.78	0.87
4000 Hz	0.13	0.37	0.51	0.60	0.75
8000 Hz	0.13	0.45	0.63	0.80	0.91
Admittance magnitude (mmho)					
250 Hz	0.32	0.65	0.81	0.97	1.22
500 Hz	0.38	0.65	0.80	0.92	1.20
1000 Hz	0.40	1.08	1.46	1.79	2.50
2000 Hz	0.66	1.04	1.36	1.88	2.86
4000 Hz	1.04	1.57	2.00	2.52	3.68
8000 Hz	1.43	2.62	3.05	4.34	7.72
Admittance phase (degrees)					
250 Hz	−1.00	12.50	23.03	38.29	62.24
500 Hz	12.60	32.88	40.19	47.17	57.90
1000 Hz	10.57	31.06	40.96	49.95	63.21
2000 Hz	−5.58	11.36	23.83	41.86	77.98
4000 Hz	22.59	48.08	59.15	69.39	83.93
8000 Hz	−8.13	29.85	48.16	66.05	84.68
Covariates					
Age (hours)	16	29	43	51	76
Gestational age (weeks)	37.0	38.4	39.2	40.1	41.2
Birth weight (grams)	2690	3190	3460	3710	4230
Head circumference (cm)	32.5	34.0	35.0	35.5	37.0
Body length (cm)	47	49	50	52	54

2.7 Appendix B: The equation for Model D_2

The logistic regression equation for Model D_2 developed in Chapter 2, to calculate the probability (Prob) that an ear has a conductive condition (fail) is provided below. Model D_2 included A_{1000}, A_{2000}, $|Y|_{1000}$, $|Y|_{2000}$, φ_{Y1000}, and φ_{Y4000} as predictors, modelled as nonlinear with five-knot restricted cubic splines. A restricted cubic spline with five knots has six "slopes" or coefficients, one for each segment (therefore, there were six coefficients for each predictor in the model). A five-knot restricted cubic spline is modelled:

$$\beta_0 + \beta_1(k-a)_+^3 + \beta_2(k-b)_+^3 + \beta_3(k-c)_+^3 + \beta_4(k-d)_+^3 + \beta_5(k-e)_+^3,$$

Where the βs are the slopes, a, b, c, d, and e are the knot locations, and k is the value of the predictor (for a particular ear). The subscript "+" after a knot term, e.g., $(x)_+$, means that the value of x is x if it is positive, otherwise, 0. Therefore, in the equation, if x is positive, its value is cubed, otherwise it is ignored. This is the mechanism that enforces continuity (i.e., makes the splines meet at the joins). The terms within parentheses are the predictor values minus the knot location, i.e., how far the predictor value is above the knot. If it is below the knot, the term is set to 0 and ignored. If it is above the knot, the coefficient is multiplied by how far above, forcing continuity. The term is cubed to make the join smooth. For an observation (i.e., a measurement from an ear), if the value for a predictor is less than the first knot, it is multiplied by β_0 and all other terms are ignored (set to 0). More terms are added as the value of the predictor passes more knots. If the predictor value (k) is higher than the first knot but lower than the second knot, the value would be k times β_0 + (β_1 × how far k is above the first knot)$^{3\ \text{(cubed)}}$, and so on.

In the equation for Model D_2 below, A_{2000} has knots at 0.18, 0.48, 0.65, 0.78, and 0.87 (rounded to two decimal places). The numbers outside the parentheses (the βs in the above equation) are the coefficients. So for A_{2000}, $\beta_0 = 6.75$, $\beta_1 = 57.91$, and $\beta_2 = 289.93$. So if, for example, A_{2000} was measured in an ear of a neonate at 0.5, the overall term for A_{2000} would be: $(6.75 \times 0.5) + 57.91(0.5 - 0.18)^3 + 289.93(0.5 - 0.48)^3$. In this case the β_3, β_4 and β_5 terms would all be set to 0 because 0.5 is lower than the 3rd knot (0.65).

The equation for Model D_2 is:

$$\text{Prob}\{\text{ear} = fail\} = \frac{1}{1+\exp(-X\hat{\beta})}, \quad \text{where}$$

$X\hat{\beta} =$

$$\begin{aligned}
&0.3523545 \\
&-0.331527A_{1000} - 11.78557(A_{1000} - 0.2019479)_+^3 \\
&+20.75623A_{1000} - 0.4604719)_+^3 + 196.1606(A_{1000} - 0.5861042)_+^3 \\
&-334.7084(A_{1000} - 0.6729156)_+^3 + 129.5771(A_{1000} - 0.7955292)_+^3 \\
&+6.747808A_{2000} - 57.91562(A_{2000} - 0.14875)_+^3 \\
&+289.9319(A_{2000} - 0.4784832)_+^3 - 347.126(A_{2000} - 0.6467917)_+^3 \\
&+62.7052(A_{2000} - 0.7775833)_+^3 + 52.40457(A_{2000} - 0.8710521)_+^3 \\
&-3.589473|\,Y|_{1000} + 3.214916(|\,Y|_{1000} - 0.4037958)_+^3 \\
&-15.66089(|\,Y|_{1000} - 1.075514)_+^3 + 23.67445(|\,Y|_{1000} - 1.464896)_+^3 \\
&-12.60408(|\,Y|_{1000} - 1.791034)_+^3 + 1.375602(|\,Y|_{1000} - 2.500048)_+^3 \\
&+3.606295|\,Y|_{2000} - 1.807441(|\,Y|_{2000} - 0.6641479)_+^3 \\
&+3.578654(|\,Y|_{2000} - 1.038218)_+^3 - 3.090144(|\,Y|_{2000} - 1.36275)_+^3 \\
&+2.113073(|\,Y|_{2000} - 1.876652)_+^3 - 0.7941422(|\,Y|_{2000} - 2.857704)_+^3 \\
&-0.06509103\varphi_{\,Y_{1000}} + 6.467163\times10^{-5}(\varphi_{\,Y_{1000}} - 10.56548)_+^3 \\
&-0.0003710368(\varphi_{\,Y_{1000}} - 31.06615)_+^3 + 0.000520529(\varphi_{\,Y_{1000}} - 40.95885)_+^3 \\
&-0.000230828(\varphi_{\,Y_{1000}} - 49.9512)_+^3 + 1.666413\times10^{-5}(\varphi_{\,Y_{1000}} - 63.20532)_+^3 \\
&-0.0254035\varphi_{\,Y_{4000}} + 2.152242\times10^{-5}(\varphi_{\,Y_{4000}} - 27.5886)_+^3 \\
&-0.0001072373(\varphi_{\,Y_{4000}} - 48.0814)_+^3 + 0.0003023625(\varphi_{\,Y_{4000}} - 59.14977)_+^3 \\
&-0.0003343566(\varphi_{\,Y_{4000}} - 69.39033)_+^3 + 0.0001177089(\varphi_{\,Y_{4000}} - 83.92554)_+^3
\end{aligned}$$

and $(x)_+ = x$ if $x > 0$, 0 otherwise

Chapter 3. Diagnosing Middle Ear Pathology in 6- to 9-Month-Old Infants Using Wideband Absorbance: A Risk Prediction Model

This chapter develops a prediction model for diagnosing middle ear pathology using wideband absorbance in 6- to 9-month-old infants. It has been previously published in the article: Myers, J., Kei, J., Aithal, S., Aithal, V., Driscoll, C., Khan, A., Manuel, A., Joseph, A., Malicka, A. N. (2018b). Diagnosing middle ear pathology in 6- to 9-month-old infants using wideband absorbance: A risk prediction model. *Journal of Speech Language and Hearing Research, 61*(9), 2386-2404.

I made substantive contributions to the article in the areas of study design, data collection, data analysis and drafting of the article, as outlined below:

Contributor	**Statement of contribution**
Joshua Myers (Candidate)	Study design (60%) Recruitment and data collection (60%) Data analysis (100%) Wrote the article (100%)
Joseph Kei	Study design (20%) Edited the article (40%)
Sreedevi Aithal	Study design (5%) Edited the article (15%)
Venkatesh Aithal	Study design (5%)
Carlie Driscoll	Study design (5%) Edited the article (15%)
Asaduzzaman Khan	Study design (5%) Edited the article (15%)
Alehandrea Manuel	Recruitment and data collection (20%)
Anjali Joseph	Recruitment and data collection (20%)
Alicja N. Malicka	Edited the article (15%)

3.1 Abstract

Purpose: The aim of this study was to develop a risk prediction model for detecting middle ear pathology in 6- to 9-month-old infants using wideband absorbance measures.

Methods: Two-hundred and forty-nine infants aged 23 to 39 weeks (median = 28 weeks) participated in the study. Distortion product otoacoustic emissions and high-frequency tympanometry were tested in both ears of each infant to assess middle ear function. Wideband absorbance was measured at ambient pressure in each participant from 226 to 8000 Hz. Absorbance results from one ear of each infant were used to predict middle ear dysfunction, using logistic regression. To develop a model likely to generalize to new infants, the number of variables was reduced using principal component analysis, and a penalty was applied when fitting the model. The model was validated using the opposite ears, and with bootstrap resampling. Model performance was evaluated through measures of discrimination and calibration. Discrimination was assessed with the area under the receiver operating characteristic curve (AUC), and calibration with calibration curves, which plotted actual against predicted probabilities.

Results: AUC of the fitted model was 0.887. The model validated adequately when applied to the opposite ears (AUC = 0.852), and with bootstrap resampling (AUC = 0.874). Calibration was satisfactory, with high agreement between predictions and observed results.

Conclusions: The risk prediction model had accurate discrimination and satisfactory calibration. Validation results indicate that it may generalize well to new infants. The model could potentially be used in diagnostic and screening settings. In the context of screening, probabilities provide an intuitive and flexible mechanism for setting the referral threshold that is sensitive to the costs associated with true, and false positive outcomes. In a diagnostic setting, predictions could be used to supplement visual inspection of absorbance for individualized diagnoses. Further research assessing the performance and impact of the model in these contexts is warranted.

3.2 Introduction

Infants with onset of otitis media in the first year of life are at greater risk of recurrent and chronic infections (Homøe et al., 1999; Kværner et al., 1997; MacIntyre et al., 2010). Diagnostic tools able to quickly and accurately identify middle ear pathology early in infancy could help to facilitate timely intervention for these children (Hunter et al., 2008b). Wideband acoustic immittance (WAI) is an emerging technology for assessing middle ear function with significant advantages over established clinical tests such as tympanometry. WAI does not require pressurization of the ear canal, and can measure middle ear function over a wider frequency range than is possible with tympanometry (Stinson et al., 1982).

The term WAI refers to a family of wideband measures including reflectance, absorbance, acoustic impedance, and acoustic admittance. Reflectance is the proportion of the forward-propagating energy that is reflected back from the middle ear. Absorbance is 1 – reflectance, and represents the proportion of energy absorbed by the middle ear (Rosowski et al., 2013). Absorbance and reflectance have been the most reported WAI measures in clinical research, as they have the desirable property of being theoretically insensitive to probe location in the ear canal (Voss et al., 2008). Numerous studies have investigated the accuracy of WAI for identifying conductive dysfunction in neonates (Aithal et al., 2015; Hunter et al., 2010; Keefe et al., 2003a; Myers et al., 2018a; Sanford et al., 2009). However, little is known about the diagnostic performance of these measures for infants beyond the first month of life. Further research is needed in this population, as WAI could be a valuable tool for identification of middle ear pathology early in infancy.

Two diagnostic performance WAI studies, to date, have included infants under 12 months of age beyond the neonatal period. Prieve et al. (2013b) found that reflectance accurately diagnosed conductive hearing loss in a sample of 70 infants aged 3 to 25 weeks (median = 10 weeks). Ellison et al. (2012) reported that absorbance and acoustic admittance performed well in detecting middle ear effusion in a case-control study using 88 children aged 0.5 to 7 years old (median = 1.3 years for cases, and 1.2 years for controls). However, results of those studies may not be directly applicable to infants in the second half of the first year of life, due to changes in the acoustic properties of the outer and middle ear over the first 12 months of infancy (Kei et al., 2013). Developmental changes over the first year of life have a significant effect on the WAI response, to the extent that age-graded diagnostic criteria are essential (Hunter et al., 2013; Keefe et al., 1993; Werner et al., 2010). Although Prieve et al. (2013b) and Ellison et al. (2012) both included some subjects in the second half of the first year of life, most infants in those studies were either under 6, or over 12 months old, respectively.

There is a lack of consensus about the effect of subject characteristics on WAI for infants beyond the neonatal period. Werner et al. (2010) reported a slight, but statistically significant ear-side effect, with reflectance lower for left ears compared to right, but this result has not been replicated by other studies (Aithal et al., 2014b; Hunter et al., 2008b; Shahnaz et al., 2014). Variations between ethnic groups have been reported (Aithal et al., 2014a; Beers et al., 2010), but using ethnic-specific norms for absorbance or reflectance has not been found to improve diagnostic accuracy (Shahnaz et al., 2013). Werner et al. (2010) reported that females had higher acoustic impedance than males, but no studies have found a statistically significant gender effect for absorbance or reflectance in infancy (Aithal et al., 2014b; Hunter et al., 2008b; Shahnaz et al., 2014; Werner et al., 2010).

The large volume of data generated by WAI presents a challenge for interpretation of results. Both qualitative and quantitative methods have been used in previous research to analyse the diagnostic properties and performance of WAI. Qualitative methods seek to identify patterns and characteristics of WAI in normal and pathological ears to aid clinical diagnosis (e.g., Hunter et al., 2008b; Sanford & Brockett, 2014). Using quantitative statistical techniques, however, could help to improve accuracy, and reduce variability between testers (Sanford & Brockett, 2014). Quantitative methods could also be automated for use in screening contexts (Sanford et al., 2009).

Both univariate and multivariate statistical methods have been used in previous quantitative diagnostic WAI research in infants and children. The goal of univariate analyses is to find the most accurate diagnostic frequency (or frequency bandwidth) in the response (e.g., Beers et al., 2010; Prieve et al., 2013b). The benefit of this approach is ease of interpretation, but multivariate modelling combining results from multiple frequencies and/or WAI measures is potentially more accurate (Ellison et al., 2012; Prieve et al., 2013a). This has been demonstrated in research in neonates and children, where multivariate methods have been shown to outperform univariate approaches (Keefe et al., 2012; Myers et al., 2018a; Piskorski et al., 1999).

However, because of the many variables (frequencies) in the response, overfitting can be an issue when multivariate modelling WAI data (Piskorski et al., 1999). A model with many predictors can have high apparent accuracy, but may not perform well when applied to new subjects (Steyerberg, 2008). Piskorski et al. (1999) used stepwise regression to reduce the number of variables in their multivariate model. However, stepwise methods can lead to arbitrary inclusion of variables, especially if predictors are correlated, as with WAI data (Harrell, 2015). Data reduction methods should ideally be blind to the outcome (the reference standard results). The simplest way of achieving this with WAI data is to decrease the frequency resolution. However, care should be taken that the bandwidth is fine enough that important diagnostic information is not lost (Keefe et al., 2015). Principal component analysis

(PCA) is another effective data reduction method. This technique was used by Keefe et al. (2003a,b) to reduce the WAI response to 5–7 variables. PCA transforms the original variables (WAI frequencies) into new variables called principal components (PCs) that are ordered such that most of the variability (information) in the data is contained in the first few PCs (Jolliffe, 2002). All of the original variables contribute to each PC, but not equally. In a multivariate model, a subset of the PCs can be used as predictors instead of the original variables. This retains most of the information from the originals, but fewer variables need to be included in the model, reducing risk of overfitting (Harrell, 2015).

In addition to limiting the number of variables, applying a penalty when fitting a model can help to improve generalizability. Previous multivariate WAI models developed in infants and children have been fitted using maximum likelihood estimation (Ellison et al., 2012; Keefe et al., 2012; Myers et al., 2018a; Piskorski et al., 1999; Sanford et al., 2009). This method finds the best fitting model for the provided data, but the model may reflect idiosyncrasies (or noise) in the data, rather than generalizable relationships (the signal), especially if the sample size is small (Harrell, 2015). Penalized maximum likelihood estimation adjusts a model for optimism by applying a penalty to the parameter estimates, and has been demonstrated to improve generalizability (Moons, Donders, Steyerberg, & Harrell, 2004; Steyerberg, Eijkemans, Harrell, & Habbema, 2001a; Steyerberg, Eijkemans, & Habbema, 2001b).

However, even with a penalty applied, there is a degree of bias in all multivariate models, as they are fitted using a particular dataset. Validation is therefore an essential component of multivariate model development, to assess how well the model is likely to generalize to new subjects (Steyerberg et al., 2001c). Performance in new samples may be lower due to overfitting, or because of differences in subject characteristics, the test environment, or equipment used in the new setting (Steyerberg, 2008). Internal validation investigates the impact of overfitting on a model, and is performed during development, utilizing the same data used to create the model (Moons et al., 2012b). External validation assesses generalizability by evaluating the model in a new sample of subjects (Moons et al., 2012a).

Previous multivariate WAI models have been internally validated only in neonates. Keefe et al. (2003a,b) and Myers et al. (2018a) used one ear of each subject for model development, and opposite ears for validation. Models in all three studies validated adequately, with a difference in AUC between development and validation samples of 0.01 to 0.04. A limitation of this approach is that the development and validation samples are correlated. Therefore, Myers et al. (2018a) also employed bootstrap resampling as a secondary form of validation which estimated bias in the AUC of 0.03. Bootstrapping uses sampling with replacement to estimate the amount of bias (or overfitting) in the model to give an "honest" estimate of model performance in new samples (Harrell et al., 1996).

Research into useful multivariate techniques for analysing WAI data is still in its infancy. Multivariate methods that can summarize results from multiple frequencies into a single quantitative parameter are desirable, as this would combine the benefits of univariate utility and multivariate performance (Myers et al., 2018a; Sanford & Brockett, 2014). Diagnostic risk prediction models have this quality, as they take multiple predictors and provide a probability (between 0 and 1) that a subject has the condition (if probability >0.5, the condition is more likely to be present than not). Providing risk estimates has significant advantages over simply classifying ears as either pass or fail. Predictions are individualized, providing information about the degree of diagnostic certainty, and allowing for grey zones (e.g., probability = 0.5; Harrell & Slaughter, 2016). Such information is lost if the clinician is given only a single pass/fail threshold. Prediction models are especially useful for conditions such as otitis media that occur on a continuum of severity, where the pass/fail cut-off for diagnosis is somewhat arbitrary (Myers et al., 2018a; Northrop et al., 1986; Palmu & Syrjänen, 2005; Vickers et al., 2008).

The aim of the present study was to develop a multivariate risk prediction model for diagnosis of middle ear pathology in 6- to 9-month-old infants using absorbance. Absorbance was chosen as the WAI measure for model development because of the desirable property of being theoretically insensitive to probe position in the ear canal. Results from one ear of each infant were used to develop the model. To create a model likely to generalize to new subjects, the number of variables was reduced using PCA, and a penalty was applied when fitting the model. The model was validated using the opposite ears, and with bootstrap resampling.

3.3 Methods

Institutional review board approval was obtained from the Townsville Health Service District Institutional Ethics Committee, and the University of Queensland Behavioural and Social Science Ethical Review Committee. This study was part of a larger project following a cohort of subjects through infancy who were recruited at birth from the maternity (healthy baby) ward of the Townsville Hospital. Seven hundred and thirty-seven infants have been enrolled in the project to date. Myers et al. (2018a) reported on 629 study participants who were tested as neonates. The present study aimed to follow up participants at around 6 months of age. Two hundred and forty-nine infants attended follow-up appointments between February 2015 and August 2017. Median age was 27.7 weeks with an interquartile range (IQR) of 26.6 to 29.5 weeks (range = 23.9 to 39.9 weeks). Two hundred and seven infants were reported by caregivers to be Caucasian (86%), 32 Asian (13%), and four others (African, South American, or unknown; 2%).

All infants either passed the newborn hearing screening (automatic auditory brainstem response), or had a finding of normal hearing sensitivity at subsequent diagnostic audiology assessment. Diagnostic audiology was performed within 6 weeks of screening, with normal hearing defined as passing a click-evoked ABR test, and also passing either TEOAEs or a tone-burst ABR test, as per the Healthy Hearing Program Audiology Diagnostic Assessment Protocol (Nicholls, 2016). The pass criterion for TEOAEs was ≥6 dB signal to-noise ratio (SNR) at 3/4 frequencies from 1000 to 4000 Hz, including at least 1000/1500 and 4000 Hz. For click-evoked ABR, a pass was a repeatable wave V present down to 20 dB nHL, and for tone burst ABR, repeatable wave V present down to 30 dB nHL at 1000 Hz, and 20 dB nHL at 4000 Hz.

3.3.1 Test procedure

Infants were tested by a research audiologist in a quiet room at a paediatric community health centre. All tests were performed using an Interacoustics Titan system connected to a laptop computer running Titan Suite software (version 3.2). Probe function was checked daily in a 2 cm^3 cavity and the system was calibrated annually by the manufacturer. Infants were tested sitting on their parent's lap. Both ears of each child were tested if possible, with the most accessible ear tested first. An appropriately sized plastic probe tip was attached to the probe, and high-frequency tympanometry (HFT), distortion product otoacoustic emissions (DPOAEs), and absorbance were tested, in no specific order. Otoscopy was also performed on each infant, to ensure that the ear canal was not occluded by wax.

HFT was measured using a 1000-Hz probe tone presented at 85 dB SPL. Pressure was swept from 200 to −400 daPa at 300 daPa/s, slowing to 100 daPa/s around the peak of the trace. Traces that were difficult to interpret due to artefact caused by activity or noise were repeated if possible. Tympanograms were classified using the method of Kei et al. (2003). This method was chosen based on the recommendations of Hoffmann et al. (2013), who found it the best performing in a sample of 577 infants <12 months old. A tympanogram was classified as "pass" if there was a single or double peak, otherwise "fail".

DPOAEs were elicited in response to pairs of primary tones (f_1 and f_2), for f_2 of 2000, 3000, 4000, and 6000 Hz. The f_1 and f_2 intensity levels were set to 65 and 55 dB SPL, respectively, and the f_2 / f_1 ratio was 1.22. An ear was classified as pass if the SNR was ≥6 dB with an emission level ≥−10 dB, at three out of four f_2 frequencies, otherwise fail. If an ear failed with noise levels ≥0 dB SPL at two or more test frequencies, the test was repeated, if possible, to avoid failing normal ears solely because of high noise floor levels (Hunter et al., 2010).

Absorbance was measured at ambient pressure at 1/24 octave frequency resolution using a 226 to 8000 Hz broadband click delivered at 96 dB peSPL. Thirty-two clicks were presented to the ear and results averaged after removal of noisy responses as described by Liu et al. (2008). Results were monitored with a visual display during testing to check for air leaks. Testing was stopped and the probe reinserted if absorbance was high (≥0.3) at frequencies below 300 Hz (Groon et al., 2015). Absorbance data were obtained using the Interacoustics Titan Research Module which saves results as a text file after each test. As absorbance <0 is theoretically impossible, when this occurred, values were set to 0, as these results were likely due to calibration error (Piskorski et al., 1999). Although multiple WAI measures can potentially be used to develop a multivariate model (e.g., Ellison et al., 2012), other measures such as admittance were not considered in the present study, in order to restrict the number of candidate variables.

3.3.2 Reference standard

HFT and DPOAE results were used to evaluate middle ear function of each infant. HFT was chosen over low-frequency (226 Hz) tympanometry, as studies comparing the two have found HFT to be more accurate in infants under 9 months old (Hoffmann et al., 2013; Zhiqi et al., 2010). DPOAEs were used in conjunction with HFT to create a more rigorous reference standard than HFT in isolation (Kei & Zhao, 2012). Although evoked otoacoustic emissions (EOAEs) are a cochlear response, they can be used to assess middle ear function, as the stimuli and emissions are transmitted through the middle ear, and significant conductive dysfunction interferes with this process (Choi et al., 1999; Zhao et al., 2003). A limitation of using otoacoustic emissions to test for middle ear pathology is that they are affected by sensory, as well as middle ear disorders. However, in the present sample it is unlikely that emissions were absent due to sensory pathology, since all subjects either passed the newborn hearing screen, or follow-up diagnostic audiological assessment.

When results of two tests are used to create a reference standard, an ear can be classified as fail if it fails either test (the "loose" criterion), or both tests (the "strict" criterion) (Turner, Frazer, & Shepard, 1984). If the component tests have higher specificity than sensitivity, the loose criterion can be used to create a reference standard that improves sensitivity without sacrificing specificity (Alonzo & Pepe, 1999). The loose criterion was, therefore, chosen for the present study, as there is evidence indicating that both HFT and EOAEs have higher specificity than sensitivity for detecting middle ear dysfunction in infants and children (Driscoll et al., 2001; Hoffmann et al., 2013; Koike & Wetmore, 1999). Hoffmann et al. (2013) reported HFT specificity of 0.90 and sensitivity of 0.77 in a sample of 577 infants under 12 months old, using otomicroscopy as the reference test. Diagnostic performance studies

of transient evoked otoacoustic emissions (TEOAEs) in children have also reported higher specificity than sensitivity using typical clinical pass/fail criteria. Driscoll et al. (2001), in a sample of 940 children 5 to 7 years old, reported specificity of 0.86 and sensitivity of 0.60 using a pass criterion of ≥6 dB TEOAE SNR, against a reference standard consisting of puretone audiometry and 226-Hz tympanometry. With the same reference standard, Koike and Wetmore (1999) reported specificity of 0.94 and sensitivity of 0.84 in a sample of 63 children aged 4 to 17 years, defining pass as >50% TEOAE reproducibility.

3.3.3 Missing data and statistical modelling

Table 3.1 shows the number of ears that passed and failed HFT, DPOAEs, and the reference standard, and provides the number of missing observations for each test, including absorbance. Ears that failed one reference test with missing results for the other were not considered missing, as they were able to be classified by the loose criterion on the basis of failing one test. Results were missing due to infants crying, or not tolerating the probe in their ear, and were considered to be "missing at random" (Harrell, 2015). Since they made up only a small proportion (0.07) of ears tested, they were removed from the sample for subsequent analyses. In total, 34 ears were removed, due to either missing absorbance, or reference standard results. The dataset after removal of these observations is called the "full sample" in this report.

Table 3.1. Number of ears passed and failed the reference tests

Test	Pass	Fail	Missing
HFT	400	80	18
DPOAEs	396	81	21
Reference Standard	374	99	25
Absorbance			24

Number of ears with missing data for each test is also shown. DPOAEs, distortion product otoacoustic emissions; HFT, high-frequency tympanometry.

The probability that an ear had middle ear dysfunction was modelled using logistic regression. As this procedure assumes that observations are independent, one ear of each infant was randomly chosen for model development (the development sample), and opposite ears were used for validation (the validation sample). Infants with results for only one ear were put into the development sample to maximize this sample size (Myers et al., 2018a). Figure 3.1 shows the progression of infants and

observations through the study, including how many ears passed and failed the reference standard for the development and validation samples.

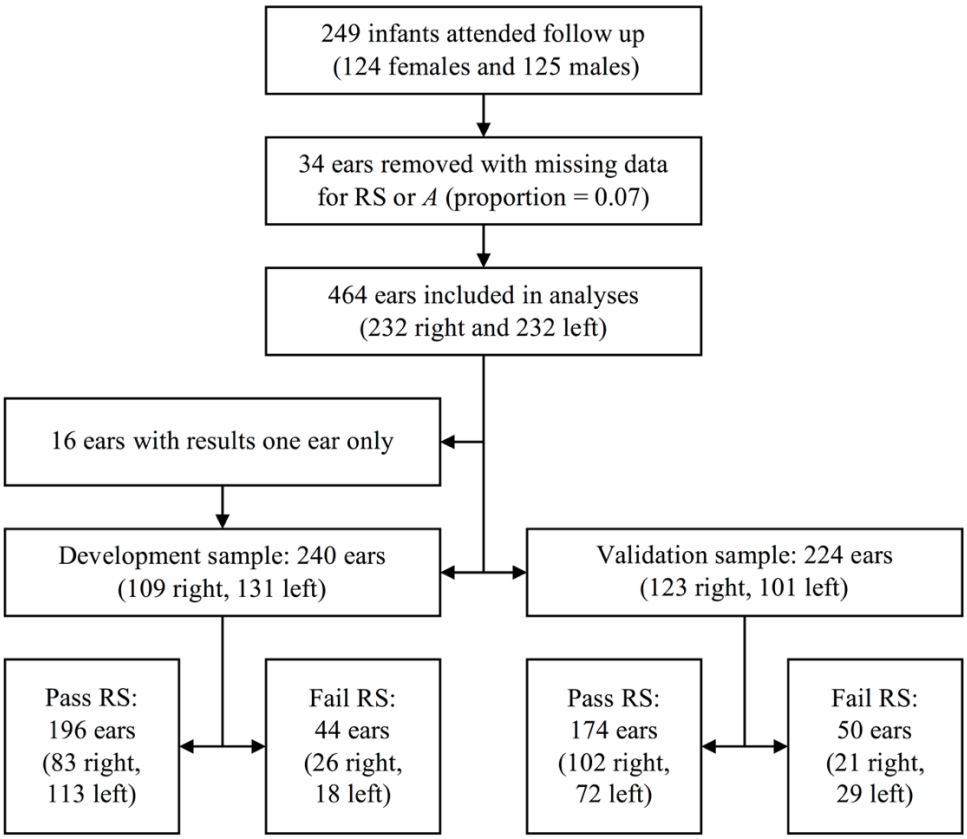


Figure 3-1. The flow of infants and observations through the study.
The pass and fail reference standard (RS) boxes (bottom row) provide the number of ears that passed and failed the reference standard for each of the samples. *A* = absorbance.

A general rule for fitting a logistic regression model is to have at least 10 observations in the smallest group (usually the fail group) for each variable included (Agresti, 2013). However, this is just a guide, and can be relaxed if the SNR is high. The shrinkage coefficient (γ) can be used to assess whether the signal is strong enough to warrant the number of predictors being modelled (Harrell, 2015):

$$\gamma = \frac{\text{model } \chi^2 - df}{\text{model } \chi^2}, \tag{3.1}$$

where, model χ^2 is the likelihood ratio χ^2 statistic (the statistical test for the model), and *df*, is the total degrees of freedom from all variables in the model. If γ suggests an issue with overfitting, further data reduction may be necessary.

As there were 44 ears that failed the reference standard in the development sample (Fig. 3.1), the 10:1 ratio suggests that the number of variables included in the model should be in the vicinity of four. Data reduction was therefore necessary, as at 1/24 octave frequency resolution, absorbance was measured at 107 frequencies. PCA was used to reduce the number of variables for modelling. The procedure transformed absorbance results into a new set of variables (PCs), ordered such that the maximum possible amount of variability was explained by the first PC (PC1), then as much of the remaining information possible was explained by the second PC (PC2), and so on. PCA did not automatically reduce the number of variables, because as many PCs were created as original variables. However, the number of predictors could be significantly reduced with minimal information loss, since most of the variability could be explained by the first few PCs. The number of PCs (starting at PC1 and adding PCs sequentially) that explained >90% of the variance were used as variables for modelling. For PCA calculations, absorbance data were first centred to have a mean of 0. It is common to also scale variables to have variance of 1, but this was not necessary for the present study because absorbance variables were already all on the same scale. Furthermore, scaling assumes that all predictors are equally important. However, previous research in infants has shown that certain frequencies in the absorbance response are more predictive than others, with the 1500 to 6000 Hz region the most diagnostically important (Ellison et al., 2012; Prieve et al., 2013b).

To make the composition of PCs more interpretable, the number of absorbance variables was reduced prior to PCA, by decreasing the frequency resolution. Reducing the number of variables made the PCs easier to interpret, since all variables (absorbance frequencies) contributed to each PC. For example, PCA on absorbance measured at 1/24 octave frequency resolution resulted in absorbance from all 107 frequencies contributing to each PC. However, this reduced to 11 variables when using 1/2 octave resolution. Rather than making an arbitrary decision, the choice of frequency resolution was systematically investigated to find the optimal resolution for modelling. First, absorbance results were averaged into 1/12, 1/6, 1/3, 1/2, and 1 octave frequency bands. Next, PCA was performed for each resolution. The number of PCs explaining >90% of the variance were then used as model variables. Since this usually resulted in more than 4 variables (PCs) being included (the limit suggested by the 10:1 rule), γ was calculated for the model at each frequency resolution, to assess whether overfitting was an issue. A value of >0.9 for γ was considered acceptable, as performance was not expected to be more than 10% poorer in new data (Harrell, 2015). Akaike's information criterion (AIC), a relative measure of model fit, was calculated for the model at each frequency resolution. A lower AIC indicates a better fitting model, if the models were fitted using the same dataset. The frequency resolution that resulted in the best fitting model was chosen as the one to proceed with.

Penalized maximum likelihood estimation was used to fit the models, with the aim of improving generalizability in new samples. The method used to choose the penalty was to maximize the modified AIC (AIC_M). AIC_M is model $\chi^2 - 2 \times$ effective *df*, where effective *df* is the degrees of freedom after penalization. In a standard model, each predictor "costs" one *df*. In a penalized model, effective *df* is lower than the number of predictors because the penalty reduces the number of parameters being estimated (Harrell, 2015). The penalty was chosen by performing a grid search over a range of possible values (0 to 5), and the value that maximized AIC_M was used as the penalty when fitting the model (Moons et al., 2004).

Including predictive covariates in a multivariate model can help to improve performance (Harrell, 2015). Given the equivocal nature of results from prior WAI research, univariate analyses of subject characteristics were performed, and any statistically significant covariates were included as multivariate predictors, along with absorbance PCs, to assess whether this improved model fit. Univariate analyses were performed for gender, age, ethnicity, and ear side. These analyses were done using the development sample, with missing observations omitted. Subject characteristics for the development sample are summarized in Table 3.2, including the number of missing observations for each covariate. For multivariate analyses, covariates with missing observations were given a value using multiple imputation so that all analyses could be performed with the same dataset, without having to remove observations with complete absorbance and reference standard results, missing only covariate data (Buuren & Groothuis-Oudshoorn, 2011). This resulted in four values being imputed for ethnicity. A model was also fitted from the development sample on the best univariate predictor. This was done to compare with the multivariate models to test whether the additional information improved model fit, and/or performance.

Table 3.2. Characteristics of neonates included in the development sample

Characteristic	Value
Age (weeks)	20 missing
Median (IQR)	27.7 (26.6–29.5)
Range	23.9–39.9
Gender (count)	0 missing
Female (%)	120 (50)
Male (%)	120 (50)
Ethnicity (count)	1 missing
Caucasian (%)	207 (86)
Asian (%)	29 (12)
Other (%)	3 (1)
Ear (count)	0 missing
Right (%)	109
Left (%)	131

The number of subjects with missing data for each characteristic is provided in the Value column. IQR, interquartile range.

The model with the lowest AIC was taken as the final model for further evaluation and validation (Burnham & Anderson, 2002). When evaluating a prediction model, it is important to assess both discrimination and calibration. Discrimination measures how well the model differentiates between normal and diseased ears, and calibration assesses the quality of predictions. The performance of the best fitting model was evaluated with AUC, and calibration with calibration curves, which plot actual against predicted probabilities. The model was validated by applying the coefficients to the validation sample, and also with bootstrap resampling. Bootstrapping involved sampling with replacement from the development sample, a "training sample", the same size as the development sample. A model was fitted to the training sample, and the coefficients applied to the original development sample (the "test sample"). The amount of bias was estimated by calculating the difference in performance measures (AUC and calibration) between the training and test samples. The process was repeated 500 times and averaged to give a stable estimate of the amount of bias in the model (Steyerberg et al., 2001c). The estimated bias was subtracted from final model's performance measures to provide a bias-corrected estimate of future model performance (Steyerberg, 2008). The best univariate predictor model was also applied to the validation sample for comparison with the final multivariate model. The difference in AUC between the univariate and multivariate models was tested statistically using the method of DeLong, DeLong, and Clarke-Pearson (1988).

Statistical modelling was done with R (R Core Team, 2017), expanded with the *rms* library for regression modelling (Harrell, 2016). This report has been written to conform to the recommendations of the TRIPOD (transparent reporting of a multivariable prediction model for individual prognosis or diagnosis) statement for reporting multivariate clinical prediction models (Collins et al., 2015).

3.4 Results

3.4.1 Absorbance data and the reference standard

Median absorbance for ears that passed the reference standard are compared with results from previous normative studies in Figure 3.2. The general shape of the response is similar to other studies with absorbance low in the low and high frequencies, and generally highest from 2000 to 6000 Hz, with a dip at 3000 Hz. However, with the exception of Aithal et al. (2014b), the magnitude of absorbance in the present study was generally smaller than for other studies.

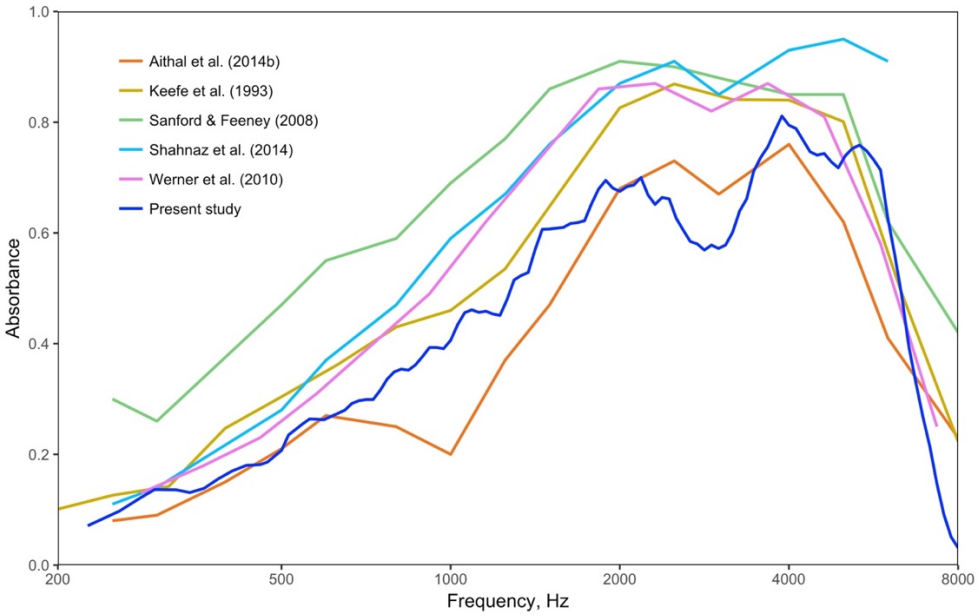


Figure 3-2. Median absorbance of normal ears compared with other studies.
Aithal et al. (2014b) reported on fourteen 6-month-old infants (27 ears); Keefe et al. (1993), eleven infants aged 6 months; Sanford and Feeney (2008), twenty subjects aged 6 months; Shahnaz et al. (2014), thirty-three ears from 6-month-old infants; and Werner et al. (2010), two hundred and sixty subjects aged 5 to 9 months. Aithal et al. reported median absorbance; the other studies mean reflectance which has been converted to absorbance for comparison. Data from the present study are from the full sample, displayed at 1/24 octave frequency resolution.

Absorbance median and IQR for ears that passed and failed the reference standard are depicted in Figure 3.3. Median absorbance for the fail group was lower than the pass group across the entire frequency range. Median absorbance for the pass group was higher than the third quartile of the fail group from 1000 to 2500 Hz, and 5000 to 7000 Hz. IQRs were almost completely separated from 1200 to 2000 Hz.

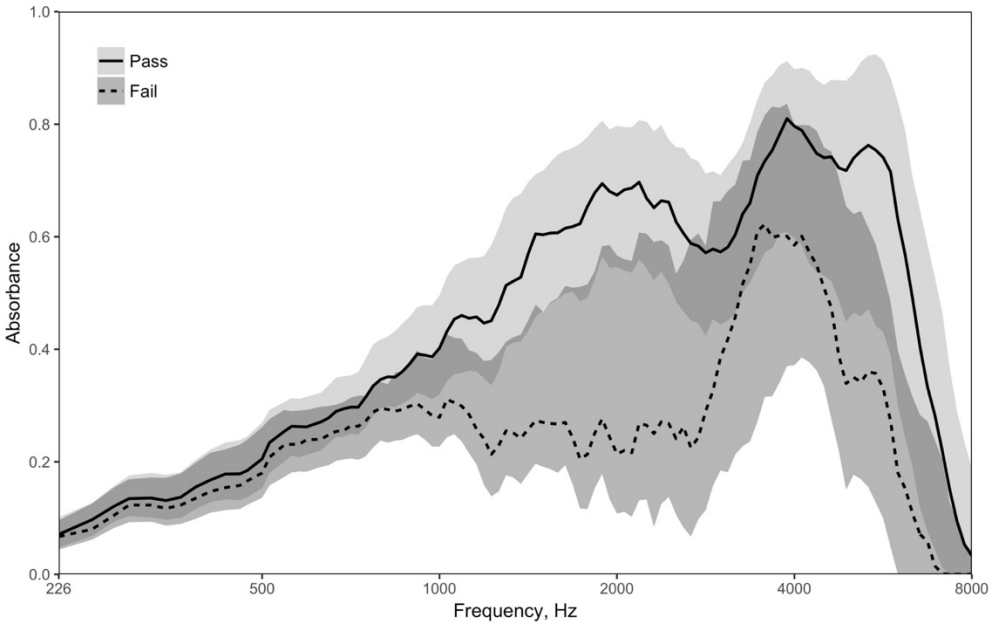


Figure 3-3. Median and interquartile range absorbance for the pass and fail groups.
The solid and dashed lines show the median, and light and medium grey shaded areas depict the interquartile range (IQR) for the pass and fail groups, respectively. The dark grey shading depicts where IQRs of the groups overlapped. Data were taken from the full sample and plotted at 1/24 octave frequency resolution.

The reference standard assumed that ears failing only one test had middle ear pathology. If these ears were not false positives, and were actually dysfunctional, we would expect average absorbance to be lower for these ears than the normal group. To assess this, median absorbance for ears that passed both HFT and DPOAEs (n = 370), failed both tests (n = 58), failed DPOAEs only (n = 19), and failed HFT only (n = 17) are shown in Figure 3.4. Compared to the normal group, ears that failed DPOAEs only had lower absorbance from 1800 to 8000 Hz, and measurements in ears that failed HFT only were lower over almost the entire frequency range (except for around 3500 Hz). Ears that failed both tests generally had lowest absorbance compared to the other groups from 700 to 7000 Hz.

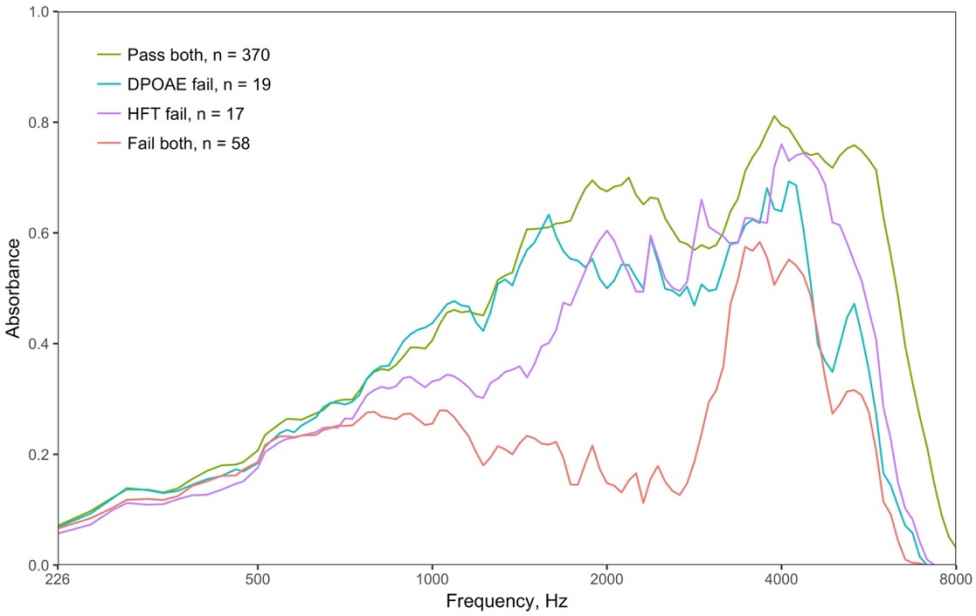


Figure 3-4. Absorbance stratified by reference test results.
Median absorbance for ears that passed both tests in the reference standard, failed both tests, failed only DPOAEs, and failed only HFT. Data were taken from the full sample, and are shown at 1/24 octave frequency resolution. DPOAE = distortion product otoacoustic emissions; HFT = high-frequency tympanometry.

3.4.2 Data reduction

Results of the PCA models for choosing the frequency resolution are shown in Table 3.3. The number of PCs needed to explain over 90% of the variance is provided, along with γ and AIC. Generally, more PCs were needed to explain over 90% of the variance as the frequency resolution increased. Four PCs explained 96% of the variance at 1 octave frequency resolution, but 6 PCs were needed to explain 92% at 1/24 octave resolution. The cumulative proportion of variance explained by the 1/2 octave resolution PCs is depicted in Figure 3.5. Half of the variance was explained by the first PC, increasing to 0.88 with four PCs, and then 0.94 with five. Apart from 1 octave, all models needed more than 4 variables (PCs) to explain over 90% of the variance. However, γ was acceptable (>0.9) for models at all frequency resolutions, indicating that overfitting was not of concern. Since the 1/2 octave model had the lowest AIC (indicating that it was the best fitting), it was chosen as the frequency resolution to proceed with (Table 3.3).

Table 3.3. Principal component analysis results for various frequency resolutions

f resolution	PCs	Variance	γ	AIC
1 octave	4	0.96	0.94	188.82
1/2 octave	5	0.94	0.95	155.03
1/3 octave	5	0.92	0.94	180.19
1/6 octave	6	0.96	0.94	174.45
1/12 octave	6	0.92	0.94	161.03
1/24 octave	6	0.92	0.93	178.59

The number of principal components (PCs) included, proportion of variance explained, γ, and AIC for the PCA models fitted using various frequency (f) resolutions. AIC, Akaike's information criterion; γ, shrinkage coefficient; PCA, principal component analysis.

The loadings for the first five PCs of the 1/2 octave PCA are provided in Table 3.4. Loadings are rounded to three decimal places for ease of interpretation. The actual loadings for all 11 PCs, as well as the centring factors, are provided in Section 3.6: Appendix A. The loadings show the degree to which absorbance at each frequency contributes to a given PC. A larger absolute value for a loading indicates a greater contribution to the PC by absorbance at that frequency. Loadings can be positive or negative, but the sign is arbitrary. Figure 3.6 is a visual depiction of the loadings, showing the contribution that absorbance at each frequency makes to each PC. The biggest contributors to PC1 were absorbance variables from 2000 to 8000 Hz. For PC2, 1000 to 2000, and 5657 Hz contributed the most.

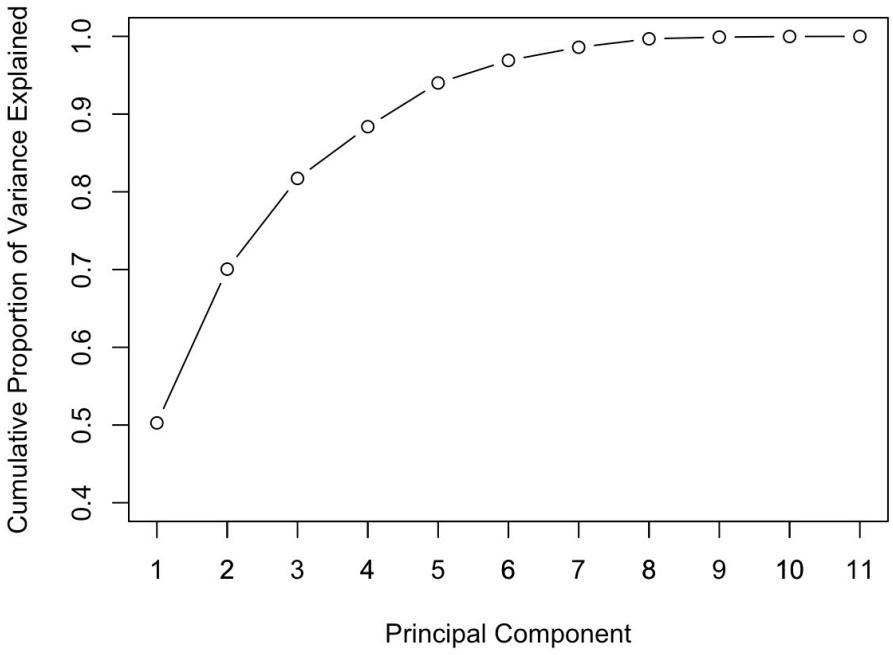


Figure 3-5. Cumulative proportion of principal components.
Cumulative proportion of variance explained (*y*-axis) by the principal components from the 1/2 octave frequency resolution principal component analysis (*x*-axis).

The effect of the choice of frequency resolution was explored in Figure 3.7. The top panel (A) depicts univariate AUC for 1/24 (the raw data) and 1/2 octave frequency bandwidths. The general shape of the AUC as a function of frequency is the similar for both resolutions. AUC is highest in the 1500 to 2000 Hz region, with a secondary peak around 6000 Hz. Highest AUC at 1/24 octave frequency resolution was 0.821 at 1682 Hz, and at 1/2 octave, 0.805 at 2000 Hz. The bottom panel (B) shows median absorbance for the pass and fail groups plotted at 1/24, 1/3, 1/2, and 1 octave bandwidths. The 1/3 and 1/2 octave resolutions show a smoothing effect compared to the raw data (1/24 octave), but the shape of the response is very similar for all three resolutions. The 1 octave bandwidth, however, tends to depart from the raw data at frequencies above 2000 Hz, especially for the pass group.

Table 3.4. Loadings for the first five principal components

Variable	PC1	PC2	PC3	PC4	PC5	CF
250 Hz	0.049	-0.102	0.177	-0.242	0.064	0.122
354 Hz	0.066	-0.134	0.236	-0.323	0.084	0.168
500 Hz	0.080	-0.171	0.275	-0.345	0.069	0.244
707 Hz	0.074	-0.250	0.281	-0.235	0.089	0.324
1000 Hz	0.067	-0.436	0.257	-0.025	0.193	0.409
1414 Hz	0.180	-0.608	0.037	0.333	0.063	0.511
2000 Hz	0.415	-0.363	-0.383	0.257	-0.316	0.595
2828 Hz	0.401	-0.003	-0.449	-0.580	-0.200	0.557
4000 Hz	0.415	0.175	-0.185	-0.120	0.591	0.677
5657 Hz	0.579	0.349	0.302	0.364	0.216	0.529
8000 Hz	0.323	0.189	0.465	-0.044	-0.632	0.183

Loadings for the first five PCs and centring factors (CF) for the absorbance variables for the 1/2 octave frequency resolution PCA. The sign of the loadings is arbitrary, and can be ignored when considering the size of a variable's contribution to a PC (i.e., consider only the absolute value). The CF column provides the centring factor used to set the mean absorbance to 0 for each frequency. The CF was subtracted from the absorbance value at a given frequency prior to performing the PCA analysis. PC, principal component; PCA, principal component analysis.

3.4.3 Covariate analyses

Results of the univariate covariate analyses are presented in Table 3.5. Ear side was significantly associated with the outcome ($p = 0.04$), and was therefore included as a candidate covariate for multivariate modelling. Figure 3.8 shows the effect of ear side (left panel) on predicted risk of middle ear dysfunction. Right ears had higher risk of middle ear dysfunction (0.24) compared to left (0.14). Although statistically significant, the size of the effect is fairly small (difference in risk between the ears of 0.1).

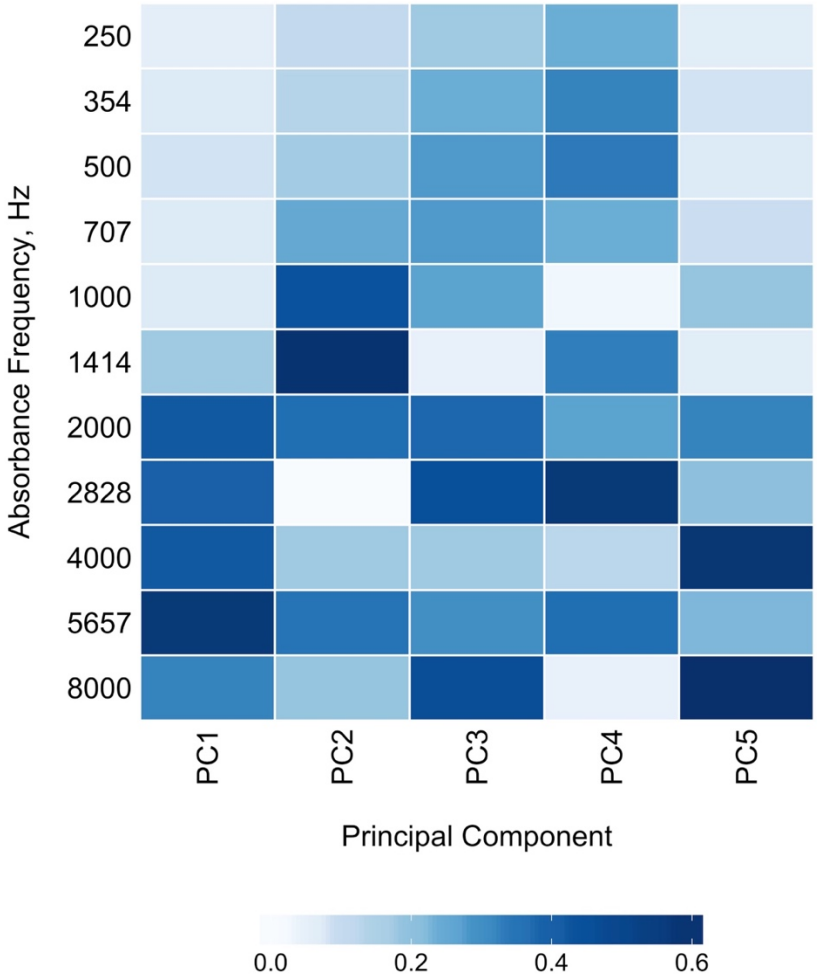


Figure 3-6. Principal component heatmap.
Heatmap of the absolute value of the loadings for the first five principal components (PC; *x*-axis) for the absorbance frequencies (*y*-axis) from the 1/2 octave resolution principal component analysis. Darker blue in a cell indicates a larger absolute value.

The effect of ear side on the absorbance response was explored in Figure 3.9 which depicts median absorbance for the pass and fail groups by ear side. Overall, median absorbance for ear side was very similar for right and left ears for both the pass and fail groups. Right ears that failed the reference standard did have lower median absorbance at the most important predictive frequencies (1500 to 2000 Hz, and 6000 Hz), but only slightly.

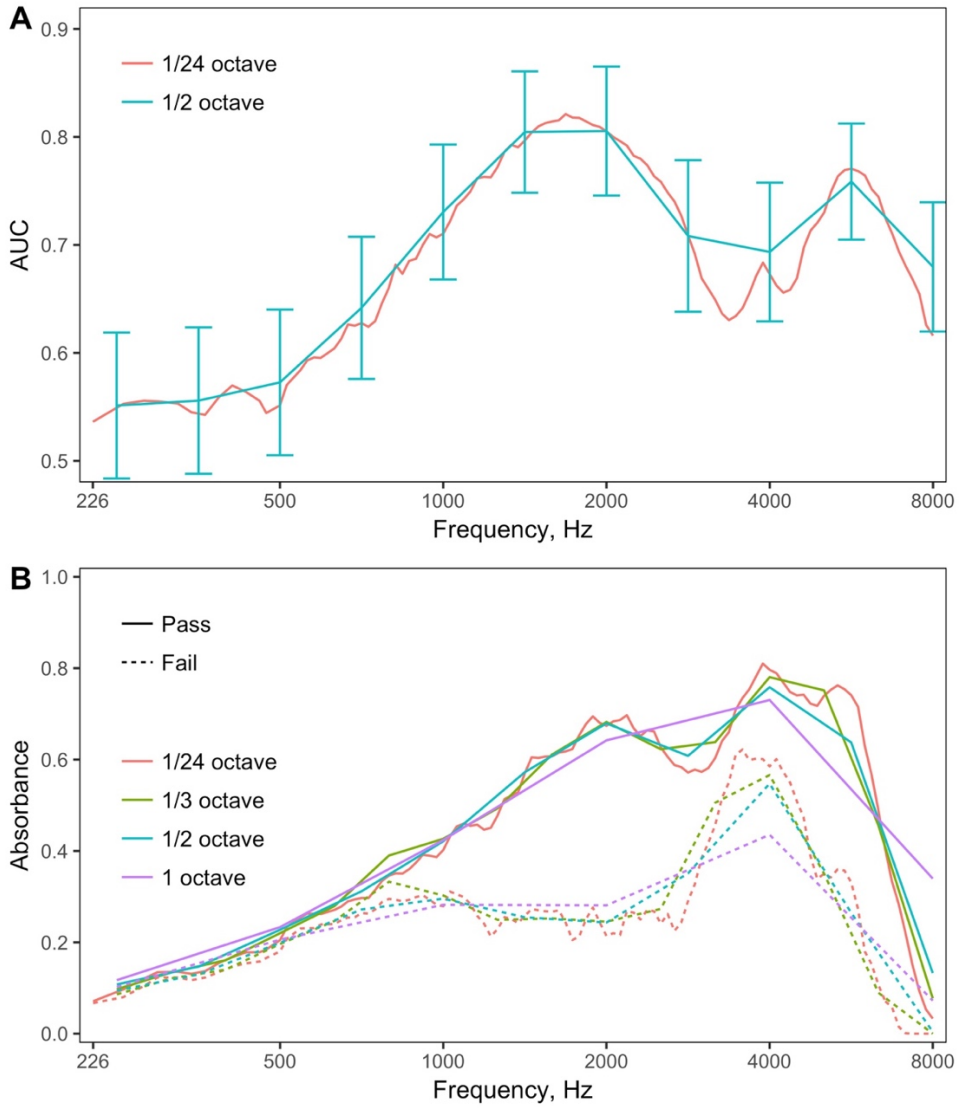


Figure 3-7. AUC and the effect of frequency resolution.
The effect of frequency resolution on univariate absorbance AUC (A, top), and median absorbance for the pass and fail groups in the full sample (B, bottom). The 95% confidence interval bars are provided for the 1/2 octave AUC results. AUC = area under the receiver operating characteristic curve.

3.4.4 Statistical modelling

Details of the fitted models, including model χ^2, AIC, and penalties are provided in Table 3.6. Model A was fitted on the best univariate predictor at 1/2 octave frequency resolution (2000 Hz). This was to compare with the multivariate models, to evaluate whether there was a multivariate advantage. The other models all used the first five PCs from the 1/2 octave PCA as predictors. Model B used only the absorbance PCs, but Model C also included ear side, the significant covariate, to assess whether this improved model fit. AIC for model C was not lower than Model B (156.77 compared to 155.03; Table 3.6), indicating that including ear side did not improve model fit. Therefore, since Model B was best fitting, it was taken as the final model for interpretation and validation.

Table 3.5. Statistics from the univariate covariate analyses

Covariate	LR χ^2	*p*-value	AUC
Age	0.69	0.405	0.551
Gender	2.80	0.094	0.570
Ear side	4.05	0.044	0.584
Ethnicity	5.66	0.059	0.580

Model LR χ^2 (the statistical test), associated *p*-values, and AUC results for the univariate covariate analyses. These models were fitted using the development sample with missing observations omitted. AUC, area under the receiver operating characteristic curve; LR χ^2, model likelihood ratio chi-squared statistic.

AUCs for Model B for the development, bootstrapped and validation samples are shown in Table 3.7. AUC for the development sample was 0.887. Bootstrapping estimated bias of 0.013, leaving a bias-corrected AUC of 0.874. Applying the model to the validation sample resulted in AUC of 0.852. AUC for Model A (the model fitted from the best univariate predictor) for the development and validation samples is also presented for comparison (Table 3.7). Model B had higher AUC than Model A for both the development and validation samples (0.852 compared to 0.785 for the validation sample). The difference in AUC between the models when applied to the validation sample was statistically significant ($p = 0.008$).

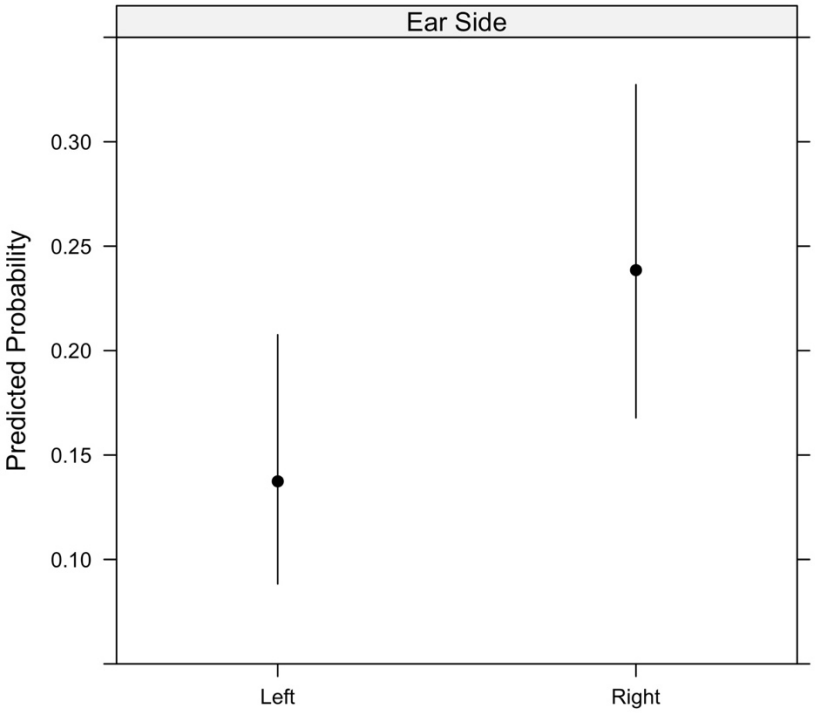


Figure 3-8. The effect of ear side on predicted probability of middle ear dysfunction.
The error bars depict the 95% confidence intervals.

Calibration curves for Model B for the development, bootstrapped and validation samples are shown in Figure 3.10. Risk estimates were close to the ideal line for the development and bootstrapped samples, with predictions slightly low for probabilities above 0.6. For the validation sample, predictions were a little low across the entire range of probabilities, more so for probabilities between 0.1 and 0.2. Overall, calibration curves were satisfactory for all samples. The equation for Model B to calculate risk predictions is provided in Section 3.7: Appendix B. A web application implementing Model B is available (https://joshmyers.shinyapps.io/WAIPredictions/) that can be used to make predictions for an infant either by manually entering absorbance values, or uploading a file exported from an Interacoustics Titan Device.

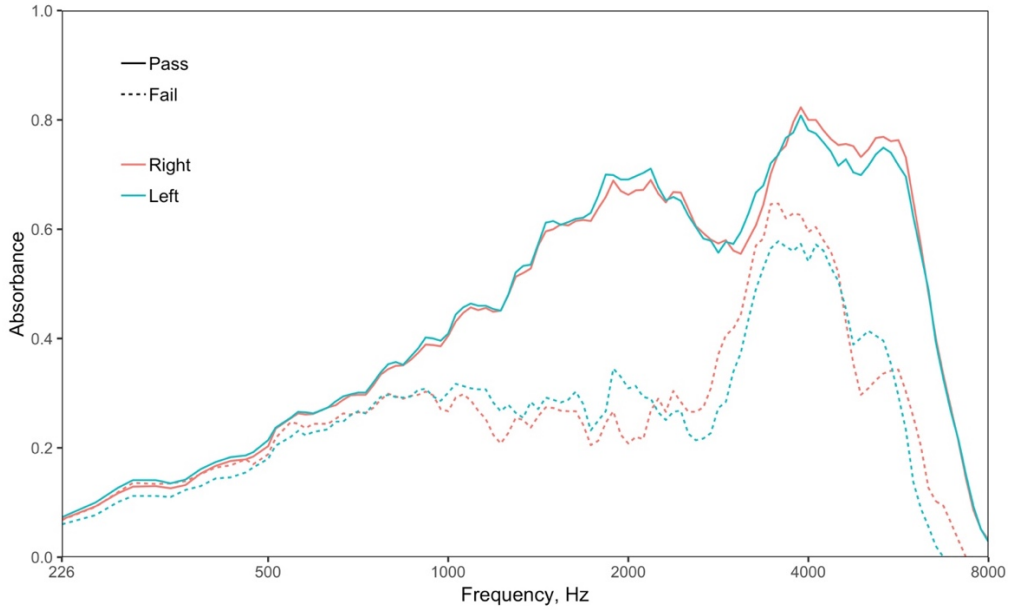


Figure 3-9. Median absorbance for the pass and fail groups by ear side.
Data are from the full sample plotted at 1/24 octave frequency resolution.

3.5 Discussion

3.5.1 Clinical application of the model

Predicted probabilities from the final model (Model B) could potentially be used clinically in both screening and diagnostic settings. In the context of screening, a referral threshold must be chosen. Previous WAI studies have recommended cut-offs based on visual inspection of results (Prieve et al., 2013b), or statistically, using the ROC curve (Keefe et al., 2012; Piskorski et al., 1999; Sanford et al., 2009). From a statistical perspective, the most efficient cut-off is the point at the top left hand corner of the ROC curve (Youden, 1950). The point where sensitivity and specificity are close to being equal (the "point of symmetry") has also been suggested as a potentially useful cut-off in diagnostic WAI research (Keefe et al., 2012; Sanford et al., 2009).

Table 3.6. Statistical results for the fitted models

	Variables in Model	LR χ^2	Penalty	*df*	γ	AIC
Model A	2000 Hz	64.88	0.50	0.98	0.98	168.73
Model B	First 5 PCs (1/2 octave *f* resolution)	88.52	1.11	4.75	0.95	155.03
Model C	Model B variables plus ear side	89.55	1.57	5.53	0.94	156.77

Variables included, model LR χ^2 (the statistical test), penalty, degrees of freedom (*df*), shrinkage coefficient (γ), and AIC for the fitted models. *p*-values for the model LR χ^2 statistic were <0.001 for all models. The models were fitted using the development sample with missing observations for covariates imputed. The penalty was chosen by performing a grid search over a range of possible penalty values (0 to 5), and the point that maximized AIC_M was used as the penalty. Model B is the same model as the 1/2 octave resolution model in Table 3.3. AIC, Akaike's information criterion; AIC_M, modified Akaike's information criterion; LR χ^2, model likelihood ratio chi-squared statistic.

An issue with the above strategies is that the threshold is chosen based on the properties of the test. The decision-analytic approach, on the other hand, takes the cost of correctly identifying, or missing a case into consideration when choosing the threshold (D. J. Hand in discussion to Briggs & Zaretzki, 2008; Steyerberg et al., 2011). However, formal decision analysis, where the utility (harm or benefit) of every possible outcome is considered, can be difficult to implement clinically, due to its complexity (Metz, 1978; Vickers, 2008). Using predicted risk to set the threshold simplifies the process, as only one value needs to be assigned (Vickers, 2008). The threshold is the point where there is enough concern (i.e., high enough risk) to warrant further action (e.g., review, or referral). It reveals the value associated with the benefits of accurately identifying the condition (true positives) compared to the harm (or cost) from unnecessary follow up (false positives). In the context of screening for middle ear dysfunction in infants, the benefits relate to timely intervention for early onset otitis media, and the costs include increased pressure on medical, audiology, and otology services to follow up infants who refer from the program. The threshold (T) is the point where the harms and benefits are equal (Van Calster et al., 2013). The odds of the chosen threshold (T/1−T) correspond to the perceived harms compared to the benefits (Vickers, 2011). For example, setting the threshold at 0.5 suggests that true and false positives are valued equally. Using a risk threshold of 0.6, implies that false positives "cost" 1.5 times more than true positives (0.6/1−0.6). The threshold is flexible, and can be adjusted depending on the goals of the screening program.

Table 3.7. Discrimination and internal validation results

	Sample	AUC [95% CI]
Model B	Development	0.887 [0.826–0.948]
	Bootstrap training	0.893
	Bootstrap test	0.880
	Bias (bootstrap training – test)	0.013
	Bias-corrected	0.874
	Validation	0.852 [0.784–0.920]
Model A	Development	0.828 [0.746–0.909]
	Validation	0.785 [0.698–0.873]

AUC for the development, bootstrapped, and validation samples for Model B and Model A. AUC, area under the receiver operating characteristic curve; CI, confidence interval.

Once a threshold is chosen, statistics such as sensitivity and specificity can be calculated to see how well the model performs at the chosen threshold (Steyerberg, 2008). Sensitivity, specificity, and predictive values for various risk thresholds from Model B are provided in Table 3.8. Thresholds were calculated from the development sample, and then applied to the validation sample. Test performance statistics are provided for the point of symmetry, and cut-offs corresponding to threshold odds of 0.5 (T = 0.33), 1 (T = 0.5) and 1.5 (T = 0.6).

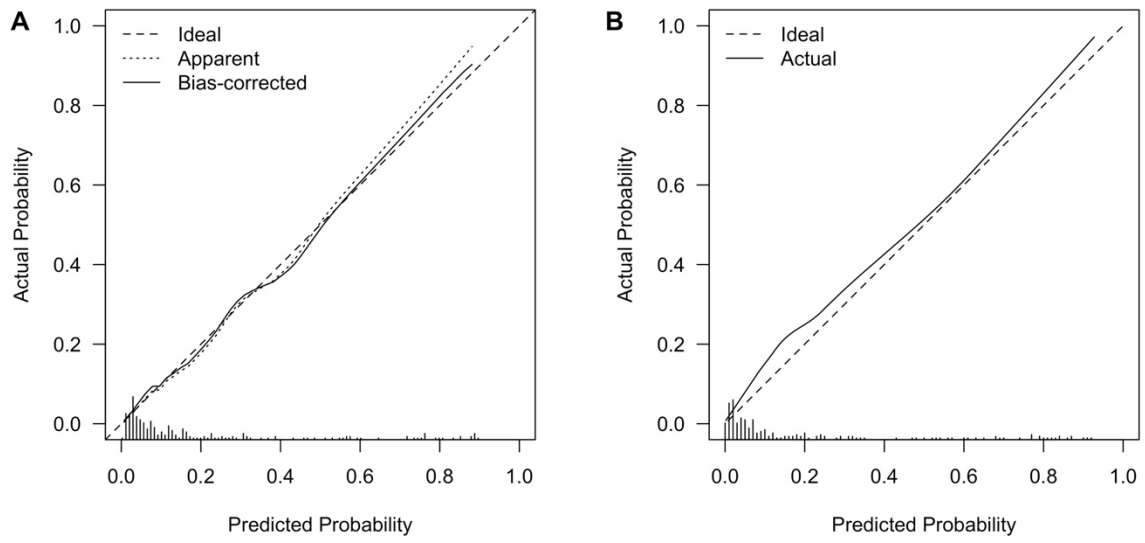


Figure 3-10. Calibration curves for Model B.
Actual versus predicted probability for Model B for: A, the development and bootstrapped (bias-corrected) samples; and B, the validation sample. Perfect (ideal) calibration is shown by the diagonal dashed lines, and the histograms inside the *x*-axes depict the distributions of predictions.

The cut-off for the point of symmetry was probability of 0.16. This threshold seems too low in the context of diagnosing middle ear pathology, as 16% risk of middle ear dysfunction in an infant hardly seems concerning enough to warrant further action. Choosing this threshold implies that true positives are 5.25 times more valuable than false positives (1−0.16/0.16). The resulting positive predictive value of 0.49 means that a clinician would be wrong approximately half the time if using this cut-off to make a positive diagnosis. As a starting point, a probability threshold of 0.5 (50% risk of middle ear dysfunction) seems reasonable, since at this point middle ear dysfunction begins to be more likely than not, and a clinician would probably want to follow this up. However, the cut-off is flexible, and can be adjusted depending on the goals of the screening program. One program may set the refer threshold at 0.5, but another, in a healthcare system with fewer resources may decide to set it higher to limit the burden on services (Myers et al., 2018a). Another consideration is that other tests (e.g., EOAEs) might be used in conjunction with absorbance to create a screening test battery. In this case, predicted risk of middle ear dysfunction could be used along with other results to create decision rules for further action. For example, a program may refer an infant directly to audiology if EOAEs are absent, and risk of middle ear pathology low. Alternatively, an infant with high risk of middle ear pathology and absent EOAEs might be scheduled for review.

Table 3.8. Diagnostic accuracy statistics at various thresholds

Threshold	Description	Sp	Se	NPV	PPV
0.16	Se ~ Sp	0.78	0.74	0.91	0.49
0.33	$Odds_T = 0.5$	0.93	0.60	0.89	0.70
0.50	$Odds_T = 1.0$	0.95	0.56	0.88	0.76
0.60	$Odds_T = 1.5$	0.98	0.52	0.88	0.87

Specificity (Sp), sensitivity (Se), and predictive values for various risk thresholds from Model B. The thresholds were calculated from the development sample and then applied to the validation sample. NPV, negative predictive value; $Odds_T$, odds at the threshold; PPV, positive predictive value.

In a diagnostic context, predicted probabilities from Model B could be used in conjunction with visual depiction of absorbance to provide individualized diagnoses. Clinical reasoning is inherently probabilistic and providing results as predictions complements this process (Sox, Higgins, & Owens, 2013). If estimated risk is high or low, a clinician may confidently make a diagnosis, and an indeterminate result (e.g., probability = 0.5), may necessitate further investigation or review. Using the model to supplement visual inspection of the data could be especially helpful for inexperienced clinicians, but even practiced users may benefit from the additional objective information. Importantly, the diagnostic decision point is flexible, and can be adjusted depending on the clinical context. For example, in a routine assessment with no parental concern, a clinician may decide to discharge if the result is borderline (probability = 0.5). However, if there are concerns about development, the same result may warrant review. In the context of the audiological test battery, if hearing thresholds are elevated, EOAEs absent, and probability of middle ear dysfunction high, a clinician may suspect conductive hearing loss. Alternatively, the same hearing thresholds and EOAE results with low probability of middle ear dysfunction may alert the clinician to a possible sensorineural hearing loss (Blankenship et al., 2018).

3.5.2 Model development and validation

AUCs for Model B for the bias-corrected and validation samples were within the 95% confidence interval (CI) of the development sample, and the calibration plots for the development, bootstrapped, and validation samples all showed high agreement between predicted and actual probabilities. These results indicate that Model B may generalize well to new infants. Model B had a clear advantage over the best univariate model (Model A) on all performance measures (AIC and AUC). Previous studies have demonstrated a multivariate performance advantage for diagnosing conductive hearing loss in children (Keefe et al., 2012; Piskorski et al., 1999), and conductive conditions in neonates (Myers et

al., 2018a), but the present study showed this also extends to diagnosing middle ear pathology in 6- to 9-month-old infants.

Using 1/2 octave frequency averaging significantly reduced the number of absorbance variables (107 to 11), but did not result in substantial information loss, as the top 1/2 octave frequency AUC was only 0.016 lower than the most accurate frequency at 1/24 octave resolution. AIC for the 1/2 octave model was lower than 1/24 octave, indicating that the reduced frequency resolution actually improved model fit. PCA was also an effective data reduction method that resulted in interpretable PCs. The best performing absorbance frequencies from both the univariate analyses, and findings from previous research were well represented in the PCs used for modelling. The most predictive absorbance frequencies from the univariate analyses were 1414, 2000 and 5657 Hz, with frequencies ≥1000 Hz having better performance than lower frequencies. Prior research has identified absorbance from 1500 to 6000 Hz as the most important region diagnostically (Ellison et al., 2012; Prieve et al., 2013b). All of the important absorbance frequencies identified in the univariate analyses, and prior research contributed significantly to the first five PCs from the 1/2 octave PCA used to fit Model B.

3.5.3 Absorbance and reference standard results

The AUC for Model B for the development sample in the present study of 0.887 (95% CI, 0.826 to 0.948) was not as high as Ellison et al. (2012), who reported AUC of 0.93 for absorbance. However, the AUC 95% CI of the present study included the result of Ellison et al. Overall, the median and IQR for the pass and fail groups from the present study compared favourably with results from Ellison et al., although there were some differences. The median for the fail group in the present study was not as low as Ellison et al. for frequencies between 500 and 2000 Hz, and differences between reference standards used in the studies may have contributed to this. Ellison et al. included only ears with confirmed middle ear effusion in the fail group, whereas the reference standard in the present study was designed to identify ears with even mild middle ear dysfunction. The IQRs for Ellison et al. had better separation between pass and fail groups from 250 to 1000 Hz, but not as clear separation from 5000 to 6000 Hz as was seen in the present study. Other factors that may have contributed to the differences between the studies include age of subjects and equipment used to measure WAI. Median age of subjects in Ellison et al. was substantially older than the present study (12 months old, compared to 6 months), and they used a prototype system for WAI measurements, whereas the present study used the commercially available Titan system.

Median absorbance for ears that failed only one test in the reference standard had lower absorbance than ears that passed both tests at many of the diagnostically important frequencies (Fig. 3.4;

Ellison et al., 2012; Prieve et al., 2013b). Previous research has demonstrated that ears with mild dysfunction (e.g., negative tympanic peak pressure on tympanometry, or moderately stiff eardrums on pneumatic otoscopy), have average absorbance that falls between normal, and ears with effusion in the 1000 to 5000 Hz region (Beers et al., 2010; Ellison et al., 2012; Hunter et al., 2008b). In the present study median absorbance for ears that failed one test fell between ears that passed or failed both tests over much of the same region, indicating that these ears had milder dysfunction than ears that failed both tests.

Ears that passed the reference standard in the present study had median absorbance similar to normative results from other studies (Fig. 3.2), although average absorbance was slightly lower in general, with the exception of Aithal et al. (2014b). As mentioned, possible factors such as age, equipment, and reference standard may account for differences.

3.5.4 Strengths, limitations, and directions for future research

Since the study sample was recruited from the general population, predicted probabilities from Model B would be most suitable for use in a similar population, such as screening in conjunction with a health check at 6-month-old immunizations. Although potentially useful in diagnostic or high-risk screening contexts also, the model may need to be updated to reflect the prevalence and typical disease severity in the new setting, as the prevalence is often higher, and disease more severe in these contexts (Moons et al., 2012a). For example, the model may need updating if being used in an otology clinic, as ear disease would likely be higher in prevalence, and more severe than typically seen in the general population.

The modelling strategy aimed to develop a model likely to generalize to new infants through limiting the number of predictors with data reduction, applying a penalty when fitting, and internally validating using two methods, to assess for overfitting. However, external validation of the model is needed to assess the degree to which differences in subject characteristics, equipment and environmental factors affect model performance (Moons et al., 2012a). Further research could evaluate the performance of Model B in a new sample of infants not used in model development. An external validation study would apply the centring factors and loadings from Section 3.6: Appendix A on absorbance results from a new sample, and then use the equation in Section 3.7: Appendix B to calculate predictions, updating the model as appropriate (Moons et al., 2012a).

A limitation of the present study was that data collection was not blinded. Both the reference tests and absorbance were measured by the same research audiologist. However, interpretation of the reference tests was objective, either meeting the pass criteria or not. Being able to complete all of the tests with a single insertion of the probe enabled results to be obtained quickly, resulting in a high

success rate for both absorbance and the reference tests in a difficult-to-test population (Shahnaz et al., 2014).

3.5.5 Summary and conclusions

The aim of this study was to develop a prediction model for diagnosing middle ear dysfunction in 6- to 9-month-old infants using wideband absorbance. Methods for data reduction, penalization and internal validation were applied, with the aim of creating a model likely to generalize to new infants. The model performed well on both measures of discrimination and calibration, and may be clinically useful. In a screening setting, predicted probabilities offer an intuitive and flexible mechanism for setting the referral threshold that takes into account the costs associated with true and false positives. In a diagnostic context, risk estimates could be used to supplement subjective interpretation of absorbance results for individualized diagnosis of middle ear pathology. Future research assessing the performance and impact of Model B in these contexts is warranted.

3.6 Appendix A: Principal component analysis loadings

The principal component analysis loadings and centring factors from Chapter 3.

	Principal Components 1 to 4			
Variable	PC1	PC2	PC3	PC4
250 Hz	0.0489836914566045	-0.10243946467333800	0.1771267479374100	-0.2424551835520600
354 Hz	0.0663486542738451	-0.13433669878202500	0.2357612482449830	-0.3231430134155570
500 Hz	0.0804672893382117	-0.17054970229456500	0.2750095104975020	-0.3449762895962940
707 Hz	0.0736980781612132	-0.25020948253632700	0.2807124529759180	-0.2345448527166250
1000 Hz	0.0666297381024985	-0.43590476463258000	0.2574997502514360	-0.0246520981772233
1414 Hz	0.1797435440830960	-0.60792929080639200	0.0368068444677677	0.3332597336042940
2000 Hz	0.4149698049365220	-0.36276695151122000	-0.3832962063068450	0.2565032260200490
2828 Hz	0.4010412337127480	-0.00273314771963196	-0.4494844247560330	-0.5802458175060310
4000 Hz	0.4148312121592830	0.17514193947667900	-0.1849045775278440	-0.1201808059206920
5657 Hz	0.5789888313208650	0.34926721285388500	0.3020703095587450	0.3636156898300010
8000 Hz	0.3227968207825770	0.18949200178531200	0.4653641231518370	-0.0443335409051452

	Principal Components 5 to 8			
Variable	PC5	PC6	PC7	PC8
250 Hz	0.0638377562597850	-0.2525423602344350	0.0386821460040066	-0.1775292523861350
354 Hz	0.0840478103032062	-0.3635941960062980	0.0451338570879222	-0.2235629632959160
500 Hz	0.0687641822818827	-0.3510140653952520	-0.0192572892379463	-0.0755906970367255
707 Hz	0.0893750248844870	-0.0131558272513994	-0.1104780771210280	0.3546463108172130
1000 Hz	0.1925244163245900	0.4510064360587900	-0.1564978991066850	0.4749467058769500
1414 Hz	0.0632618032655485	0.2391620260320740	0.1141465261962870	-0.6154130982293710
2000 Hz	-0.3156395548928020	-0.4838480367047310	0.0881635702300124	0.3680319468761840
2828 Hz	-0.2002537391584820	0.3237930091401330	-0.3512556093681700	-0.1789451049560760
4000 Hz	0.5906260637547510	0.1011477500393730	0.6151917977597290	0.0967997034473142
5657 Hz	0.2164550044223570	-0.1001250811505480	-0.5064788799898650	-0.0751285704732589
8000 Hz	-0.6321825584446530	0.2466285716519680	0.4245400261569100	0.0149667177802476

	Principal Components 9 to 11 and Centring Factors			
Variable	PC9	PC10	PC11	Centre
250 Hz	0.55420103171458800	-0.53403922025553200	0.45138213760472400	0.122301979166667
354 Hz	0.29709489542237500	0.23832452229678800	-0.69368619845686700	0.167580166666667
500 Hz	-0.41806514005709500	0.46788707362349100	0.49201421180052500	0.243896157407407
707 Hz	-0.49462495670947000	-0.58454252661234300	-0.25625932870834200	0.323728939393939
1000 Hz	0.38858967917302100	0.30031919715417100	0.08307638540554890	0.409447847222222
1414 Hz	-0.16142710008727700	-0.08004527231496300	-0.01637822556669920	0.511355659722222
2000 Hz	0.07748863891831260	0.02269057471360320	0.00365807858279349	0.594627951388889
2828 Hz	-0.00619028377946813	-0.00328947192439493	-0.00652190003473820	0.556790173611111
4000 Hz	-0.04042054854413880	-0.00215626597675115	0.00822327075701300	0.676837847222222
5657 Hz	0.01588102305875460	-0.00375149867861264	-0.00350228436270827	0.529095694444444
8000 Hz	0.01319267276134030	0.01587395311821680	0.00102896873706815	0.183177152777778

Centre is the centring factor. PC, principal component; PCA, principal component analysis

3.7 Appendix B: The equation for Model B

The logistic regression equation for the final model (Model B) in Chapter 3, using principal components (PC) as predictors to calculate the probability (Prob) that an ear has middle ear pathology (fail) is:

$$\text{Prob}\{\text{ear} = fail\} = \frac{1}{1 + \exp(-X\beta)}, \quad \text{where}$$

$$\begin{aligned} X\hat{\beta} = \\ &-2.231122 - 3.425094\,\text{PC1} + 2.527118\,\text{PC2} + 0.1173496\,\text{PC3} \\ &-4.588192\,\text{PC4} + 0.9285594\,\text{PC5} \end{aligned}$$

Chapter 4. Diagnosing Middle Ear Dysfunction in 10- to 16-Month-Old Infants Using Wideband Absorbance: An Ordinal Prediction Model

This chapter develops an ordinal prediction model for diagnosing middle ear dysfunction using wideband absorbance in 10- to 16-month-old infants. It has been previously published in the article: Myers, J., Kei, J., Aithal, S., Aithal, V., Driscoll, C., Khan, A., Manuel, A., Joseph, A., Malicka, A. N. (2019b). Diagnosing middle ear dysfunction in 10- to 16-month-old infants using wideband absorbance: An ordinal prediction model. *Journal of Speech Language and Hearing Research, 62*(8), 2906-2917.

I made substantive contributions to the article in the areas of study design, data collection, data analysis and drafting of the article, as outlined below:

Contributor	**Statement of contribution**
Joshua Myers (Candidate)	Study design (60%) Recruitment and data collection (60%) Data analysis (100%) Wrote the article (100%)
Joseph Kei	Study design (20%) Edited the article (40%)
Sreedevi Aithal	Study design (5%) Edited the article (15%)
Venkatesh Aithal	Study design (5%)
Carlie Driscoll	Study design (5%) Edited the article (15%)
Asaduzzaman Khan	Study design (5%) Edited the article (15%)
Alehandrea Manuel	Recruitment and data collection (20%)
Anjali Joseph	Recruitment and data collection (20%)
Alicja N. Malicka	Edited the article (15%)

4.1 Abstract

Purpose: To develop an ordinal prediction model for diagnosing middle ear dysfunction in 10- to 16-month-old infants using wideband absorbance.

Methods: Wideband absorbance, tympanometry, and distortion product otoacoustic emissions (DPOAEs) were measured in 358 ears of 186 infants aged 10 to 16 months (mean age = 12 months). An ordinal reference standard (normal, mild and severe middle ear dysfunction) was created from the tympanometry and DPOAE results. Absorbance from 1000 to 5657 Hz was used to model the probability of middle ear dysfunction with ordinal logistic regression. Model performance was evaluated using measures of discrimination and calibration. Discrimination was assessed with the *c*-index, and calibration with calibration curves. Performance measures were adjusted for overfitting (bias) using bootstrap resampling. Probabilistic and simplified methods for interpreting the model are presented. The probabilistic method displays the probability of ≥mild, and ≥severe middle ear dysfunction, and the simplified method presents the condition with the highest probability as the most likely diagnosis (normal, mild or severe middle ear dysfunction).

Results: The *c*-index of the fitted model was 0.919 (0.914 after correction for bias) and calibration was satisfactory for both the mild and severe middle ear conditions. The model performed well for the probabilistic method of interpretation, and the simplified (most likely diagnosis) method was accurate for normal and severe cases, but diagnosed some cases with mild middle ear dysfunction as normal.

Conclusions: The model may be clinically useful, and either the probabilistic or simplified paradigm of interpretation could be applied, depending on the context. In situations where the main goal is to identify severe middle ear dysfunction and ease of interpretation is highly valued, the simplified interpretation may be preferable (e.g., in a screening clinic that may not be concerned about missing some mild cases). A diagnostic clinical environment, however, may benefit from using the probabilistic method of interpretation. Future research could investigate the clinical impact of the model and the degree to which it generalizes to a new sample of infants.

4.2 Introduction

Early onset of otitis media in infancy increases risk of frequent and persistent infections through childhood (Howie, Ploussard, & Sloyer, 1975; MacIntyre et al., 2010; Marchant et al., 1986; Shurin, Pelton, Donner, & Klein, 1979). Quick and accurate tools for assessing middle ear function in infants could help provide timely diagnosis and appropriate management for affected children (Hunter et al., 2008b; Myers et al., 2018b). Wideband acoustic immittance (WAI) is an emerging test of middle ear function that is quick to administer and has several advantages over current routine clinical tests such as tympanometry. WAI can be tested at either ambient or pressurized conditions. It elicits a broadband response, assessing middle ear function over much of the frequency range important for understanding speech (Keefe et al., 1993; Keefe & Levi, 1996). WAI is an umbrella term that refers to a family of broadband middle ear tests including absorbance, reflectance, and acoustic admittance. Absorbance and reflectance have been the most used WAI measures in clinical research, because they are relatively insensitive to probe location in the ear canal (Voss et al., 2008).

Numerous diagnostic WAI studies have shown high predictive accuracy in newborns (Aithal et al., 2015; Hunter et al., 2010; Keefe et al., 2003a; Myers et al., 2018a; Sanford et al., 2009). However, much less is known about the diagnostic performance of WAI in infants beyond the neonatal period. Although preliminary studies in infants have reported strong performance (Ellison et al., 2012; Myers et al., 2018b; Prieve et al., 2013b), more diagnostic research in this age group is necessary, as WAI could be a valuable tool for identifying middle ear pathology early in childhood. Furthermore, research is needed on different age groups through infancy, as there are large developmental effects on the WAI response that are not yet complete by 24 months of age (Keefe et al., 1993; Kei et al., 2013; Myers et al., 2019c). Ellison et al. (2012) included some 12-month-old infants in their study, but the age of the 88 children in the sample ranged from 6 months to 7 years. Hence, further research targeting infants aged around 12 months is needed.

There is growing evidence that WAI can detect mild pathology such as Eustachian tube dysfunction, as well as more severe conditions such as otitis media with effusion (Aithal et al., 2018; Beers et al., 2010; Ellison et al., 2012; Hunter et al., 2008b; Myers et al., 2018a, 2018b; Robinson et al., 2016; Shaver & Sun, 2013; Voss et al., 2012; Werner et al., 2010). Studies of infants and children have found average absorbance or reflectance for mild cases falls between results for healthy ears and ears with severe middle ear dysfunction from 1000 to 6000 Hz (Beers et al., 2010; Ellison et al., 2012; Hunter et al., 2008b; Myers et al., 2018b).

Beers et al. (2010) measured reflectance in 142 children (average age = 6 years) diagnosed either by an audiological test battery (tympanometry, transient evoked otoacoustic emissions and puretone audiometry), or an otologist using pneumatic otoscopy and otomicroscopy. They found a systematic increase in ambient reflectance between 1000 and 5000 Hz from normal status, to negative middle ear pressure, to middle ear effusion. Ellison et al. (2012) measured absorbance in 88 children aged 6 months to 7 years (average age = 1 year). They reported that absorbance from 1500 to 3000 Hz decreased as stiffness of the tympanic membrane increased (assessed with pneumatic otoscopy). Hunter et al. (2008b) measured reflectance in 97 children aged 3 days to 47 months against a test battery consisting of otoscopy, tympanometry, and distortion product otoacoustic emissions (DPOAEs). They found that average reflectance from 1000 to 4000 Hz measured in ears with negative middle ear pressure fell between ears with normal and poor status. Myers et al. (2018b) measured absorbance in 249 infants aged 6- to 9-months against a reference standard consisting of 1000-Hz tympanometry and DPOAEs. They showed that infants with mild middle ear conditions (defined as failing one test) had average absorbance from 1500 to 6000 Hz that fell between normal (passed both tests) and severe conditions (failed both tests).

Previous quantitative diagnostic WAI studies in infants and children have used a binary (pass/fail) outcome to assess test performance (Aithal et al., 2015; Beers et al., 2010; Ellison et al., 2012; Hunter et al., 2010; Keefe et al., 2003a; Keefe et al., 2012; Myers et al., 2018a, 2018b; Piskorski et al., 1999; Prieve et al., 2013b; Sanford et al., 2009). However, when using a binary outcome, it can be difficult to know how to treat ears with mild dysfunction. Some studies have left these ears out of the statistical analyses of diagnostic performance (Beers et al., 2010; Ellison et al., 2012). However, leaving out mild cases can cause test performance results to be biased, as the test is only being assessed on easy-to-diagnose cases at the extreme ends of the disease spectrum (Bossuyt et al., 2003). Therefore, when using a binary outcome, mild cases should be included in either the pass or fail group. Myers et al. (2018a, 2018b) included mild cases in the fail group, with the rationale of wanting to create a reference standard sensitive to mild dysfunction. However, creating a dichotomous outcome from a disease that lies on a spectrum can lead to loss of information.

Another approach would be to create a reference standard with more than two categories. Numerous diagnostic WAI studies have used binary logistic regression models (Keefe et al., 2003a; Myers et al., 2018a, 2018b; Piskorski et al., 1999), but the outcome in logistic regression does not necessarily need to be dichotomous, and can be extended to incorporate multiple categories. The fact that previous research has shown that average absorbance systematically decreases as severity of middle ear dysfunction increases, indicates that an ordinal model may be appropriate (Harrell, 2015). Ordinal models,

where the outcome is an ordered scale (e.g., normal, mild, severe) can also be easier to interpret and have increased power compared to binary outcome models (Agresti, 2013).

The aim of this study was to develop an ordinal prediction model for diagnosing middle ear dysfunction in 10- to 16-month-old infants using wideband absorbance.

4.3 Methods

This study was part of a larger project that followed 753 infants from birth (Myers et al., 2018a, 2018b, 2019c). The project was approved by the Townsville Health Service District Institutional Ethics and the University of Queensland Behavioural and Social Science Ethical Review committees. This study presents results from 220 subjects who attended follow up appointments at approximately 12 months of age. All infants either passed the neonatal hearing screen (automated auditory brainstem response [ABR]), or had a diagnosis of normal hearing sensitivity at a follow-up diagnostic hearing evaluation. Diagnostic audiology was performed within 6 weeks of screening, with normal hearing sensitivity defined as passing click-evoked ABR, and also either tone-burst ABR (1000 and 4000 Hz), or transient evoked otoacoustic emissions.

4.3.1 Test procedure

Participants were tested in a quiet office in a paediatric community health centre. All tests were performed using an Interacoustics Titan device that was calibrated annually by the manufacturer. Infants sat on their parent's lap, and both ears were tested if possible. A suitably sized plastic probe tip was connected to the probe, and 226-Hz tympanometry, DPOAEs, and absorbance were tested, in no particular order. Each ear was also examined with otoscopy to ensure that the ear canal was not occluded with wax.

Tympanometry measured peak compensated static admittance using a 226-Hz probe tone presented at 85 dB SPL. Pressure was swept from 200 to −400 daPa at 300 daPa/s slowing to 100 daPa around the peak of the tympanogram. Traces were classified as: "type A" if there was a peak with $|Y| \geq 0.3$ mmho between −150 to 50 daPa, "type C" if a peak $|Y| \geq 0.3$ mmho occurred at pressure <−150 daPa, and otherwise "type B" (Roush, Bryant, Mundy, Zeisel, & Roberts, 1995). DPOAEs were elicited using pairs of primary tones (f_1 and f_2) for f_2 of 2000, 3000, 4000, 6000 Hz. The f_2 / f_1 ratio was 1.22, and f_1 and f_2 intensity levels were 65 and 55 dB SPL, respectively. Results were classified as "pass" if the signal-to-noise ratio was ≥6 dB, with DPOAE level ≥−10 dB SPL for at least 3 of the f_2 frequencies, otherwise "fail" (Gorga et al., 2005). Absorbance was measured at ambient pressure at 1/24 octave resolution with a broadband click delivered at 96 dB peSPL. Thirty-two clicks were presented to the

ear and averaged after removal of artefacts as described by Liu et al. (2008). A graph of absorbance was monitored during testing to check for air leaks. The test was stopped and the probe reinserted if absorbance was ≥0.3 at low frequencies (<300 Hz; Groon et al., 2015). Absorbance data were averaged into 1/2 octave frequency bands for statistical modelling to limit the potential number of predictor variables (Myers et al., 2018b).

4.3.2 Reference standard and missing data

The ordinal reference standard was created from the results of tympanometry and DPOAEs to create a more rigorous reference standard than either test in isolation (Kei & Zhao, 2012). Although DPOAEs are a cochlear response, middle ear dysfunction affects the process of forward and reverse transmission of the stimuli and emissions through the conductive pathway, and therefore reflect middle ear, as well as sensory function (Choi et al., 1999; Zhao et al., 2003). A limitation of using DPOAEs to assess middle ear function is that they may be absent due to cochlear, as well as middle ear disorders. However, sensory disorders are unlikely in the present sample, since all subjects passed the newborn hearing screen or had a finding of normal hearing at subsequent audiology assessment.

Ears were classified as: 1) "normal" if they passed DPOAEs with type A tympanograms; 2) "mild" middle ear dysfunction, likely Eustachian tube dysfunction, if they had type C tympanograms (regardless of DPOAE results), failed DPOAEs with type A tympanograms, or passed DPOAEs with type B tympanograms; or 3) "severe" middle ear dysfunction, likely middle ear effusion, if they failed DPOAEs with type B tympanograms. These classifications were chosen based on results of previous research. Beers et al. (2010) and Hunter et al. (2008b) found that ears with negative middle ear pressure had higher average reflectance than healthy ears, but not as high as ears with middle ear effusion. Myers et al. (2018a, 2018b) showed that ears that failed only one test in a test battery consisting of tympanometry and DPOAEs had absorbance that fell between ears that passed both tests and failed both tests (a fail on tympanometry being a flat trace). Therefore, the mild group in this study consisted of ears that had negative middle ear pressure on tympanometry (type C) or failed one test in the test battery (i.e., type B and passed DPOAEs, or type A and failed DPOAEs).

Table 4.1 presents the tympanometry, DPOAE, and reference standard results, and also the number of ears with missing data for each test. In total, out of the 220 infants who attended follow up assessments, there were 358 ears from 186 infants with complete reference standard and absorbance data. Data were missing due to the infant crying, or not tolerating the probe in her ear. The sample with missing absorbance and reference standard data removed for analyses is called the "study sample" in

this report. Characteristics of the study sample and reference standard results are presented in Table 4.2.

Table 4.1. Reference test results

Test	Results	Missing
Tympanometry (type A, C, B)	254, 17, 114	55
DPOAEs (pass, fail)	288, 95	57
Absorbance		63

Test results from the 440 ears of 220 infants that attended follow up, including number of missing values. DPOAEs, distortion product otoacoustic emissions.

4.3.3 Statistical modelling

The probability that an ear had mild or severe middle ear dysfunction was modelled using proportional odds ordinal logistic regression. The assumption of proportional odds is that the same regression coefficients can be used to predict the outcome regardless of the level of the reference standard being predicted. This assumption was assessed in the study sample by plotting the mean absorbance at each frequency stratified by levels of the reference standard with and without assuming proportional odds (Harrell, 2015). Models were fitted with data from both ears of an infant, if available, and Huber-White robust covariance matrix estimates were used to account for correlations between the ears. This method assumes that groups of correlated observations in the sample are independent, rather than the individual observations (Hardin, 2005), and works well when there is a large number of small clusters, as was the case in this study (Harrell, 2015).

Table 4.2. Subject characteristics

Characteristic	Value
Age (weeks)	2 missing
Median (IQR)	54 (52–57)
Range	43–70
Gender (count)	0 missing
Female (%)	85 (46)
Male (%)	101 (54)
Ethnicity (count)	0 missing
Caucasian (%)	158 (85)
Asian (%)	19 (10)
Other (%)	9 (5)
Reference standard (count)	0 missing
Pass (right, left)	232 (114, 118)
Mild (right, left)	54 (26, 28)
Severe (right, left)	72 (35, 37)

Characteristics of the 186 infants, and reference standard results for the 358 ears in the study sample. IQR, interquartile range.

Univariate (UV) and multivariate (MV) models were fitted, to compare whether there was an advantage to including multiple absorbance variables in the model. Multivariate models can be more difficult to interpret, but may be worth the extra complexity if the resulting model is more accurate, as has been found in previous research (Keefe et al., 2012; Myers et al., 2018a, 2018b; Piskorski et al., 1999). Both univariate and multivariate models used absorbance at 1/2 octave frequency resolution as predictor variables. Variables for the multivariate models were selected based on previous research in infants and children, which has found absorbance from 1000 to 6000 Hz to have an ordinal association with middle ear dysfunction (Beers et al., 2010; Ellison et al., 2012; Hunter et al., 2008b; Myers et al., 2018b; Werner et al., 2010). Three multivariate models were fitted: 1) Model MV_A only included absorbance variables as predictors, assumed to have a linear association with the outcome (i.e., simple ordinal logistic regression); 2) Model MV_B included the same variables as MV_A, with the assumption of linearity relaxed, i.e., variables were allowed to have a non-linear, flexible relationship with the ordinal outcome using restricted cubic splines (Harrell, 2015; Myers et al., 2018a); and 3) Model MV_C was the same as MV_A, with ethnicity (Caucasian and non-Caucasian), gender, and ear side included as covariates, to see if including demographic information improved model fit (Harrell, 2015). A general guide for fitting multivariate logistic regression models is to have at least 10 observations in the small-

est group for each variable. However, if the signal-to-noise ratio is high, a higher observation-to-predictor ratio can be accommodated without risking overfitting. We used the shrinkage coefficient (γ) to assess whether too many parameters were being estimated for the strength of the signal in the data (Harrell, 2015):

$$\gamma = \frac{\text{model } \chi^2 - df}{\text{model } \chi^2}, \tag{4.1}$$

where, model χ^2 is the likelihood ratio χ^2 statistic (the statistical test), and *df*, is the degrees of freedom. We took a value of γ >0.9 to be acceptable for a model, as this indicates that performance is not expected to be more than 10% worse when applied to a new sample (Harrell, 2015).

Akaike's information criterion (AIC), a relative measure of model fit, was used to compare models. A lower AIC suggests a better fitting model, if models were fitted using the same sample. The model with the lowest AIC was taken as the final model for further evaluation and interpretation (Burnham & Anderson, 2002). Performance of the final model was assessed with measures of discrimination and calibration. Discrimination assessed the ability of the model to discriminate between different levels of the reference standard, and calibration evaluated the quality of predictions (Steyerberg, 2008). It is important that a prediction model be well calibrated, since decisions are being made based on the predicted probabilities (Myers et al., 2018a). For example, approximately 30% of infants with predicted risk of 0.3 should actually have the condition (Steyerberg et al., 2010). Discrimination was assessed using the Somers' D statistic, which was converted to the *c*-index (a generalized area under the receiver operating characteristic curve) for ease of interpretation: *c*-index = 0.5(Somers' D + 1) (Harrell, 2015). Calibration was assessed with calibration curves, which plotted actual against predicted probabilities for the mild and severe middle ear pathology conditions (Myers et al., 2018a).

Performance measures of the final model were internally validated for overfitting with bootstrap resampling. Bootstrapping sampled with replacement from the study sample, a sample the same size as the study sample (a "training sample"). A model was fitted on the training sample, and the coefficients applied to the original study sample (the "test sample"). The amount of bias (or overfitting) was estimated by calculating the difference in performance measures (*c*-index and calibration) between the training and test samples. This process was repeated 500 times and averaged estimate the amount of bias in the model (Steyerberg et al., 2001c). The bias estimate was then subtracted from final model's *c*-index and calibration metrics to provide a bias-corrected estimate of model performance in new samples (Steyerberg, 2008).

Analyses were performed with R (R Core Team, 2017) expanded with the *rms* library for regression modelling (Harrell, 2016). This report has been written to conform to the recommendations of the TRIPOD (transparent reporting of a multivariate prediction model for individual prognosis or diagnosis) statement for reporting clinical prediction models (Collins et al., 2015). The data and code for the experiments are available online (https://github.com/Josh-Myers/Ordinal-Model-12-Months).

4.4 Results

Figure 4.1A shows mean absorbance as a function of frequency stratified by levels of the reference standard (normal, mild, and severe). The mean of the mild group lay between the normal and severe groups from approximately 600 to 7000 Hz. Figure 4.1B again depicts mean absorbance for the normal (type A and pass DPOAEs) and severe groups (type B and fail DPOAEs), but further stratifies the mild group by all possible combinations of reference standard results: ears with type A that failed DPOAEs, ears with type C that passed/failed DPOAEs, and ears with type B that passed DPOAEs. Mean absorbance for all subgroups fell between the normal and severe groups from 1000 to 2500, and 3500 to 4000 Hz.

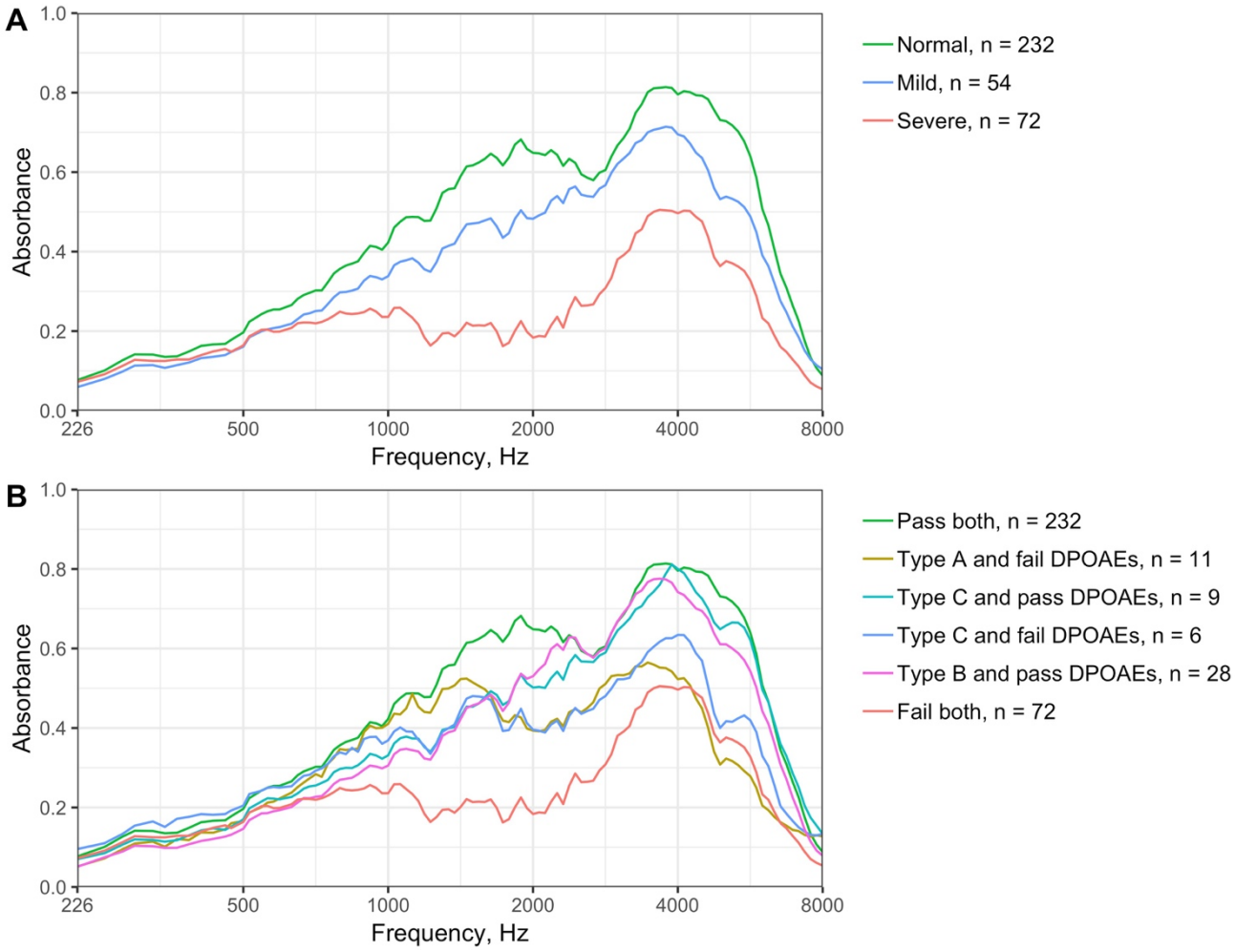


Figure 4-1. Mean absorbance for levels of the reference tests.
Mean absorbance stratified by: A) the reference standard results, and B) all possible combinations of tympanometry and distortion product otoacoustic emissions (DPOAEs) results. Data are from the 358 ears in the study sample presented at 1/24 octave frequency resolution.

Results from the univariate models are presented in Table 4.3. Absorbance at 2000 Hz had the highest LR χ^2 and lowest AIC (lower AIC indicates a better fitting model), and 1414 Hz, the highest *c*-index. Absorbance at 1000 and 4000 Hz had the third and fourth best performance, respectively, on all metrics.

Table 4.3. Statistics for the univariate models

Model	LR χ^2	*p*-value	AIC	*c*-index
$UV_{250\ Hz}$	4.40	0.036	638.12	0.578
$UV_{354\ Hz}$	5.46	0.019	637.06	0.581
$UV_{500\ Hz}$	16.97	<0.001	625.55	0.624
$UV_{707\ Hz}$	67.65	<0.001	574.87	0.734
$UV_{1000\ Hz}$	149.76	<0.001	492.77	0.830
$UV_{1414\ Hz}$	202.14	<0.001	440.39	0.861
$UV_{2000\ Hz}$	209.45	<0.001	433.08	0.855
$UV_{2828\ Hz}$	86.94	<0.001	555.59	0.719
$UV_{4000\ Hz}$	103.92	<0.001	538.61	0.787
$UV_{5657\ Hz}$	63.16	<0.001	579.36	0.725
$UV_{8000\ Hz}$	10.59	0.001	631.94	0.611

LR χ^2 statistics with associated *p*-values, AIC, and *c*-index results from the univariate (UV) models. AIC, Akaike's information criterion; LR χ^2, likelihood ratio chi-squared statistic.

Plots assessing the assumption of proportional odds for the absorbance predictors in the multivariate models (1000 to 5657 Hz) are shown in Figure 4.2. The solid lines and circles represent the simple stratified means, and the dashed lines the expected values if the assumption of proportional odds is met. The trend in the solid lines should be monotonic to satisfy the assumption of ordinality. The stratified means were monotonic for all variables, and the expected values were very close to the simple means for all variables except for 2828 Hz. Overall, the assumption of proportional odds was satisfied for these variables.

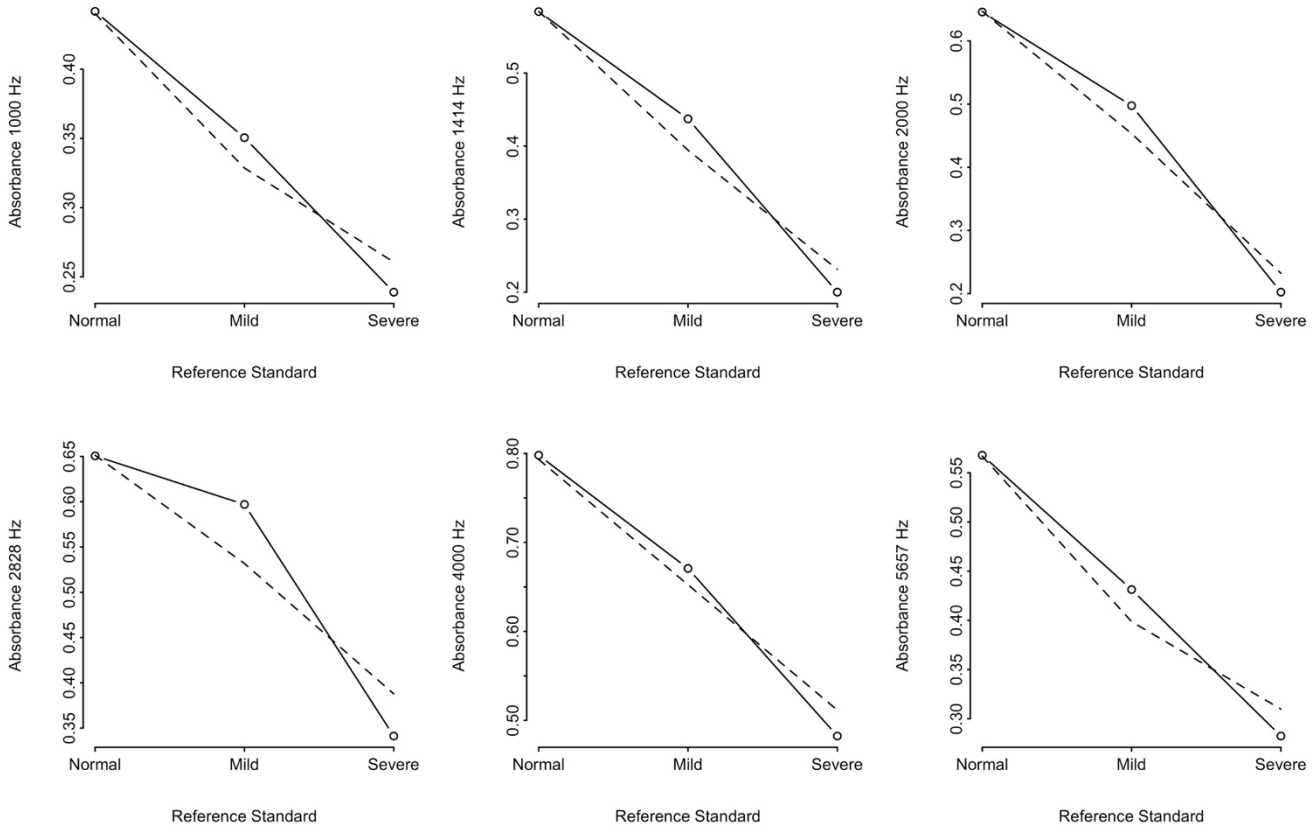


Figure 4-2. Testing the assumption of proportional odds.
Plots assessing the assumption of proportional odds for the absorbance predictors used in the multivariate models. The circles connected by solid lines represent the simple stratified means (i.e., the raw data), and the dashed lines the expected values if the assumption of proportional odds is met. The trends in the simple stratified means should be monotonic to satisfy the assumption of ordinality. The trends for the stratified means are monotonic for all variables, and the expected values are very close to the simple means for all variables except for absorbance at 2828 Hz.

Results from the multivariate models are presented in Table 4.4. The γ was acceptable for all models (>0.9) and Model MV_A (the simple linear model) was better fitting (with lower AIC) than both Model MV_B (the nonlinear model), and Model MV_C (the model including covariates). Model MV_A was also better fitting than the best fitting univariate model ($UV_{2000\ Hz}$). Model MV_A was, therefore, taken as the final model for further evaluation of discrimination and calibration.

Table 4.4. Statistics for the multivariate models

Model	Predictor Variables	LR χ^2	*df*	γ	AIC
MV_A	Absorbance from 1000 to 5657 Hz	298.92	6	0.98	353.61
MV_B	Model MV_A variables nonlinear	324.50	24	0.93	364.02
MV_C	Model MV_A variables + covariates	302.03	9	0.97	356.49

LR χ^2, AIC, and *c*-index results from the multivariate (MV) models. The *p*-values for the LR χ^2 statistics were <0.0001 for all models. The absorbance predictor variables were 1000, 1414, 2000, 2828, 4000, 5657 Hz at 1/2 octave frequency resolution. Covariates in Model MV_C were ear side, gender and ethnicity. AIC, Akaike's information criterion; LR χ^2, likelihood ratio chi-squared statistic; γ, shrinkage coefficient; *df*, degrees of freedom.

The importance of predictors in Model MV_A is investigated in Table 4.5. Absorbance at 2000 Hz contributed the most to the model (highest LR χ^2), followed by 1000, and then 5657 Hz. Absorbance at 1414, 2828, and 4000 Hz were all relatively strong univariate predictors (*c*-index >0.7), but did not substantially contribute to Model MV_A, possibly due to redundancy effects. For example, absorbance at 1000 and 2000 Hz were important predictors in the model, but not the inter-octave variable, 1414 Hz, even though this variable had the highest *c*-index of the univariate models. This may be because 1414 Hz did not contribute a lot of new information not already contained in the adjacent variables. However, we retained the seemingly unimportant variables (1414, 2828 and 4000 Hz), since they may still be contributing, and removing variables based on statistical testing is not advised (Gelman & Hill, 2007).

Table 4.5. Statistics for Model MV_A

Variable	LR χ^2	*df*	*p*-value
1000 Hz	11.69	1	<0.001
1414 Hz	1.33	1	0.248
2000 Hz	13.04	1	<0.001
2828 Hz	<0.01	1	0.961
4000 Hz	2.68	1	0.102
5657 Hz	9.57	1	0.002
TOTAL	72.57	6	<0.001

Statistical analysis for the final model (MV_A), showing the LR χ^2 statistics, associated *p*-values, and degrees of freedom (*df*) for each predictor variable in the model. LR χ^2, likelihood ratio chi-squared statistic; MV, multivariate.

Apparent and bias-corrected (bootstrapped) *c*-index values for Model MV_A are presented in Table 4.6. Apparent performance was 0.919, with estimated bias of 0.003, leaving a bias-corrected *c*-index of 0.914. This was higher than the best performing univariate model ($UV_{1414\ Hz}$, *c*-index = 0.861). Appar-

ent and bias-corrected calibration curves for Model MV_A, for predicting mild and severe middle ear pathology are presented in Figure 4.3. Predictions for mild middle ear dysfunction (A) were slightly low between 0.2 and 0.4, but were overall satisfactory. Predictions for severe dysfunction (B) had a nonlinearity, with predictions too low from 0.1 to 0.3, but the curve was satisfactory for predictions above 0.3. The nonlinearity is not a cause for concern, however, since probabilities falling in this region are unlikely to warrant further action, i.e., <0.3 probability of middle ear dysfunction would not be of enough concern to require review or referral (Myers et al., 2018b).

Table 4.6. Discriminative ability of Model MV_A

Sample	*c*-index
Apparent	0.919
Bootstrap training	0.921
Bootstrap test	0.917
Bias (bootstrap training – test)	0.003
Bias-corrected	0.914

The *c*-index results for the apparent, bootstrapped and bias-corrected samples for the final model (MV_A). MV, multivariate.

The equation for Model MV_A to make predictions is provided in Section 4.6: Appendix. A web application implementing the model is available online (https://joshmyers.shinyapps.io/WAIPredictions/) that can make predictions using a file exported from a Titan Device, or by manually entering absorbance values at 1/2 octave frequency resolution.

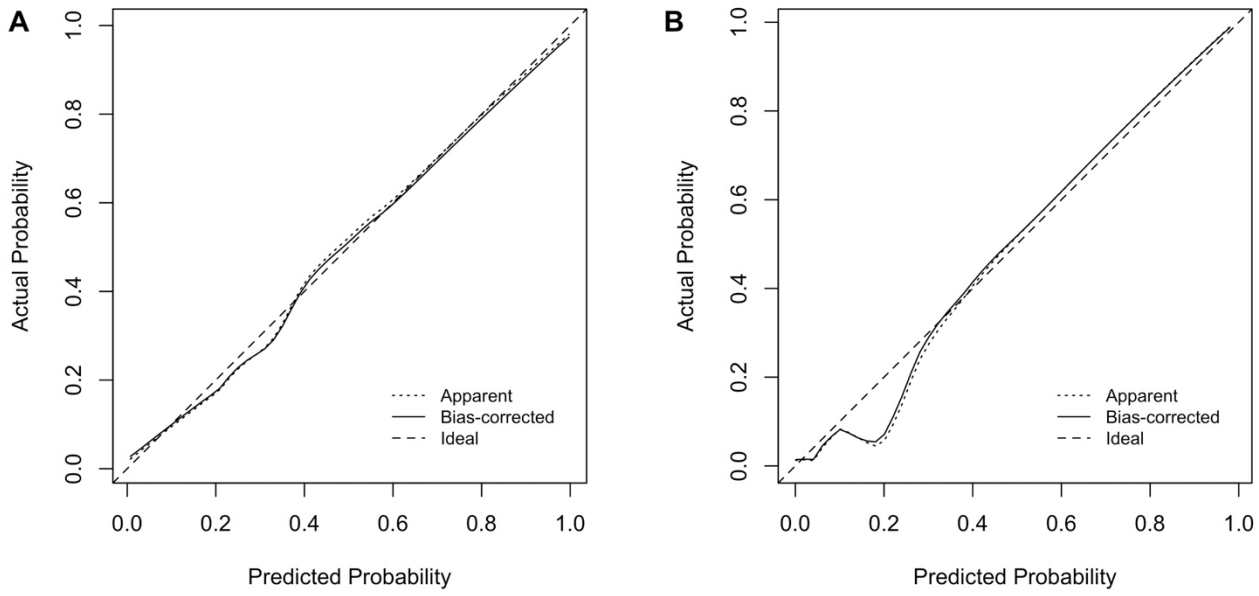


Figure 4-3. Calibration curves for Model MV_A.
Actual against predicted probability for the apparent and bias-corrected samples for A) the probability of ≥mild middle ear dysfunction; and B) the probability of ≥severe middle ear dysfunction. Ideal calibration is depicted by the diagonal dashed lines.

4.5 Discussion

4.5.1 Clinical application of the model

There are 2 main ways that results from model MV_A could be presented for clinical use. One option would be to present the probability (P) of ≥ mild and/or ≥ severe dysfunction: P(ME ≥ j|X), where ME is the condition of the middle ear, j the level of the reference standard (mild or severe middle ear dysfunction), and X the absorbance predictor variables in the model (1000 to 5657 Hz). Probabilities for one or both levels of the reference standard could be presented. For example, if only wanting to test for severe middle ear dysfunction, you could just present P(ME ≥ severe). The decision threshold would be the point where there is enough concern to warrant further action (Myers et al., 2018a). This could be set automatically in a screening program (e.g., P > 0.5), or on a case-by-case basis in a diagnostic context (Myers et al., 2018b).

An alternative, simplified, approach would be to calculate the probability of normal, mild and severe dysfunction: P(ME = j|X). P(ME = severe) is P(ME ≥ severe); P(ME = mild) is P(ME ≥ mild) – P(ME ≥ severe); and P(ME = normal) is 1 – P(ME ≥ mild) (since probabilities sum to 1). Probabilities for all three conditions could be presented, or the condition with the highest probability could be displayed as the most likely condition predicted by the model. Two examples are provided in Figure

4.4. Each panel shows the absorbance results for one ear of an infant, as well as the probabilistic P(ME ≥ j|X) and simplified P(ME = j|X) model interpretation results.

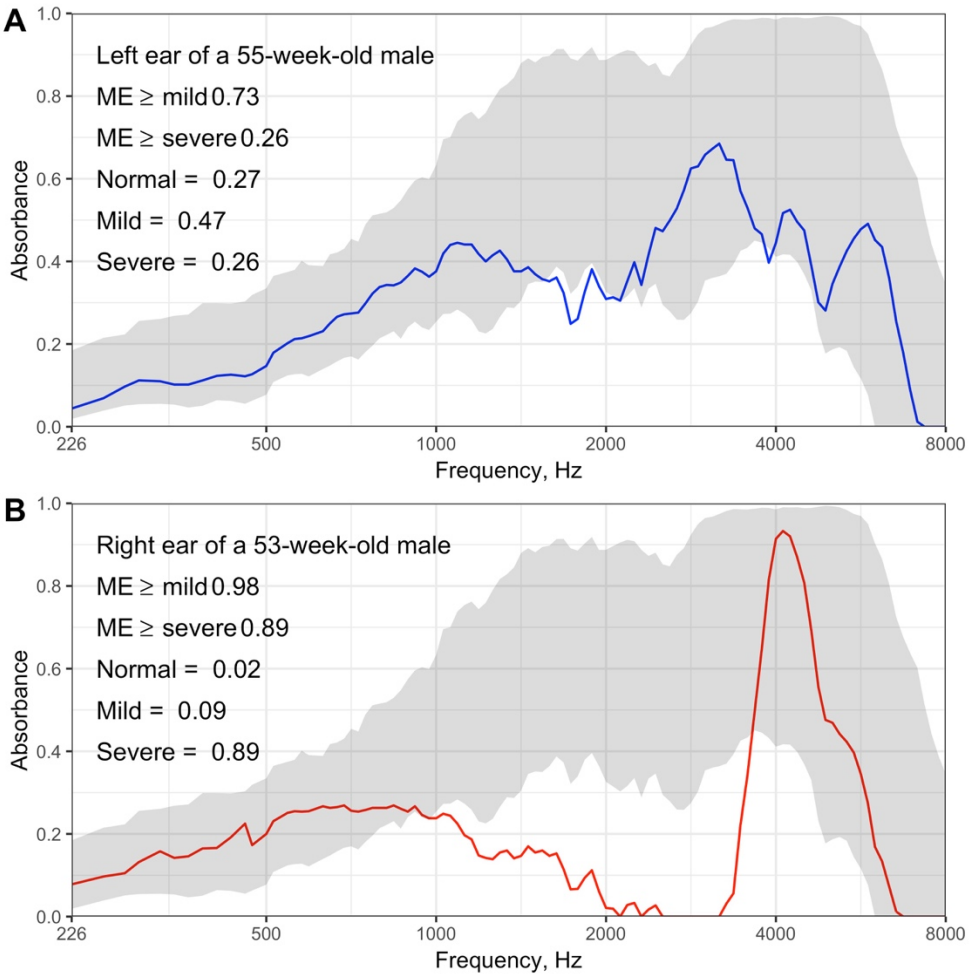


Figure 4-4. Examples of applying the model to individual cases.
Each panel shows the absorbance results for one ear of an infant, as well as the probabilistic P(ME ≥ *j*|X) and simplified P(ME = *j*|X) model interpretations. *j*, reference standard level (e.g., mild or severe); ME, middle ear condition; P, probability; X, absorbance predictors in model MV_A (1000 to 5657 Hz).

Table 4.7 shows the most likely P(ME = j|X) diagnosis (simplified interpretation) predicted by the model compared to the reference standard labels for the study sample. Note that the range of probabilities (in the "Predictions" column) used to classify ears never included the area of nonlinearity in calibration (0.1 to 0.3 for severe dysfunction). Looking down the columns of the table, this method of interpretation did very in well identifying normal ears (only 10 of 232 misdiagnosed). Severe cases were also correctly identified in most instances (10 of 72 misdiagnosed). However, over half of the mild cases were mislabelled (only 20 of 54 correctly diagnosed). Looking across the rows, of the 255 predicted to be normal, 33 were incorrectly labelled (28 as mild, and 5 as severe); of the 31 predicted to have mild dysfunction, 11 were misdiagnosed (6 as normal and 5 as severe); and of the 72 predicted to

have severe dysfunction, 10 were mislabelled (4 as normal and 6 as mild). In general, the most likely diagnosis predicted by the simplified interpretation was correct. Ears predicted to be normal were normal 87% of the time, mild predictions were 65% correct, and severe predictions 86% correct. However, this method of interpretation was insensitive to mild dysfunction, because it resulted in some mild cases being misdiagnosed as normal.

Table 4.7. Diagnostic accuracy of the simplified model interpretation

	Reference Standard		
Predictions (Range)	Normal	Mild	Severe
Normal (0.44–1.00)	222	28	5
Mild (0.43–0.47)	6	20	5
Severe (0.43–1.00)	4	6	62

The most likely diagnosis (simplified method of interpretation) predicted by the model (rows) compared to the corresponding reference standard labels (columns). The range of predictions used for a particular diagnosis is provided in parentheses in the left-hand column. For example, ears that were given a diagnosis of "mild" by the most likely diagnosis method (i.e., mild had a higher probability than normal or severe) had predictions ranging from 0.43 to 0.47.

In diagnostic clinical contexts, where it is important to identify cases of both mild and severe dysfunction, it may be beneficial to use the more complex probabilistic $P(ME \geq j|X)$ paradigm for interpreting results, choosing the cut off for ≥mild and ≥severe based on the level of risk where further action is warranted (Myers et al., 2018b). However, if the main goal is to identify severe cases of middle ear dysfunction, and ease of interpretation is preferable, the simplified method may be suitable (e.g., a screening program may wish for a simple interpretation and not be concerned about missing some mild cases).

4.5.2 Absorbance results and model development

Average absorbance for ears with mild dysfunction fell between normal and severe ears from 600 to 7000 Hz, which is largely consistent with results of previous research (Beers et al., 2010; Ellison et al., 2012; Hunter et al., 2008a; Myers et al., 2018b). This was true for all possible combinations of reference standard results, including ears with type A tympanograms but absent DPOAEs (Figure 4.1), which indicates that DPOAEs in these ears were most likely absent due to middle ear, not sensory dysfunction.

Model MV_A had powerful discriminative ability (bias-corrected *c*-index = 0.914), which although not as high as the 0.93 reported for absorbance by Ellison et al. (2012), was comparable, and within the margin of error reported by Ellison et al. Calibration of Model MV_A was overall satisfactory. There

was a nonlinearity in the low predictions (0.1 to 0.3) in the calibration curve for predicting severe dysfunction, but calibration was acceptable over the regions relevant for clinical decision making (Myers et al., 2018b). Bias-corrected (bootstrapped) results were very close to apparent performance measures for both discrimination and calibration, indicating that the model was not significantly overfitting the data and may generalize well to new infants.

Consistent with previous studies, we found that the multivariate model outperformed the best univariate predictor (Keefe et al., 2012; Myers et al., 2018a, 2018b; Piskorski et al., 1999). Model MV_A had higher *c*-index and lower AIC than the top two univariate models, $UV_{1414\ Hz}$ and $UV_{2000\ Hz}$. Also consistent with other reports, we found that adding covariates such as ethnicity and gender did not improve performance (Beers et al., 2010; Myers et al., 2018a, 2018b; Shahnaz et al., 2013). Contrary to Myers et al. (2018a), however, allowing predictors to have a nonlinear association with the outcome did not improve model fit in this study. Age may be a contributing factor to this difference, as Myers et al. (2018a) developed a model for neonates, compared to 12-month-old infants in the present study, and developmental changes in the outer and middle ear over the first year of life have a substantial effect on the WAI response (Kei et al., 2013). Furthermore, Myers et al. (2018a) used a binary pass/fail outcome, compared to the ordinal outcome used in this study.

4.5.3 Strengths, limitations, and directions for future research

Developing an ordinal model better captured the spectrum of middle ear disease compared to a binary outcome, and it eliminated the problem of how to treat mild cases, as it can be difficult to know whether these should be classified as normal or diseased if using a dichotomous outcome.

The modelling strategy was intended to reduce risk of overfitting by limiting the number of predictors and bootstrap resampling correcting for bias indicated that the model may generalize well to new samples. However, the model would need to be applied to a new sample of infants to assess the extent to which differences in environment, subject characteristics, and equipment affect model performance (Myers et al., 2018b).

A limitation of this study was that it was not blinded, as the same researcher collected both the reference standard and absorbance data. Even though interpretation of the reference tests was objective, this may have introduced bias. Also, the reference standard used was not the gold standard for identifying middle ear disease in infants. The mild group may not have been homogenous, ears in this group may have had Eustachian tube dysfunction, partial middle ear effusion, or both (Shaver & Sun, 2013). Future research could create a more rigorous ordinal reference standard using examination by an otologist with surgical confirmation for middle ear effusion.

4.5.4 Summary and conclusions

We developed an ordinal prediction model for identifying middle ear dysfunction using wideband absorbance. The model may be clinically useful, and either the probabilistic or simplified paradigm of model interpretation could be applied, depending on the context. In situations where the main goal is to identify severe cases of middle ear dysfunction and ease of interpretation is preferable, the simplified method may be suitable (e.g., a screening program that may not be concerned about missing some mild cases). However, diagnostic clinical environments may benefit from using the more complex probabilistic method of interpretation. Future research could investigate the impact of the model and the degree to which it generalizes to a new sample of infants.

4.6 Appendix: The equation for Model MV_A

The ordinal logistic regression equation for the final model (MV_A) in Chapter 4, using absorbance (*A*) at 1000, 1414, 2000, 2828, 4000, and 5657 Hz to calculate the probability (P) of middle ear (ME) dysfunction for the levels of the reference standard (*j*) is:

$$\mathrm{P}\{\mathrm{ME} \geq j\} = \frac{1}{1 + \exp(-\alpha_j - X\beta)}, \quad \text{where}$$

$$\begin{aligned} \hat{\alpha}_{\mathrm{mild}} &= 8.356191 \\ \hat{\alpha}_{\mathrm{severe}} &= 6.291253 \end{aligned}$$

$$\begin{aligned} X\hat{\beta} = & \\ & -8.087252\, A_{1000\ \mathrm{Hz}} - 1.749161\, A_{1414\ \mathrm{Hz}} - 5.054438\, A_{2000\ \mathrm{Hz}} \\ & +0.05685856\, A_{2828\ \mathrm{Hz}} - 1.646836\, A_{4000\ \mathrm{Hz}} - 2.703953\, A_{5657\ \mathrm{Hz}} \end{aligned}$$

Chapter 5. Longitudinal Development of Wideband Absorbance and Admittance Through Infancy

This chapter investigates developmental effects on wideband acoustic immittance measures through infancy. It has been previously published in the article: Myers, J., Kei, J., Aithal, S., Aithal, V., Driscoll, C., Khan, A., Manuel, A., Joseph, A., Malicka, A. N. (2019c). Longitudinal development of wideband absorbance and admittance through infancy. *Journal of Speech Language and Hearing Research, 62*(7), 2535-2552.

I made substantive contributions to the article in the areas of study design, data collection, data analysis and drafting of the article, as outlined below:

Contributor	**Statement of contribution**
Joshua Myers (Candidate)	Study design (60%) Recruitment and data collection (60%) Data analysis (100%) Wrote the article (100%)
Joseph Kei	Study design (20%) Edited the article (40%)
Sreedevi Aithal	Study design (5%) Edited the article (15%)
Venkatesh Aithal	Study design (5%)
Carlie Driscoll	Study design (5%) Edited the article (15%)
Asaduzzaman Khan	Study design (5%) Edited the article (15%)
Alehandrea Manuel	Recruitment and data collection (20%)
Anjali Joseph	Recruitment and data collection (20%)
Alicja N. Malicka	Edited the article (15%)

5.1 Abstract

Purpose: To study the normal longitudinal development of wideband absorbance and admittance measures through infancy.

Methods: Two-hundred and one infants who passed the newborn hearing screen (automated auditory brainstem response) were tested at birth, then followed up at approximately 6, 12, and 18 months of age. Most infants were of either Caucasian (86%) or Asian (11%) descent. At each test session, infants passed tympanometry and distortion product otoacoustic emissions tests. High-frequency (1000-Hz) tympanometry was used at birth and 6 months, and low-frequency (226-Hz) tympanometry at 12 and 18 months of age. Wideband pressure reflectance was also measured at each session, and analysed in terms of absorbance, admittance at the probe tip, and admittance normalized for differences in ear canal area. Multilevel hierarchical models were fitted to the absorbance and admittance data to investigate for effects of age, ear side, gender, ethnicity, and frequency.

Results: There were considerable age effects on wideband absorbance and admittance measurements over the first 18 months of life. The most dramatic changes occurred between birth and 6 months and there were significant differences between all age groups in the 3000 to 4000 Hz region. There were significant ethnicity effects that were substantial for certain combinations of ethnicity, age, and frequency (e.g., absorbance at 6000 Hz at 12 months).

Conclusions: There are large developmental effects on wideband absorbance and admittance measures through infancy. For absorbance, we recommend separate reference data be used at birth, 6 months, and 12–18 months. For admittance (both normalized, and at the probe tip), we advise using separate normative regions for each age group (neonates, 6, 12, and 18 months).

5.2 Introduction

The outer and middle ear are not fully mature at birth, and continue to develop throughout infancy. The ear canal increases in length and diameter (Keefe et al., 1993), and stiffens as the bony part of the canal wall lengthens (Wright, 1997). The tympanic membrane thins and increases in inclination, and the middle ear cavities increase in size and pneumatization (Hunter & Shahnaz, 2014). The mass and resistance of the conductive pathway (outer and middle ear) decrease, and stiffness increases, due to loss of mesenchyme, ossification of the ear canal, tightening of the ossicular joints and tympanic ring, and change in composition and orientation of the tympanic membrane (Wilson, 2012).

These developmental changes affect the results of aural acoustic tests for assessing middle ear function. Traditional low-frequency (226 Hz) tympanometry (LFT) is not accurate for diagnosing middle ear pathology in young infants, due to energy absorption by the more compliant ear canal wall of these infants at low frequencies (Hunter et al., 2008a; Mazlan & Kei, 2012). High-frequency tympanometry (HFT) using a 1000-Hz probe tone is more effective in infants under 7 months of age (Baldwin, 2006; Zhiqi et al., 2010) but is insensitive to conductive dysfunction in neonates (Margolis et al., 2003; Sanford et al., 2009; Swanepoel et al., 2007). Wideband acoustic immittance (WAI) is an emerging technology for middle ear assessment with several advantages over traditional tests such as tympanometry. WAI is more sensitive to conductive dysfunction in neonates compared to HFT (Hunter et al., 2010; Sanford et al., 2009), can assess middle ear function over a wide range of frequencies (e.g., 226 to 8000 Hz), and does not require pressurization of the ear canal, which can cause significant distention of the compliant ear canal wall of young infants (Aithal et al., 2015; Prieve et al., 2013b; Vander Werff et al., 2007). These qualities make WAI a promising tool for assessing middle ear function in infants (Aithal et al., 2015; Ellison et al., 2012; Hunter et al., 2010; Keefe et al., 2003a; Myers et al., 2018a, 2018b; Prieve et al., 2013b; Sanford et al., 2009).

The term WAI refers to a family of tests derived from the complex pressure reflectance (*PR*), including energy reflectance (*R*), energy absorbance (*A*), acoustic admittance (*Y*), and acoustic impedance (*Z*). PR is the ratio of the reflected (incoming) to forward-propagating (outgoing) acoustic pressure wave amplitude (Keefe & Levi, 1996). $R = |PR|^2$, and measures the proportion of energy reflected back from middle ear. $A = 1 - R$, and represents the proportion of energy absorbed by the conductive pathway. *Y* can be calculated from *PR*:

$$Y = Z_c^{-1} \frac{1 - PR}{1 + PR}. \tag{5.1}$$

Z_c, the characteristic impedance, is $\rho c/S_{tube}$, where ρ is the density of air, c is the speed of sound, and S_{tube} is the cross-sectional area of the calibration tube used to estimate ear canal area when calculating *PR* (Keefe et al., 2015). *Y* is a complex measure that can be expressed in polar form with magnitude ($|Y|$) and phase (φ_Y), or rectangular form in terms of its real and imaginary parts, conductance (*G*), and susceptance (*B*), respectively ($Y = G + iB$). *Z* is the reciprocal of *Y*, and is also complex, with a real part, resistance, and an imaginary part, reactance.

It is generally accepted that maturational factors during infancy significantly affect WAI measurements (Kei et al., 2013). Thus, there is a need to study the effect of age on the WAI response so that changes due to normal development can be distinguished from changes caused by dysfunction of the conductive pathway (Sanford & Feeney, 2008). Furthermore, such studies can give insight into the maturational developmental of the outer and middle ear, as WAI measures outer and middle ear function over a wide range of frequencies (Aithal et al., 2014b). A number of studies have investigated the effect of age on ambient *A* or *R* through infancy. Significant age effects have been found over the first few days of life (Hunter et al., 2010; Keefe et al., 2000; Myers et al., 2018a; Sanford et al., 2009), the first 6 months of life (Aithal et al., 2014b; Shahnaz et al., 2014), from 2 to 9 months (Werner et al., 2010), the first year of life (Hunter et al., 2016), and from 1 to 24 months of age (Keefe et al., 1993). There is general agreement that the most rapid changes occur in the first 3 to 6 months of infancy (Aithal et al., 2014b; Hunter et al., 2016; Shahnaz et al., 2014), but development is not yet complete by 24 months of age (Keefe et al., 1993). While longitudinal studies have shown differences with age, some cross-sectional studies have not. Hunter et al. (2008b) found no significant differences except for *R* at 6000 Hz in a sample of 97 subjects aged 3 days to 47 months, and Merchant et al. (2010) reported no significant differences in R between neonates and 1-month-old infants.

Much less is known, however, about the effect of age on other WAI measures such as *Y* and *Z*. Knowledge of the effect of maturation is important for these measures also, as there is interest in using *Y* and *Z*, as well as *A* and *R*, for the diagnosis of middle ear pathology in infants (Aithal et al., 2017; Ellison et al., 2012; Keefe et al., 2003a; Myers et al., 2018a; Sanford et al., 2009; Voss, Herrmann, Horton, Amadei, & Kujawa, 2016). Keefe et al. (1993) studied changes in *Z* in a sample of infants aged 1 to 24 months (n = 15 at 1 month, 18 at 3 months; 11 at 6 months; 23 at 12 months; and 11 at 24 months). They found that resistance decreased with age through infancy. Growth in ear canal area accounted for much of the variability between age groups, and stiffening of the ear canal walls and growth of the tympanic cavity were also significant factors. Werner et al. (2010) measured *Z* in 458 infants aged 2 to 9 months, and found a significant age effect on resistance and reactance in agreement with the results of Keefe et al. Sanford and Feeney (2008) studied developmental effects on *Y* measured

under pressurized conditions in a cross-sectional study of 60 infants aged 4 to 27 weeks, and found a trend of increasing $|Y|$ with age and also increasing φ_Y below 1000 Hz, indicating increased stiffness with age.

However, there are limitations to these studies, and further research is needed. Apart from a case history, Keefe et al. (1993) did not use a reference standard to assess middle ear function in their participants, and Werner et al. used LFT as the reference test in young infants (<6 months old). Sanford and Feeney studied the effect of maturation on pressurized Y using an appropriate reference standard (distortion product otoacoustic emissions [DPOAEs] and HFT), up to 6 months of age, but there remains need to study maturational effects on ambient Y measures against an appropriate reference test through infancy. The aim of this study was to investigate the normal longitudinal development of ambient A and Y from birth to 18 months.

5.3 Methods

This study was part of a larger project recruiting infants at birth and following them up through infancy (Myers et al., 2018a, 2018b). A total of 753 infants have been recruited to the project. All healthy babies born in the maternity ward were eligible to be recruited to the study, which excluded high-risk neonates in the Special Care Nursery and Neonatal Intensive Care Unit. The present study presents results from 201 infants that were assessed to have normal middle ear function at birth, and attended at least one follow up session with a finding of normal middle ear function in at least one ear. Characteristics of the study sample are presented in Table 5.1, and the number of infants, ears, and age range for each group are provided in Table 5.2. There is a high proportion of neonates born via C-section (43%) because infants were exclusively recruited from the Maternity Ward. There is also a Birth Centre at the Townsville Hospital where women return home the same day after giving birth, but we did not recruit from there due to time restraints. Neonates usually stay in the Maternity Ward because the mother needs to stay overnight in hospital, often due to having a C-section.

Ethical approval for the study was obtained from the Townsville Health Service District Institutional Ethics Committee, and the University of Queensland Behavioural and Social Science Ethical Review Committee.

Table 5.1. Characteristics of infants in the study

Characteristic	Value
Gender (count)	0 missing
Female (%)	97 (48.3)
Male (%)	104 (51.7)
Ethnicity (count)	1 missing
Caucasian (%)	173 (86.0)
Asian (%)	22 (10.9)
Oceanian (%)	3 (1.4)
South American (%)	1 (0.5)
African (%)	1 (0.5)
Gestational age (weeks)	1 missing
Median (IQR)	39 (38.2–40.2)
Range	34–41.6
Birth type (count)	0 missing
Vaginal (%)	114 (56.7)
C-section (%)	87 (43.3)
Birth weight (grams)	0 missing
Median (IQR)	3460 (3180–3810)
Range	2390–4550
Head circumference (cm)	1 missing
Median (IQR)	35 (33.5–36)
Range	24.8–38.2
Birth length (cm)	0 missing
Median (IQR)	50 (49–52)
Range	37–58.5

The number of infants with missing data for each characteristic is provided in the Value column. IQR, interquartile range.

All ears included in this study passed a battery of tests. Neonates passed automated auditory brain-stem response (AABR), HFT and DPOAEs. Infants in the 6-month group passed HFT and DPOAEs, and infants in the 12- and 18-month groups passed LFT and DPOAEs. Otoscopy was also performed on the 6- to 18-month-old infants to ensure that the ear canal was not occluded by wax.

Table 5.2. Sample size and age of infants in each age group

	Number	Ears (right, left)	Median age (IQR, range, units)
Neonate	201	328 (169, 159)	46 (34–56, 12–163, hours)
6 months	160	265 (138, 127)	27 (26–29, 23–38, weeks)
12 months	112	196 (95, 101)	54 (52–58, 46–64, weeks)
18 months	81	139 (70, 69)	80 (78–81, 76–97, weeks)

Number of infants, number of ears, and age of infants in each age group. IQR, interquartile range.

5.3.1 Test procedure

All tests were performed by a research audiologist, except for AABR, which was done by a nurse as part of the newborn hearing screening program. Neonates were tested in the maternity unit at Townsville Hospital, and the other age groups were tested in a quiet office in a paediatric community health center. AABR was performed with a Natus ALGO 3 Newborn Hearing Screener which presented clicks at 35 dB nHL and used a template matching algorithm to pass or fail ears. AABR was always tested first on the neonates, so as not to interfere with the hearing screening program. All other tests were done using Interacoustics Titan devices which were annually calibrated by the manufacturer and checked daily with a 2-cm^3 cavity. Both ears of participants were tested if possible. The most accessible ear was tested first and tympanometry (HFT or LFT depending on the age group), DPOAEs and WAI were performed in no particular order.

Tympanometric measurements plotted peak compensated $|Y|$ as a function of pressure for a 1000-Hz (HFT) or 226-Hz (LFT) probe tone, delivered at 85 dB SPL. Pressure was swept from 200 to −400 daPa at 300 daPa/s slowing to 100 daPa at the peak of the tympanogram. The pass criterion for HFT was a peaked trace extending above a baseline drawn between the positive and negative extremes of the tympanogram (Baldwin, 2006; Kei et al., 2003). The pass criteria for LFT was a peaked trace between −150 and 50 daPa with $|Y| \geq 0.2$ mmho for the 12-month, and ≥0.3 mmho for the 18-month age groups (Roush et al., 1995). DPOAEs were recorded for primary tones f_1 and f_2 for f_2 of 2000, 3000, 4000, 6000 Hz, with an f_2 / f_1 ratio of 1.22, and intensity levels of 65 and 55 dB SPL for f_1 and f_2, respectively. The pass criteria were signal-to-noise ratio of ≥6 dB, and DPOAE level ≥−10 dB for at least 3 of the f_2 frequencies (Gorga et al., 2005).

Ambient *PR* was measured at 1/24 octave frequency resolution in response to 32 broadband clicks (226 to 8000 Hz) presented at 96 dB peSPL. Results from the clicks were averaged after noisy responses were removed as described by Liu et al. (2008). To check for air leaks, a graph of A was monitored during testing. The test was stopped and the probe reinserted if A was high (≥0.7 for neonates and ≥0.3

for other age groups) at low frequencies (<300 Hz; Groon et al., 2015). A, $|Y|$ and φ_Y data were obtained with the Interacoustics Titan Research Module, which saved results as a text file after each test. Y data were also converted into G and B, since it is useful to depict Y in rectangular, as well as polar form. Because $|Y|$, G and B are affected by the volume of air between the probe tip and the eardrum, these variables were also normalized to reduce variability caused by differences in ear canal cross-sectional area (Allen et al., 2005). We investigated developmental effects on both normalized Y and Y at the probe tip in this study, because commercially available WAI systems for clinical use provide results in one of these formats. In this report, normalized variables are denoted with a subscript "n" (e.g., $|Y|_n$), and variables measured at the probe tip, with a subscript "t" (e.g., G_t). Y_t variables were normalized by multiplying the value measured at the tip by Z_{ce}, which is the characteristic impedance of the ear canal for an individual ear (e.g., $|Y|_n = |Y|_t \times Z_{ce}$). Z_{ce} is $\rho c/S$, where S is the acoustically estimated ear canal cross-sectional area, which is ρc/<resistance>, with <resistance> being the acoustic resistance averaged across frequency (Keefe et al., 1993). The values for ρ and c for the calculations were taken from Benade (1968) for a temperature of 22 °C.

5.3.2 Statistical analyses

Multilevel hierarchical models were used for statistical analyses, due to the nested and longitudinal nature of the data, as multiple measurements were made at frequencies in ears that were nested within infants. WAI data were averaged into 1/2 octave bandwidths for modelling, the frequency bandwidths are presented in Table 5.3. The response was the WAI measure (e.g., A, or $|Y|_t$), and the predictors were age group (neonates, 6 months, 12 months, 18 months), ethnicity (Caucasian or non-Caucasian), gender, ear side, and frequency (modelled as a factor). An interaction was included between frequency and age group because we expected the effect of age to vary with frequency (Hunter et al., 2016; Keefe et al., 1993). An interaction was also included between age and ethnicity because Hunter et al. (2016) found this interaction to be significant. Additionally, Ethnicity × Frequency, and Age × Ethnicity × Frequency interactions were included, because ethnicity, and Age × Ethnicity relationships may be moderated by frequency (Kenny, 2011; Shahnaz et al., 2013).

Table 5.3. Frequency bandwidths used for half-octave averaging

1/2 octave f (Hz)	min (Hz)	max (Hz)
250	226.00	297.30
354	297.31	420.45
500	420.46	594.60
707	594.61	840.90
1000	840.91	1189.21
1414	1189.22	1681.79
2000	1681.80	2378.41
2828	2378.42	3363.59
4000	3363.60	4756.83
5657	4756.84	6727.17
8000	6727.18	8000.00

Minimum (min) and maximum (max) frequencies (f) used to average WAI data into 1/2 octave bandwidths

We initially planned to use linear mixed models for all WAI measures, but preliminary models showed violations of the assumption of normally distributed residuals for all models. For A, the issue was that since it is a proportion, and (theoretically) bounded by 0 and 1, it was not normally distributed. Therefore, we used a beta generalized linear mixed model, which is appropriate for modelling a proportion. A limitation of the beta distribution is that it cannot include values in the response ≤0. Therefore, for this model, values of A that were ≤0 were set to 0.0001. Since $A < 0$ is theoretically impossible, these results were likely due to calibration error, and adding a small fraction to 0 values to make them positive was thought to be acceptable. Similarly φ_Y, which was a circular measurement, was not normally distributed, since the outer/middle ear is a passive system, which (theoretically) bounds φ_Y by −90 and 90° (Keefe et al., 1993). This non-normality was not able to be improved with transformation, so instead of modelling φ_Y, we modelled Y_n in rectangular form instead with linear mixed models (G_n and B_n). We also modelled $|Y|_t$ and $|Y|_n$ to assess the effects of age on $|Y|$ after removing the variability accounted for by ear canal area. S was also modelled to investigate age effects on acoustically estimated ear canal area.

Preliminary $|Y|_t$, $|Y|_n$, G_n and B_n models showed a heteroscedastic pattern in the residuals, and the S model, a nonlinear pattern, indicating that transforming these variables before modelling may improve linearity. Various transformations were trailed including log, square root, and Box-Cox. The Box-Cox

method transforms a variable (x) using a parameter λ, such that $x(\lambda) = x^{\lambda} - 1/\lambda$. The value for λ was chosen by searching over a range of possible values (−2 to 2; Box & Cox, 1964). A constant was added to G_n and B_n before transformation to make them positive (+1 for G_n and +2 for B_n). The transform that resulted in the most normal looking residuals was used as the transform for that model. The Box-Cox transformation was used for G_n, and log transform for $|Y|_t$, $|Y|_n$, B_n and S. Residuals for all models appeared suitably normal after transformation. Model terms (variables and interactions) were analysed for statistical significance with χ^2 tests by comparing two models, one with the variable or interaction, and one without. Significant terms were further analysed by plotting estimated marginal (EM) means (a.k.a least squares means) with their corresponding 95% confidence intervals. Variables that had been transformed were first back-transformed onto the original scale before plotting to aid interpretability.

Analyses were performed with R (R Core Team, 2017) extended with the *glmmTMB*, *lme4*, *lmerTest*, *emmeans*, and *ggplot2* packages (Bates, Mächler, Bolker, & Walker, 2015; Brooks et al., 2017; Kuznetsova, Brockhoff, & Christensen, 2017; Lenth, 2018; Wickham, 2009). The data and code for the experiments are available online (https://github.com/Josh-Myers/Longitudinal-WAI).

5.4 Results

5.4.1 WAI results

Median A for each age group as a function of frequency is shown in Figure 5.1. A measured in neonates was higher than other age groups from 226 to 2200 Hz and 6200 to 8000 Hz. Average A increased with age from 3000 to 5500 Hz. Figure 5.2 depicts median Y_t in both polar (panels A and B), and rectangular form (panels C and D). Median $|Y|_t$ increased with age across almost the entire frequency range (Figure 5.2A). Neonates had lower median φ_Y compared with other age groups up to 3200 Hz, and higher φ_Y from 3200 to 7000 Hz. From 1800 to 3000 Hz median φ_Y increased with age, and from 3200 to around 6000 Hz, median φ_Y decreased as age increased (Figure 5.2B). Median G_t generally increased with age, particularly in the 3000 to 6000 Hz region (Figure 5.2C). The higher G_t for neonates below 500 Hz was consistent with energy loss through the compliant neonatal ear canal wall (Keefe et al., 1993). B_t systematically increased with age from 400 to 3500 Hz, and then generally decreased with age from 4000 to 6000 Hz. A zero crossing resonance was observed at around 6000 Hz for 6-month-old infants, which shifts to 4500 Hz by 18 months of age (Figure 5.2D).

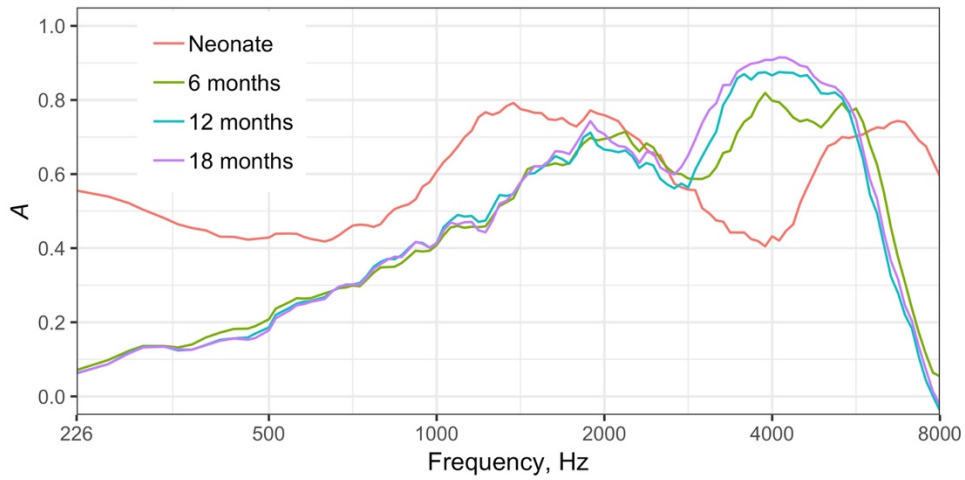


Figure 5-1. Median absorbance (A) for each age group plotted at 1/24 octave frequency resolution.

Y_n results are presented in Figure 5.3, and show that much of the variability in $|Y|_t$, G_t and B_t between age groups can be explained by growth in ear canal area (Keefe & Levi, 1996). Compared to other age groups, neonates had higher median $|Y|_n$ and G_n up to around 2000 Hz (Figure 5.3A and B), similar to median A results. All normalized measures showed increasing Y_n with age around the 3000 Hz region. B_n results showed that almost all of the variability in B_t up to 1200 Hz could be accounted for by ear canal area. Median B_n demonstrated increasing stiffness with age from 2000 to 3500 Hz, then generally decreasing at higher frequencies as age increased (Figure 5.3C). The middle ear efficiently transmits energy when $G > B$ (Allen et al., 2005). Figure 5.4 shows $G_n - B_n$ for each age group. Regions of resonance are indicated where $G_n - B_n > 0$. These regions correspond to areas where φ_Y was less than 45° (Keefe et al., 2015). For neonates, there are 3 regions where $G_n > B_n$: 226 to 600 Hz, 1000 to 3200 Hz, and 7000 to 7800 Hz. For the other age groups there are 2 regions: around 2000 Hz, and then 3500 to 7000 Hz for 6 months, and 3200 to 7000 at 12 and 18 months. These regions relate to maxima in median A for each age group (Figure 5.1), and may represent diagnostically important frequencies, apart from the low frequency region for neonates, which is caused by a resonance in the ear canal wall (Hunter et al., 2010; Keefe et al., 1993; Keefe & Levi, 1996; Myers et al., 2018a, 2018b; Prieve et al., 2013b).

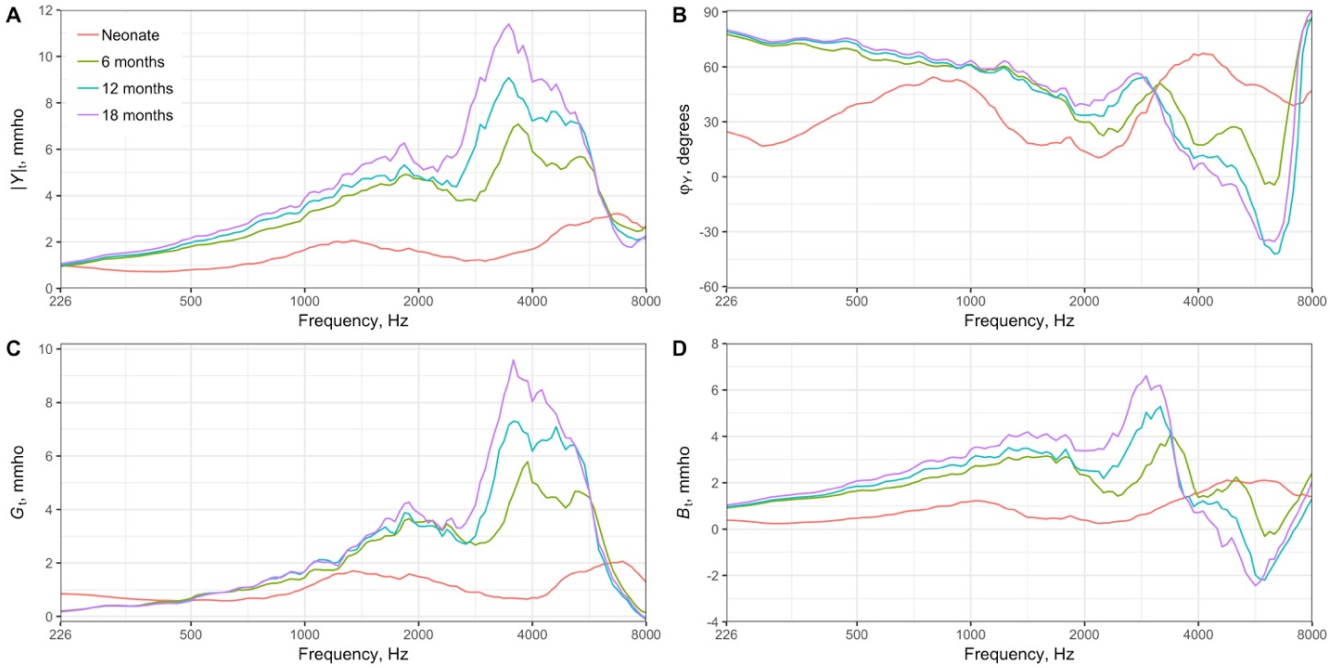


Figure 5-2. Median admittance at the probe tip (Y_t) for each age group.
Results are depicted in both polar (A and B) and rectangular form (C and D), plotted at 1/24 octave frequency resolution. *B* = susceptance; *G* = conductance; φ_Y = admittance phase angle; $|Y|$ = admittance magnitude.

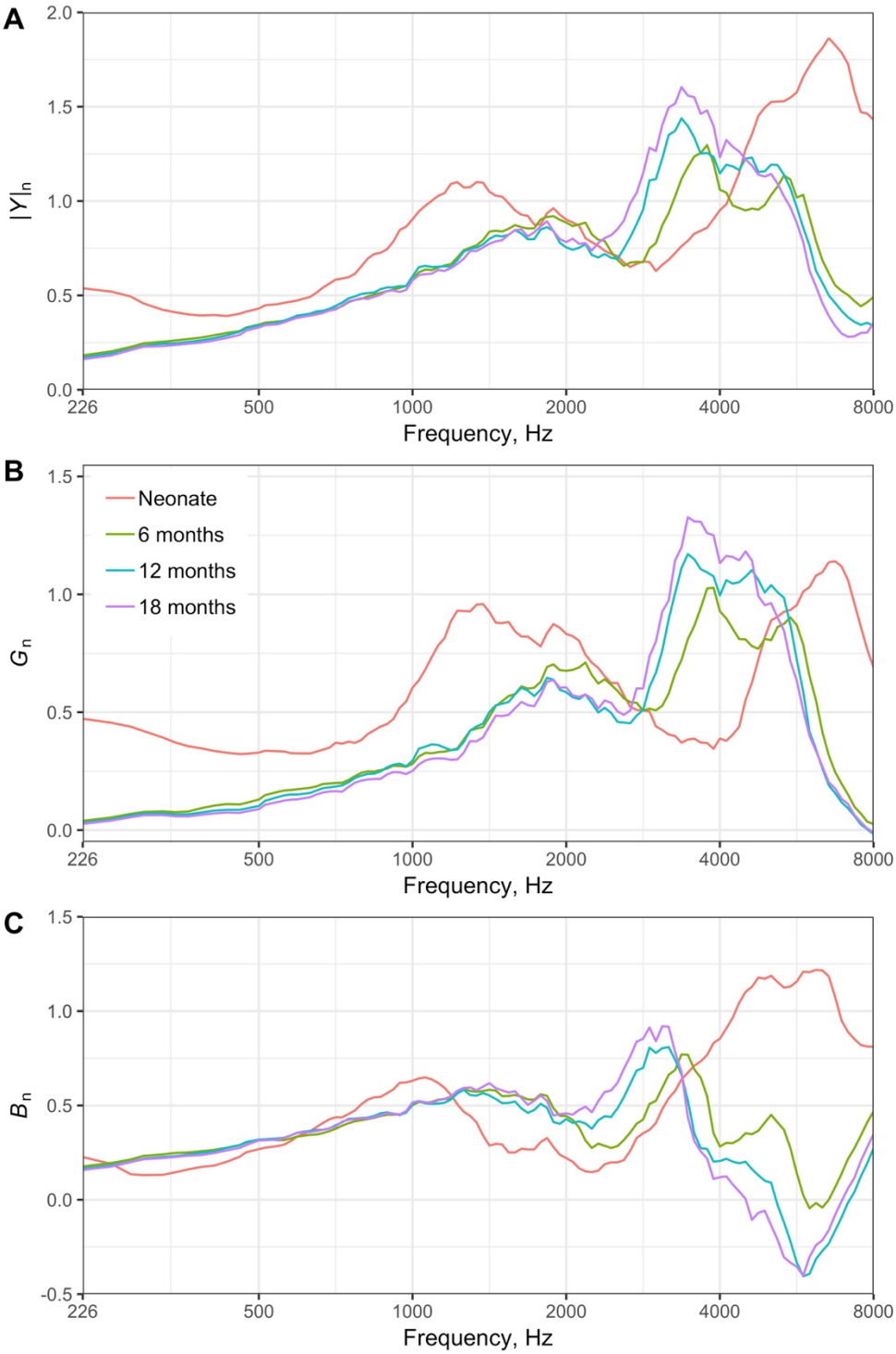


Figure 5-3. Median normalized admittance for each age group.
The top panel (A) shows admittance magnitude ($|Y|_n$), middle (B) conductance (G_n), and bottom (C) susceptance (B_n). Plots are at 1/24 octave frequency resolution.

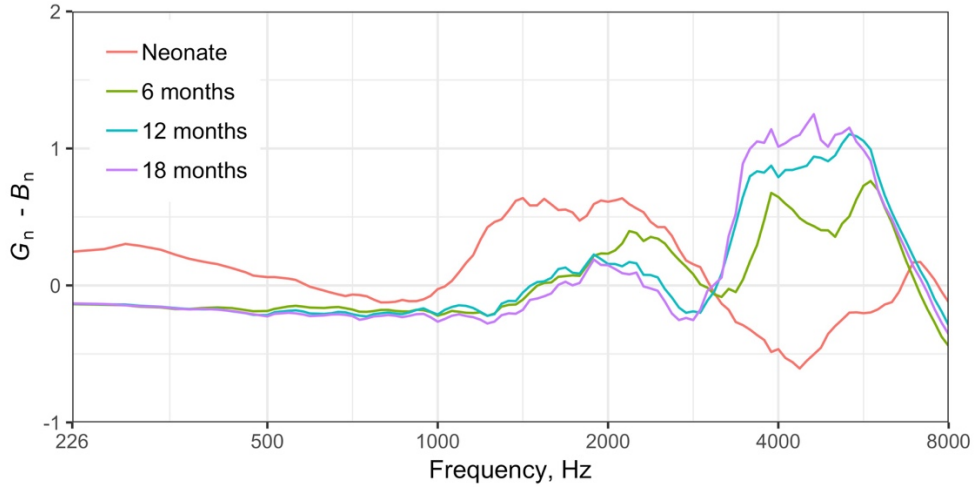


Figure 5-4. Normalized conductance (G_n) − susceptance (B_n) by age group.
Regions where $G_n > B_n$ occur when $G_n - B_n > 0$. Data are plotted at 1/24 octave resolution.

5.4.2 Statistical models

Results of statistical analyses testing model terms (variables and interactions) are presented in Table 5.4. Age, ethnicity, and frequency were significant for all WAI models (A, $|Y|_t$, $|Y|_n$, G_n and B_n). Interactions between age, ethnicity and frequency were also significant for all of these models, except for the 3-way interaction (Age × Ethnicity × Frequency) for the G_n and B_n models, as well as the Age × Ethnicity interaction in the B_n model. For the S model, only age was significant. To assess which ages and frequencies were significant, and the magnitude of effects, EM means and their 95% confidence intervals were plotted. Non-overlapping confidence intervals between groups is evidence of a statistically significant effect. Results were back-transformed onto the original scale before plotting to aid interpretation.

Table 5.4. Results of statistical significance testing

	Model	Statistic	Age	Ethnicity	Ear	Gender	f	Age × Ethnicity	Age × f	Ethnicity × f	Age × Ethnicity × f
A	β GLMM	χ^2	5138.58	99.27	0.05	0.68	9545.56	65.48	4074.33	77.11	47.35
		DF	66	44	1	1	80	33	60	40	30
		p-value	<0.001	<0.001	0.829	0.410	<0.001	<0.001	<0.001	<0.001	<0.001
$\|Y\|_t$	LMM T = Log	χ^2	7893.50	100.35	0.19	2.31	10501.89	60.34	2955.61	80.00	45.23
		DF	66	44	1	1	80	33	60	40	30
		p-value	<0.001	<0.001	0.659	0.129	<0.001	0.003	<0.001	<0.001	0.037
$\|Y\|_n$	LMM T = Log	χ^2	4879.88	91.60	0.34	0.87	11383.02	52.05	3286.92	90.42	51.14
		DF	66	44	1	1	80	33	60	40	30
		p-value	<0.001	<0.001	0.558	0.351	<0.001	0.019	<0.001	<0.001	<0.001
G_n	LMM T = BC	χ^2	5849.90	80.17	0.018	2.89	10033.78	53.07	4042.54	70.09	42.99
		DF	66	44	1	1	80	33	60	40	30
		p-value	<0.001	<0.001	0.894	0.089	<0.001	0.015	<0.001	0.002	0.059
B_n	LMM T = Log	χ^2	3615.00	89.49	<0.01	0.62	4089.04	35.42	3282.18	83.30	35.05
		DF	66	44	1	1	80	33	60	40	30
		p-value	<0.001	<0.001	0.982	0.430	<0.001	0.355	<0.001	<0.001	0.241
S	LMM T = Log	χ^2	1460.64	6.55	0.25	0.90		2.12			
		DF	6	4	1	1		3			
		p-value	<0.001	0.161	0.618	0.344		0.549			

Statistics are shown for ear model term (i.e., variables and interactions). Transformations (T) used on variables before modelling are provided in the Model column. Constants of 1 and 2 were added to G_n and B_n, respectively, to make all values positive so they could be transformed. For the G_n Box-Cox transform, $\lambda = -1.39$. A, absorbance; B_n, normalized susceptance; β GLMM, beta generalized linear mixed model; χ^2, chi-squared statistic; DF, degrees of freedom; f, frequency; G_n, normalized conductance; LMM, linear mixed model; S, acoustically estimated ear-canal area; $|Y|_n$, normalized admittance magnitude; $|Y|_t$, admittance magnitude at the probe tip.

The effect of age on EM means for A, $|Y|_t$ and Y_n is depicted in Figure 5.5. For A, neonates were significantly different from other age groups except for 2000 to 3000, and 5000 to 6000 Hz, and the size of the effect was larger than differences between other age groups (Figure 5.5A). Note that our interpretation interpolates between 1/2 octave data points (e.g., 5000 Hz was not a variable in the model), but we think this is reasonable, as the shape of the WAI response is similar for 1/24 and 1/2 octave frequency resolutions (Myers et al., 2018b). All age groups were significantly different from each other at 4000 Hz, and 12-month-old infants were different from other age groups at 6000 Hz. For $|Y|_t$, neonates were significantly different at all frequencies except for 7000 to 8000 Hz (Figure 5.5B). Once normalized for ear canal area, the size of the effect up to 4000 Hz was much smaller (Figure 5.5C). All age groups were significantly and substantially different for $|Y|_t$, from 3000 to 4000 Hz, but once normalized, only 3000 Hz remained significant, and the size of the effect was diminished. The effect of age on EM means for G_n and B_n are presented in Figure 5.6. For G_n, the neonate group was

significantly different across almost the entire frequency range, apart from where results crossed over at around 2500 and 5000 Hz (Figure 5.6A). The 12-month group, was different from other age groups in the 6000 Hz region, and the 18-month group was significantly different from the 6-month group from 3000 to 4000 Hz. For B_n, neonates were significantly different for frequencies above 1500 Hz (apart from around 3200 Hz where there was a cross over), the 6-month group was different from other groups above 3500 Hz, and the 6-month group was different from 18-months at 3000 Hz (Figure 5.6B). EM means for the *S* model are presented in Table 5.5, and show that the neonate group had significantly and substantially lower acoustically estimated ear canal area compared to other age groups, and the 6-month group was slightly but significantly different from the 12- and 18-month groups.

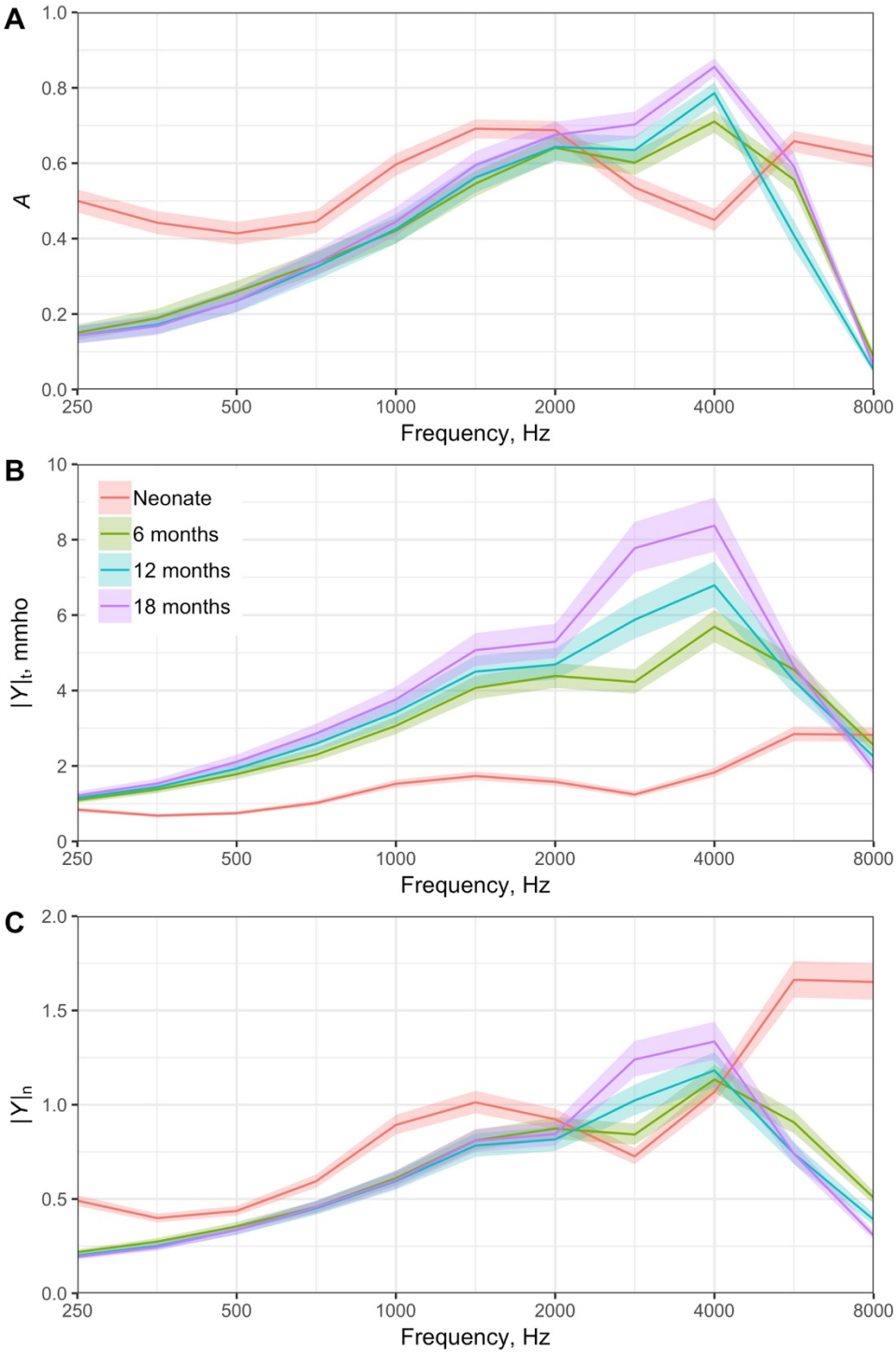


Figure 5-5. Estimated marginal means for absorbance and admittance.
Estimated marginal means (lines) and 95% confidence intervals (shaded areas) for absorbance (*A*; A, top), admittance magnitude at the probe tip ($|Y|_t$; B, middle), and normalized admittance magnitude ($|Y|_n$; C, bottom) for each age group plotted at 1/2 octave frequency resolution.

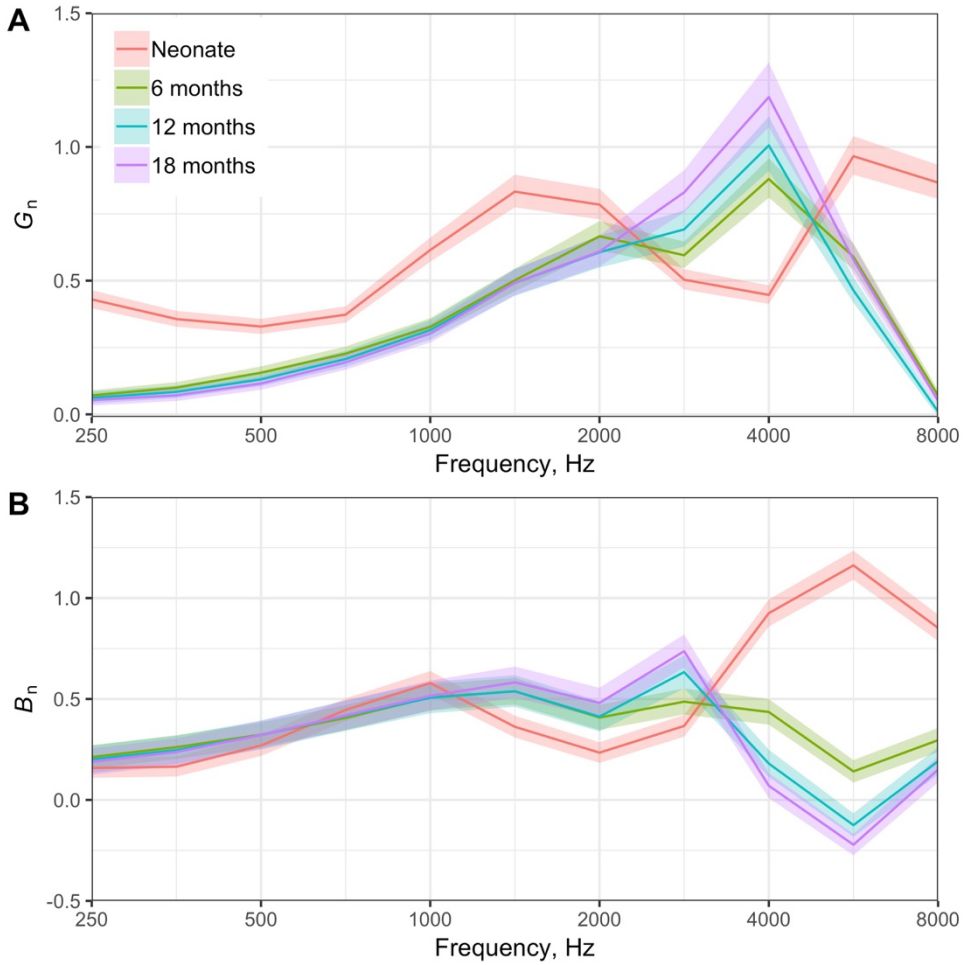


Figure 5-6. Estimated marginal means for normalized conductance and susceptance.
Estimated marginal means (lines) and 95% confidence intervals (shaded areas) for normalized conductance (G_n; A, top), and normalized susceptance (B_n; B, bottom) for each age group plotted at 1/2 octave frequency resolution.

Table 5.5. Estimated marginal mean from the ear-canal area model

	Neonate	6 months	12 months	18 months
Lower	6.71	19.51	22.14	24.34
Mean	7.06	20.64	23.66	25.95
Upper	7.43	21.84	25.28	27.67

Estimated marginal mean and 95% confidence intervals from the ear-canal area model for each age group. Results are in units of mm^2, as they have been back-transformed to the original scale to aid interpretability. "Lower" and "upper" are the lower and upper bounds of the 95% confidence intervals for the estimated marginal means.

To investigate the significant ethnicity effects and interactions, EM means by ethnicity (Caucasian compared to non-Caucasian) for the different age groups as a function of frequency for each model are shown in Figures 5.7 to 5.11. For A, confidence intervals did not overlap for the neonates at 1500 Hz (Figure 5.7), and for the 12-month group there was a large effect (>10% difference) at 6000 Hz. For $|Y|_t$ there was a significant effect at 6000 Hz for the 12-month age group, and 4000 to 8000 Hz for 18

months, which was relatively large from 4000 to 6000 Hz (Figure 5.8). For $|Y|_n$, the only significant effect was at 6000 Hz for the 12- and 18-month groups (Figure 5.9), and for G_n at 6000 Hz for the 12-month group (Figure 5.10). For B_n, there was an effect at 4000 Hz, and 6000 to 8000 Hz for 12 months, and 4000 and 6000 Hz for 18 months (Figure 5.11). Age could be a confounding factor in these results if one ethnic group was generally tested at a younger or older age within an age group. Therefore, *t*-tests were conducted for each age group to investigate whether there were significant differences in age between Caucasians and non-Caucasians within age groups (Hunter et al., 2016). The results are presented in Table 5.6 and show that age was not a significant factor between ethnic groups for any age group.

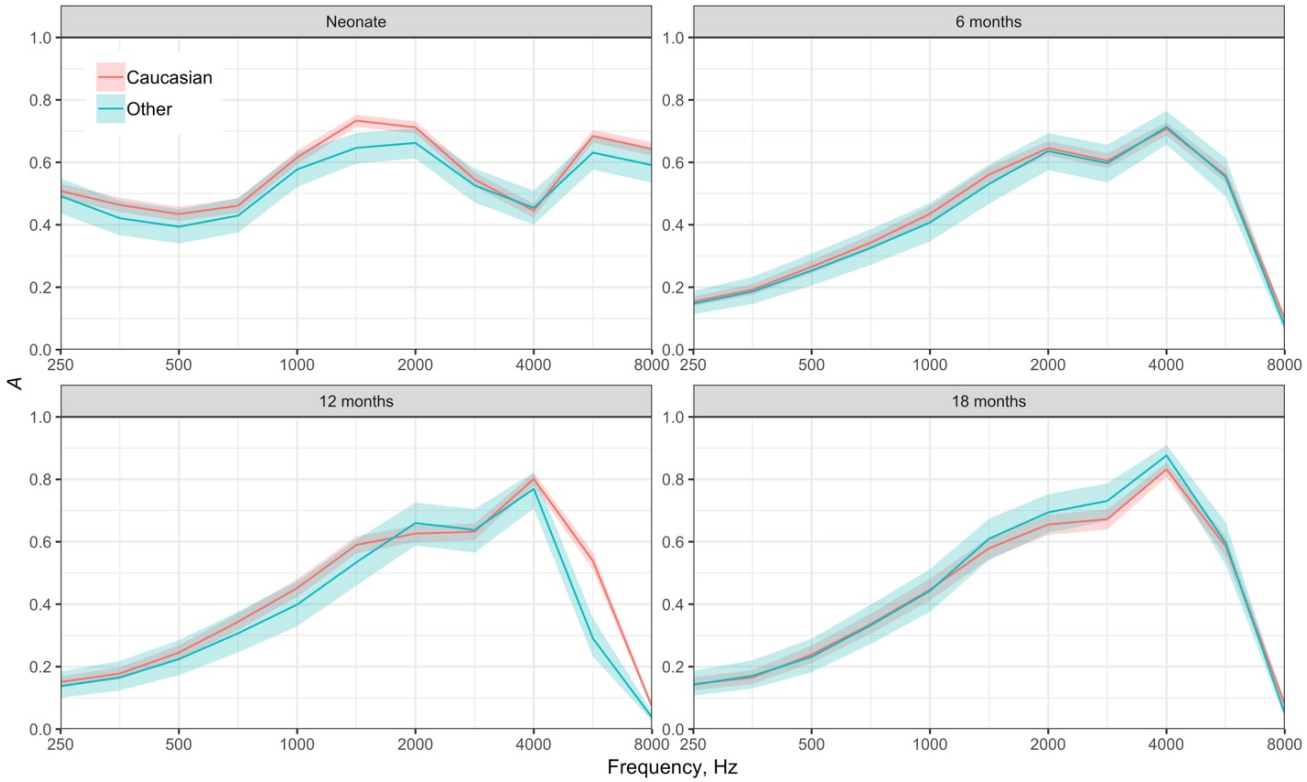


Figure 5-7. Absorbance estimated marginal means showing the ethnicity interaction.
Estimated marginal means (lines) and 95% confidence intervals (shaded areas) from the absorbance (*A*) model by age group showing the relationship between ethnicity, age and frequency.

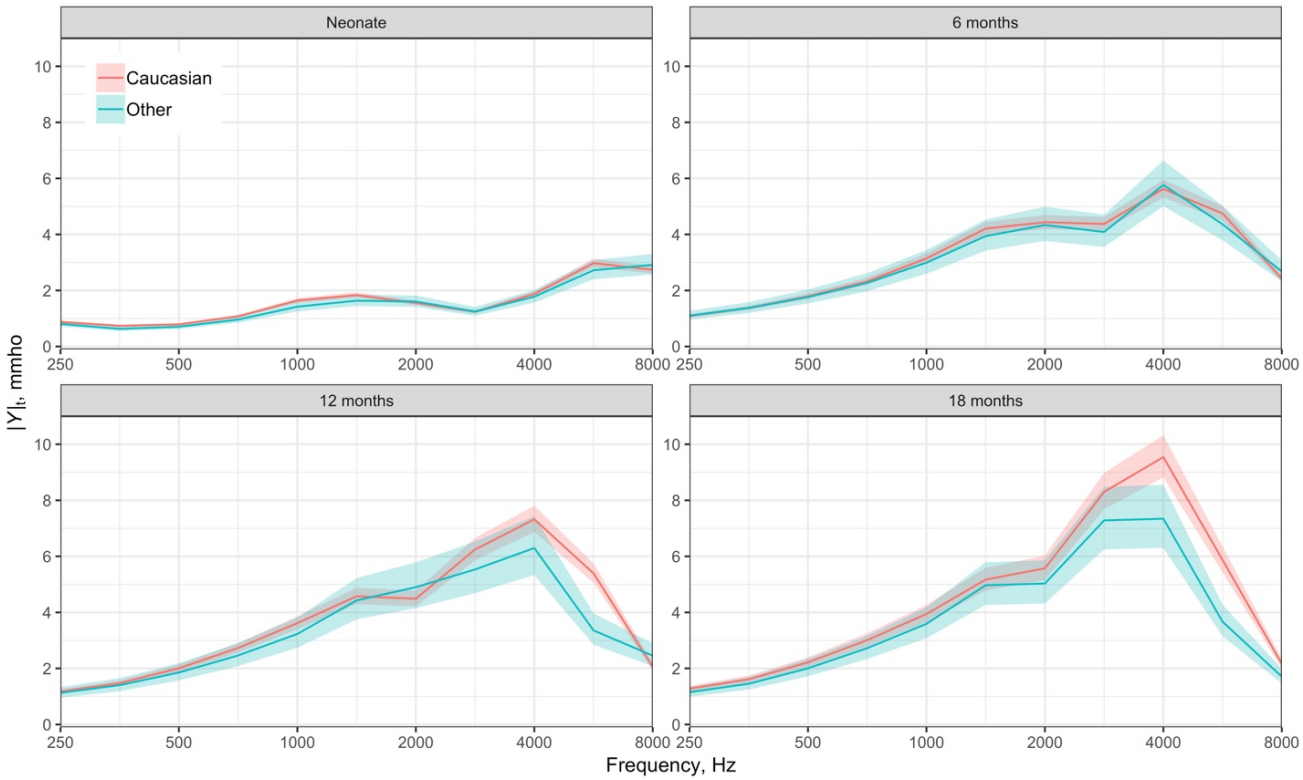


Figure 5-8. Admittance magnitude at the tip estimated marginal means showing the ethnicity interaction.
Estimated marginal means (lines) and 95% confidence intervals (shaded areas) from the admittance magnitude at the probe tip ($|Y|_t$) model by age group showing the relationship between ethnicity, age and frequency.

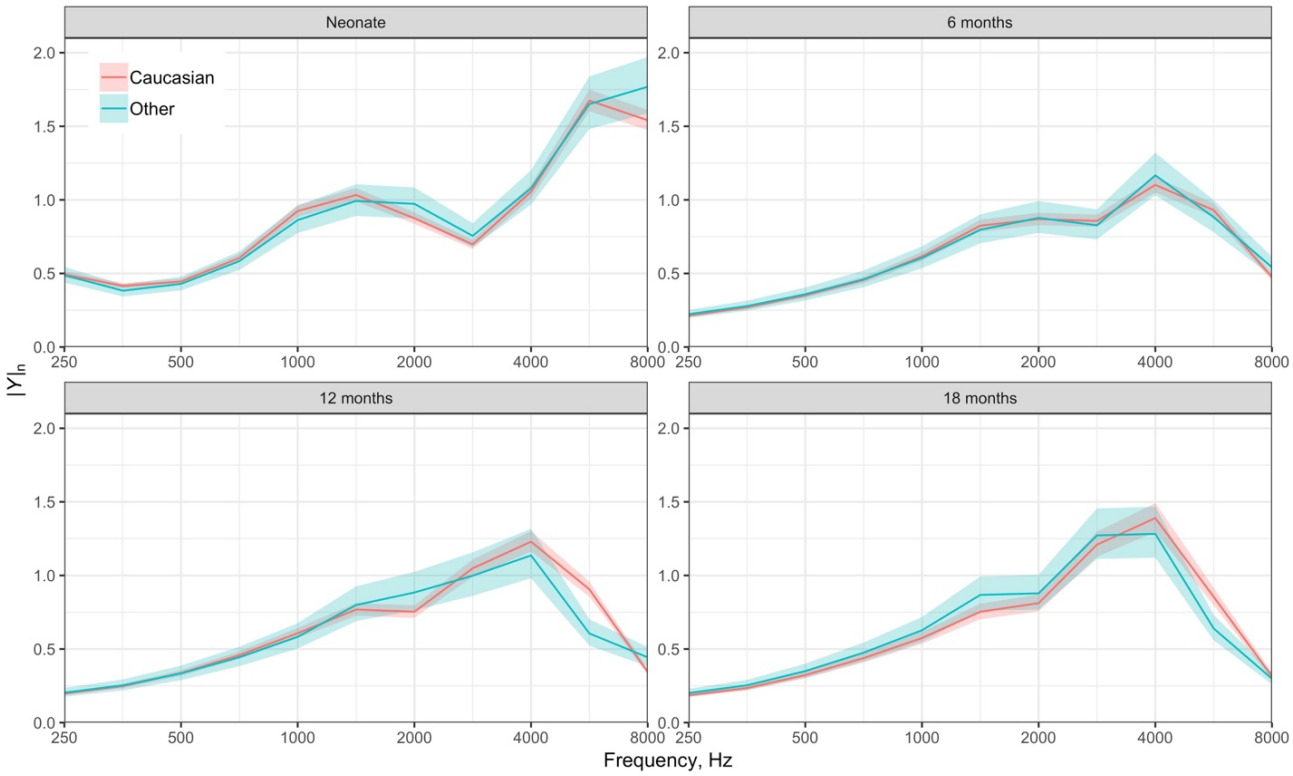


Figure 5-9. Normalized admittance magnitude estimated marginal means showing the ethnicity interaction.
Estimated marginal means (lines) and 95% confidence intervals (shaded areas) from the normalized admittance magnitude ($|Y|_n$) model by age group showing the relationship between ethnicity, age and frequency.

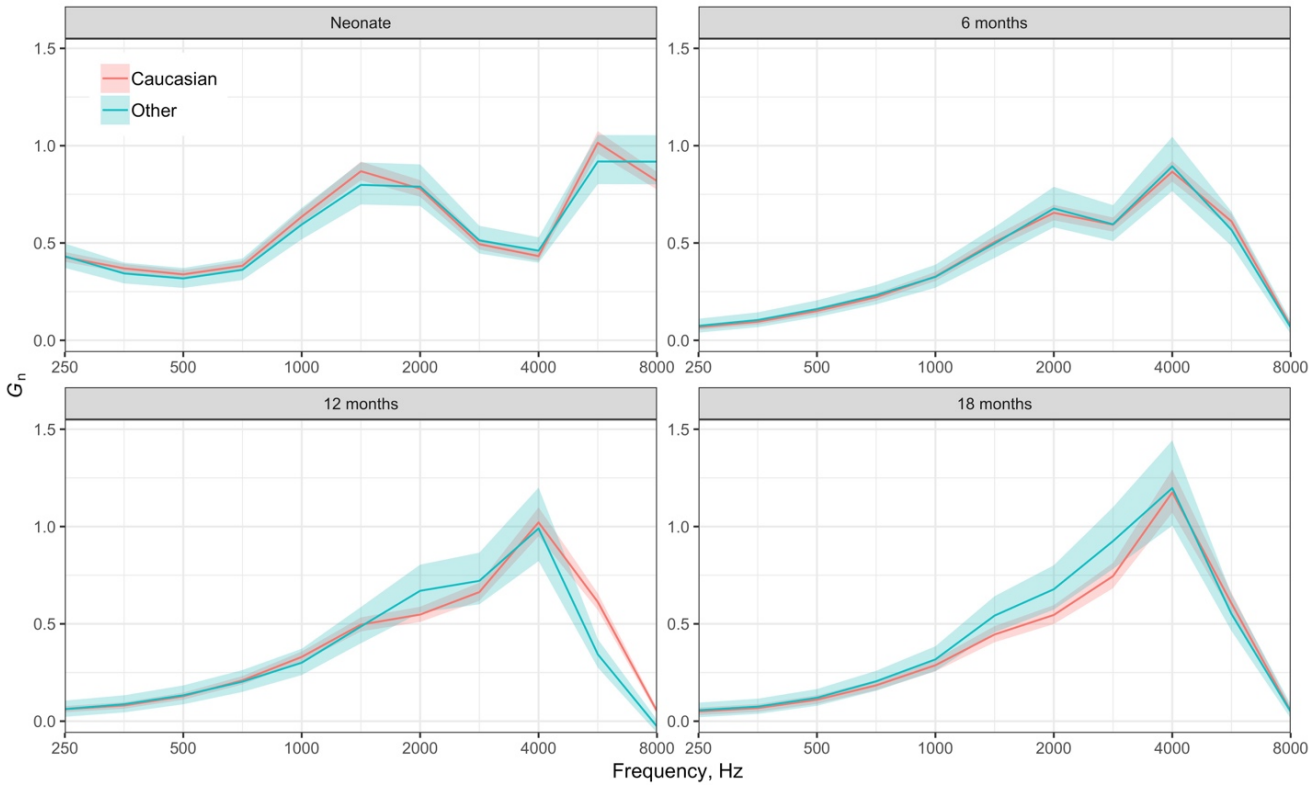


Figure 5-10. Normalized conductance estimated marginal means showing the ethnicity interaction.
Estimated marginal means (lines) and 95% confidence intervals (shaded areas) from the normalized conductance (G_n) model by age group showing the relationship between ethnicity, age and frequency.

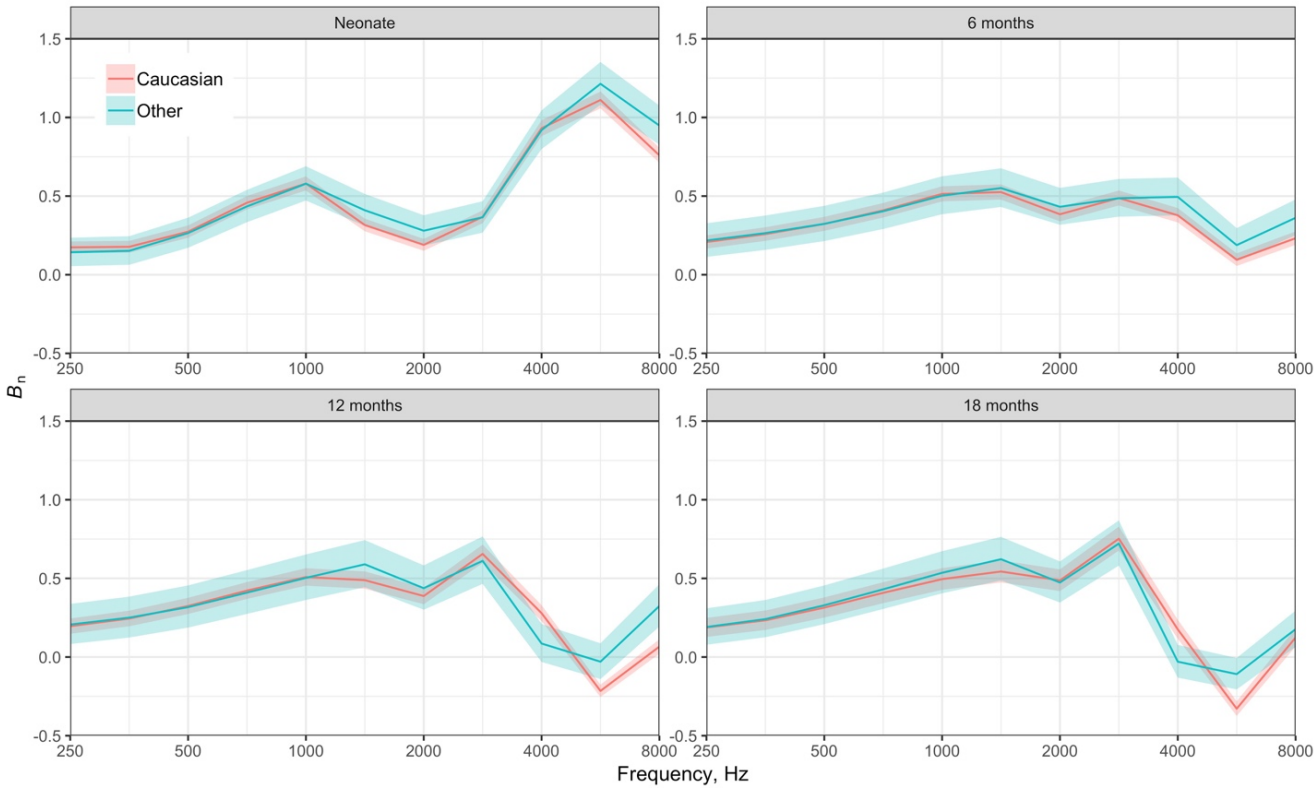


Figure 5-11. Normalized susceptance estimated marginal means showing the ethnicity interaction.
Estimated marginal means (lines) and 95% confidence intervals (shaded areas) from the normalized susceptance (B_n) model by age group showing the relationship between ethnicity, age and frequency.

Table 5.6. Results of *t*-tests between age group and ethnicity

Age	*t*-value	DF	*p*-value
Neonate	0.77	36.25	0.446
6 months	1.03	27.31	0.312
12 months	0.46	18.63	0.651
18 months	−1.16	32.98	0.254

Results of *t*-tests investigating whether the relationship between age group and ethnicity was statistically significant within age groups. DF, degrees of freedom.

An interactive web application is available (https://joshmyers.shinyapps.io/WAINorms/), which has the normative reference data for each age group (overall or ethnic-specific) for various frequency resolutions. The application also shows the effect of frequency resolution on group average results, and has results from individual infants that attended all of the follow up sessions.

5.5 Discussion

5.5.1 Comparison of WAI measurements with previous studies

Median A, $|Y|_t$ and φ_Y from this study are depicted along with results from selected studies in Figures 5.12 to 5.14. Average A in this study generally compared well with other studies (Figure 5.12). For neonates, A results from our study are very similar to the majority of other reports, with three maxima at around 250, 1500–2000, and 6000–7000 Hz. For the 6-month age group, results from our study are generally lower than other studies, but overall compare well. Average A for the 12- and 18-month age groups compared favourably with previous research also. Median $|Y|_t$ in this study compared well with other studies, although $|Y|_t$ for our study was lower than other reports above 4000 Hz for the 6- to 18-month age groups (Figure 5.13). Average φ_Y for this study generally compared well up to 2000 Hz for all age groups, but at higher frequencies, our φ_Y results were higher than reported by other research for the 6- to 18-month groups (Figure 5.14). Table 5.7 compares S from this study with other reports, and shows that our findings were smaller than those of Keefe et al. (1993), but were very similar to results from Hunter et al. (2016). However, Hunter et al. reported an Ethnicity × Age effect, which was not significant in our study. Differences in the Hunter et al. study were for infants mostly of African descent, though, while the ethnicity was majority Asian descent (other than Caucasian) in the present study. Other differences between results in this study and previous research could be due to a range of factors including differences in probe design and equipment, calibration procedures, methods for estimating ear canal area, age groupings, study design, and methodology (Aithal et al., 2014b; Shahnaz et al., 2014).

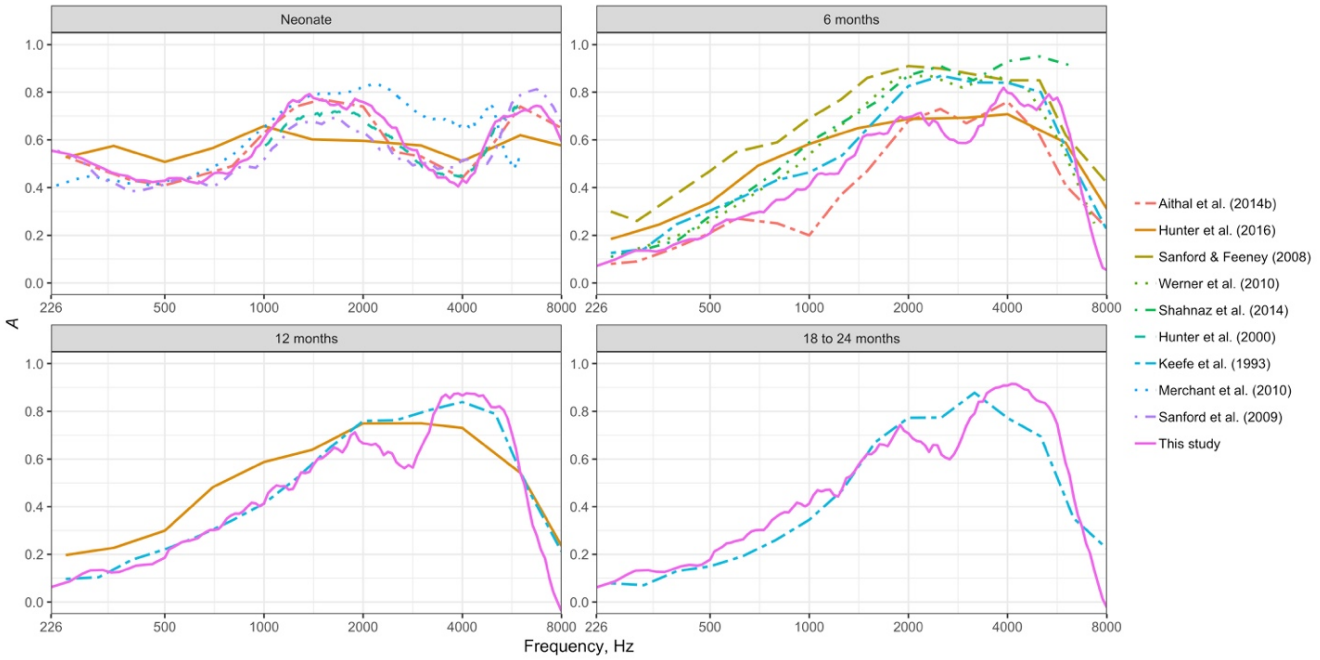


Figure 5-12. Comparison of absorbance with other studies.

Median absorbance (*A*) for this study compared with results from other studies reporting ambient *A* or reflectance (*R*) at similar ages. Aithal et al. (2014b) measured median 1/3 octave *A* from 35 ears from 35 neonates, and 27 ears of 14 infants aged 6 months; Hunter et al. (2016) presented 1/2 octave *A* estimated marginal means from one ear of 129 neonates, 95 infants aged 6 months, and 93 infants aged 12 months; Sanford and Feeney (2008) reported 1/3 octave *R* for one ear of 20 infants aged 6 months; Werner et al. (2010) measured 1/12 octave *R* averaged into 15 analysis bands from 260 infants aged 5 to 9 months; Shahnaz et al. (2014) measured 1/3 octave *R* in 33 ears of 6-month-old infants; Hunter et al. (2010) reported median *R* data at 23 Hz frequency resolution for 324 neonates; Keefe et al. (1993) presented mean 1/3 octave *R* for 11 infants aged 6 months, 23 aged 12 months, and 11 aged 24 months; Merchant et al. (2010) measured mean *R* at 23 Hz frequency resolution in 8 neonates; and Sanford et al. (2009) results are median 1/12 octave *A* from 53 neonatal ears measured on the second day of life. Studies that reported *R* have been converted to *A* for ease of comparison. For the 18 to 24 months panel, infants in the this study were approximately 18 months old and participants in Keefe et al. (1993) were 24 months of age. Results from this study are presented at 1/24 octave frequency resolution.

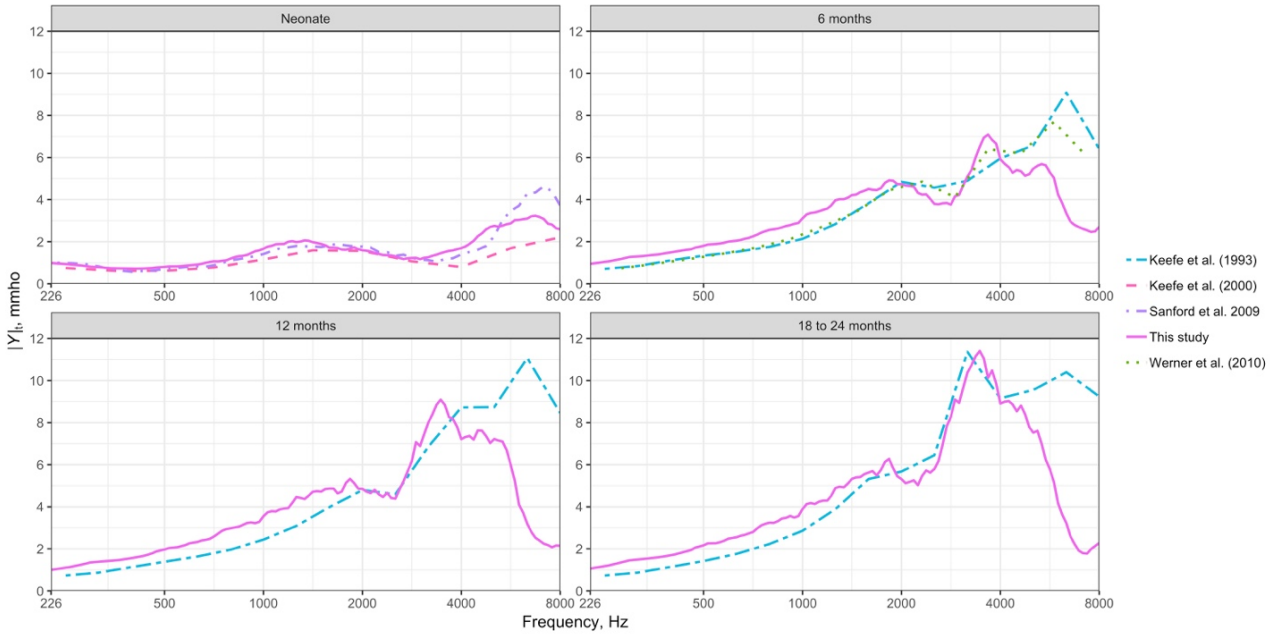


Figure 5-13. Comparison of admittance magnitude with other studies.
Median $|Y|_t$ from this study compared with results from other studies that measured Y_t or Z_t at similar ages. Keefe et al. (1993) presented mean 1/3 octave Z_t for 11 infants aged 6 months, 23 aged 12 months, and 11 aged 24 months; Keefe et al. (2000) reported median 1/2 octave Y_t from 2081 neonatal ears; Sanford et al. (2009) results are median 1/12 octave Y_t from 53 neonatal ears measured on the second day of life; and Werner et al. (2010) reported 1/12 octave Z_t averaged into 15 analysis bands from 260 infants aged 5 to 9 months. Studies reporting Z_t have been converted to Y_t for ease of comparison. For the 18 to 24 months panel, infants in the this study were approximately 18 months old and participants in Keefe et al. (1993) were 24 months of age. Results from this study are presented at 1/24 octave frequency resolution. Y_t = admittance measured at the probe tip; Z_t = impedance measured at the probe tip.

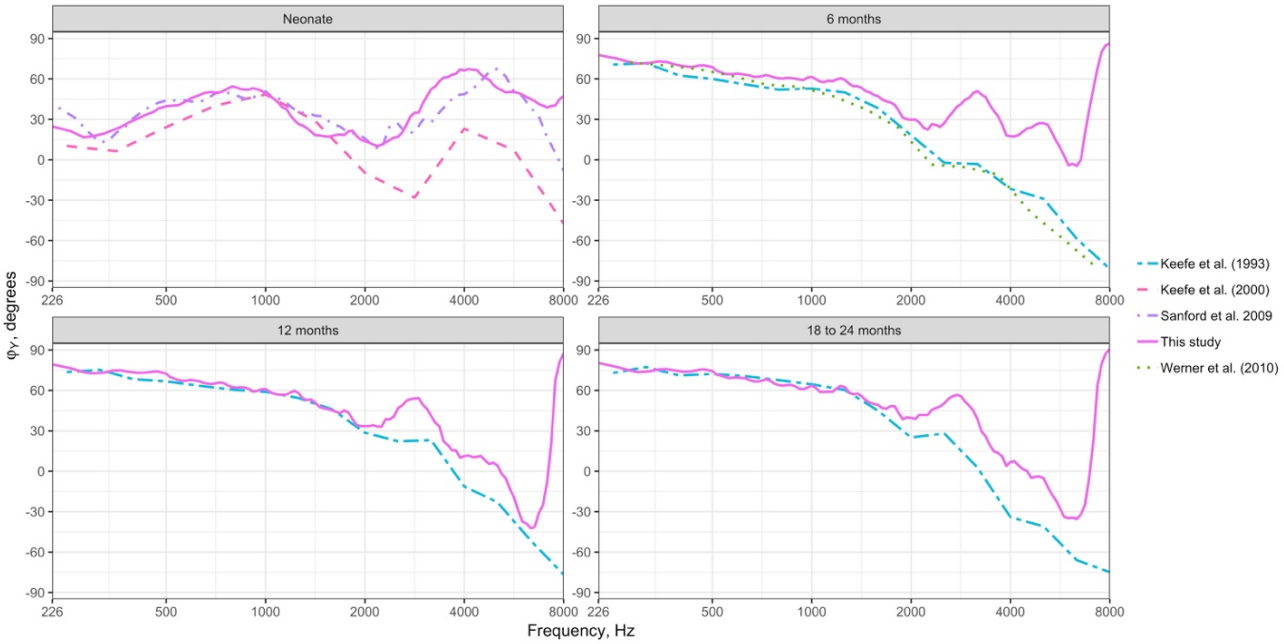


Figure 5-14. Comparison of admittance phase angle with other studies.
Median admittance phase angle (φ_Y) from this study compared with results from other studies that measured admittance or impedance at similar ages. Studies reporting impedance have been converted to admittance to facilitate comparison. Details of the studies are given in Figure 5.13. Results from this study are depicted at 1/24 octave resolution.

Table 5.7. Comparison of average acoustic ear-canal area estimate with other studies

	Hunter et al. (2016)	Keefe et al. (1993)	This study
Neonate	6		7.51
6 months	24	31.17	22.22
12 months	22	38.48	25.66
18 to 24 months		47.57	28.76

Units are in mm^2. Estimates from this study and Keefe et al. (1993) are the mean, and results from Hunter et al. (2016) are estimated marginal means. For the 18 to 24 months, infants in this study were approximately 18 months old and participants from Keefe et al. were 24 months of age.

5.5.2 Developmental effects on WAI

Developmental effects on *A* and *Y* in this study compared well with results from previous research, although there were differences. In our study, from birth to 6 months of age, *A* increased in the 4000 Hz region, and decreased below 2000 Hz and above 6000 Hz (Figure 5.1) in agreement with most published reports (Aithal et al., 2014a; Hunter et al., 2016; Keefe et al., 1993; Sanford & Feeney, 2008; Shahnaz et al., 2014). However, our data showed a significant effect of increasing *A* with age from 6 to 18 months at 4000 Hz, which has not been found in other studies (Hunter et al., 2016; Hunter et al., 2008b; Keefe et al., 1993). Our $|Y|_t$ results compared favourably with other studies, with a trend of

increasing $|Y|_t$ with age (Keefe et al., 1993; Sanford & Feeney, 2008; Werner et al., 2010). However, average $|Y|_t$ in our study decreased substantially beyond 4000 Hz for the 6- to 18-month age groups, whereas in Keefe et al. (1993) it continued to trend upwards until above 6000 Hz. For φ_Y, our results showed a general pattern of decreasing zero crossing with age, largely consistent with results from Keefe et al., although in our study this occurred at a higher frequency. We found a significant ethnicity effect and interaction with age, for both A and Y, in agreement with Hunter et al. (2016), and the effect was substantial for certain combinations of age, ethnicity and frequency (e.g., A at 6000 Hz for 12 month-old infants; see Figure 5.7).

5.5.3 Maturational factors contributing to developmental changes in WAI

Consistent with previous reports, we have found that growth and development of the outer and middle ear significantly affect WAI measured in infants (Aithal et al., 2014b; Hunter et al., 2016; Keefe et al., 1993; Sanford & Feeney, 2008; Shahnaz et al., 2014; Werner et al., 2010). Contributing factors include stiffening of the ear canal, growth in ear canal length and diameter, clearance of mesenchyme and fluid from the middle ear, growth of tympanic cavity, and growth and development of the ossicles (Keefe et al., 1993). The higher A, $|Y|_n$ and G_n for neonates at low frequencies observed in this study is consistent with energy absorbed by the compliant and lossy neonatal ear canal wall (Aithal et al., 2014b; Hunter et al., 2016; Keefe et al., 1993; Keefe & Levi, 1996; Sanford & Feeney, 2008; Shahnaz et al., 2014). It is generally accepted that the strongest effect is for frequencies below 1000 Hz, which is consistent with our results (Keefe et al., 1993; Keefe & Levi, 1996; Sanford & Feeney, 2008; Shahnaz et al., 2014). The effect decreases over the first few months of life as the ear canal wall stiffens and ossifies (Keefe et al., 1993; Wilson, 2012).

The degree to which changes in ear canal area explain differences in Y_t measurements between age groups can be seen by comparing Y_t with Y_n (Keefe & Levi, 1996). In our study, differences in ear canal area explained most of the variability in B_t up to 1200 Hz for all age groups, and for $|Y|_t$, almost all of the variability from 6- to 18-months up to 2000 Hz. We conclude that ear canal area growth was an important factor for differences in Y_t between age groups, consistent with the results of Keefe et al. (1993) and Keefe and Levi (1996). Ear canal length increases with age through infancy, if this significantly contributed to differences in $|Y|_n$, we would expect smaller $|Y|_n$ at low frequencies for younger infants, since $|Y|_n = kL / 2$, where L is ear canal length, and k is the ratio of the angular frequency ($2\pi f$) to the speed of sound (Keefe et al., 1993; Keefe, Bulen, Campbell, & Burns, 1994). Given that $|Y|_n$ was larger for neonates than other age groups, and there was no difference from 6- to 18-months at low

frequencies, we conclude that growth in ear canal length was not a significant factor in differences in $|Y|_n$ observed in our study, consistent with the results of Keefe et al. (1993).

The mass and resistance of the middle ear decrease as mesenchyme, amniotic fluid and debris clear after birth (Kei et al., 2013; Wilson, 2012). In our study, the greater mass for neonates compared with other age groups from 1200 to 3500 Hz for B_n (Figure 5.3C), and increase in G_n from 3000 to 5000 Hz is likely at least in part due to aeration of the middle ear cavity (Figure 5.3B). The increasing stiffness with age for B_n from 2000 to 3500 Hz is consistent with increasing stiffness of the middle ear possibly due to fusion of the tympanic ring, changes in the orientation and fibres of the tympanic membrane, and tightening of the ossicular joints (Figure 5.3C; Kei et al., 2013; Wilson, 2012). Growth of the middle ear cavity is thought to be a contributing factor to increasing Y with age (Holte, Margolish, & Cavanaugh, 1991). The tympanic cavity increases in size through infancy, and its volume also increases due to pneumatization, as mesenchyme is absorbed and amniotic fluid and debris clear (Wilson, 2012). This growth could be a contributing factor to increasing $|Y|_n$ with age from 3000 to 4000 Hz in our study (Figure 5.3A). The increase in mass with age above 3700 Hz for B_n could also be due to increase in middle ear size, as a smaller cavity is acoustically stiffer (Figure 5.3C; Keefe et al., 1993; Keefe & Levi, 1996). At high frequencies (above 6000 Hz), the increase in mass seen for B_n may be due to development of the ossicles, which change in orientation as the orientation of the eardrum changes, increase in size and weight, and ossify throughout infancy (Aithal et al., 2014b; Allen et al., 2005; Keefe et al., 1993; Sanford & Feeney, 2008; Shahnaz et al., 2014; Wilson, 2012).

5.5.4 Clinical significance

Regions of resonance are indicated where $G > B$, and these areas related to maxima in median A for each age group (Allen et al., 2005). Therefore, A at frequencies where $G_n > B_n$ may be important predictors of middle ear status (Figure 5.4), apart from the low frequency region for neonates (<600 Hz), which is caused by an ear canal resonance (Keefe et al., 1993). This would indicate that A in the 1500 to 2000 Hz region is diagnostically important for neonates, which is consistent with previous reports (Aithal et al., 2015; Hunter et al., 2010; Myers et al., 2018a). For older age groups our data indicate that 2000 Hz may also be important, as well as A at frequencies from 3500 to 6000 Hz, which is also generally consistent with results of other studies (Keefe et al., 1993; Myers et al., 2018b; Prieve et al., 2013b).

There is a general consensus in the WAI literature that separate reference data are needed for different age groups (Kei et al., 2013). However, exactly which ages require their own normative data is an area of ongoing research. Aithal et al. (2014b) and Hunter et al. (2016) recommended separate

normative *A* regions be used for neonates compared to 1-month old infants, but Merchant et al. (2010) found no significant differences between these age groups. Hunter et al. also suggested that *A* reference data could be collapsed for infants aged 6 to 15 months. Our *A* results, however, indicate a systematic age effect at 4000 Hz, with a reasonably large difference in EM means between the 6- and 18-month groups (>10%; Figure 5.5A), indicating that it may be beneficial to have separate normative data for 6-month-old infants. The difference between 12 and 18 months at 4000 Hz was also statistically significant, but overall, it seems reasonable to collapse reference data across those age groups. As mentioned, a possible reason for differences between studies could be different methods for estimating ear canal area. In our study, when calculating *A*, the Titan system used the area of the calibration tube as the estimated ear canal area, whereas Hunter et al. used an acoustic estimate of ear canal area.

Most clinical WAI research has been focused on *A* and *R* since these measures are theoretically insensitive to probe location in the ear canal. There is interest, however, in using *Y* as well as *A* in diagnostic applications both for objectively interpreted predictive models (Ellison et al., 2012; Myers et al., 2018a; Piskorski et al., 1999; Sanford et al., 2009), and subjectively interpreted visual displays of results (Aithal et al., 2017; Allen et al., 2005; Sanford & Brockett, 2014; Sanford & Feeney, 2008; Voss et al., 2016). Knowledge of developmental effects on *Y* are essential for both of these applications. Our results indicate that for subjective interpretation of results, different reference data would be necessary for each age group (neonates, 6, 12, and 18 months) for both Y_t and Y_n. Likewise, when using *Y* as a variable in a predictive model, these age effects should be taken into account by either creating age-specific models, or including interaction terms that allow the interpretation of *Y* to vary with age.

For all WAI measures there were also substantial ethnicity effects for certain Age × Frequency interactions. For example, there was over 10% difference between EM means for Caucasian and non-Caucasians for *A* at 6000 Hz at 12 months, and for $|Y|_t$ at 4000 to 6000 Hz at 18 months, indicating that having separate ethnicity reference data or diagnostic criteria may be worthwhile, at least for certain age groups. However, taking ethnic differences into account has not been found to improve diagnostic performance in previous studies of infants and school-aged children (Beers et al., 2010; Myers et al., 2018a, 2018b; Shahnaz et al., 2013).

5.5.5 Strengths, limitations and directions for future research

The longitudinal design of this study was a strength, but a limitation was that there were no follow up appointments between birth and 6 months of age, where the largest WAI developmental effects occur. Also, the DPOAE reference data used in this study were developed primarily in adults (Gorga et al., 2005), but recent research has shown that these may not be optimal for use in infants (Blankenship et

al., 2018; Hunter et al., 2018). Future research could assess developmental effects on WAI using age-appropriate otoacoustic emissions normative criteria.

5.5.6 Summary and conclusions

There were large developmental effects on measurements of A and Y through infancy. The most dramatic changes occurred between the neonate and 6-month age groups, but there were significant differences between all age groups at certain frequencies. In the 3000 to 4000 Hz region, A, $|Y|_t$, $|Y|_n$, and G_n increased with age. B_n and φ_Y increased with age from 2000 to 3000 Hz and decreased from 4000 to 6000 Hz. There were significant ethnicity effects moderated by age and frequency that were substantial for certain Age × Frequency combinations. For A, we recommend separate normative regions be used for neonates, 6 months, and 12–18 months. For Y_t and Y_n, separate reference data should be used for all age groups (neonates, 6, 12, and 18 months). Further research will be needed to ascertain whether using ethnic-specific normative data leads to improved diagnostic performance.

Chapter 6. Diagnosing Conductive Dysfunction in Infants Using Wideband Acoustic Immittance: Validation and Development of Predictive Models

This chapter develops an ordinal model for diagnosing middle ear dysfunction using wideband absorbance that controls for the effect of age in infants aged 6 to 23 months, and externally validates the neonate model developed in Chapter 2 in a new sample. It has been previously published in the article: Myers, J., Kei, J., Aithal, S., Aithal, V., Driscoll, C., Khan, A., Manuel, A., Joseph, A., Malicka, A. N. (2019a). Diagnosing conductive dysfunction in infants using wideband acoustic immittance: Validation and development of predictive models. *Journal of Speech Language and Hearing Research, 62*(9), 3607-3619.

I made substantive contributions to the article in the areas of study design, data collection, data analysis and drafting of the article, as outlined below:

Contributor	**Statement of contribution**
Joshua Myers (Candidate)	Study design (60%) Recruitment and data collection (60%) Data analysis (100%) Wrote the article (100%)
Joseph Kei	Study design (20%) Edited the article (40%)
Sreedevi Aithal	Study design (5%) Edited the article (15%)
Venkatesh Aithal	Study design (5%)
Carlie Driscoll	Study design (5%) Edited the article (15%)
Asaduzzaman Khan	Study design (5%) Edited the article (15%)
Alehandrea Manuel	Recruitment and data collection (20%)
Anjali Joseph	Recruitment and data collection (20%)
Alicja N. Malicka	Edited the article (15%)

6.1 Abstract

Purpose: To validate the wideband acoustic immittance (WAI) model developed by Myers et al. [*Ear and Hearing*, *39*(6), 1116-1135 (2018)] in a new sample of neonates, and to develop a prediction model for diagnosing middle ear dysfunction in infants aged 6 to 18 months using wideband absorbance.

Methods: Tympanometry, distortion product otoacoustic emissions (DPOAEs) and WAI were measured in 124 neonates, and longitudinally in 357 infants at 6, 12 and 18 months of age. High-frequency (1000-Hz) tympanometry was measured in neonates and at 6 months, and 226-Hz tympanometry at 12 and 18 months. Tympanometry and DPOAEs were used to assess middle ear function of each infant. Results from the neonates were applied to the diagnostic WAI model developed by Myers et al. (2018), and a new prediction model was developed using results from the 6- to 18-month infants. Absorbance was averaged into 1/2 octave bandwidths and 1000, 1414, 2000 and 5657 Hz were used as predictor variables in the model. Results from one ear of infants in each age group (6, 12 and 18 months) were used to develop the model. The amount of bias (overfitting) was estimated with bootstrap resampling, and by applying the model to the opposite ears (the test sample). Performance was assessed using measures of discrimination (*c*-index) and calibration (calibration curves).

Results: The Myers et al. (2018) model had a *c*-index of 0.837 and was accurately calibrated when applied to a new sample of neonates. The model developed for 6- to 18-month infants had satisfactory calibration, and apparent, bias-corrected and test-sample *c*-index of 0.884, 0.867, and 0.887, respectively.

Conclusions: The Myers et al. (2018) model validated well to a new sample of neonates, and the model developed for 6- to 18-month infants was both discriminating and accurately calibrated. The models may be clinically useful, and further research validating, updating and assessing their clinical impact is warranted.

6.2 Introduction

Infants with early onset of otitis media have increased risk of recurrent and persistent infections through childhood (Corbeel, 2007; Damoiseaux et al., 2006; Karma et al., 1989b). Diagnostic tools able to detect middle ear disease at a young age could help to identify affected infants for appropriate treatment (Hunter et al., 2008b; Myers et al., 2018b). Wideband acoustic immittance (WAI) is an innovative, high-resolution test of middle ear function that is suitable for use in infants from birth (Keefe et al., 2000). The term WAI encompasses a family of broadband measures including pressure reflectance (*PR*), energy reflectance (*R*), energy absorbance (*A*), and acoustic admittance (*Y*). *PR* is the ratio of the reflected to forward-propagating acoustic pressure wave amplitude (Keefe & Levi, 1996). *R* is $|PR|^2$, and is the proportion of energy reflected back from the middle ear. *A* is $1 - R$, and represents the proportion of energy absorbed by the middle ear (Rosowski et al., 2013). *Y* is the ratio of volume velocity to acoustic pressure, and is a complex measure with magnitude ($|Y|$) and phase angle (φ_Y) (Keefe & Levi, 1996).

Diagnostic studies have shown that WAI can accurately diagnose conductive dysfunction in neonates (Aithal et al., 2015; Hunter et al., 2010; Keefe et al., 2003a; Myers et al., 2018a; Sanford et al., 2009), and infants (Ellison et al., 2012; Hunter et al., 2008b; Myers et al., 2018b, 2019b; Prieve et al., 2013a). Using WAI clinically typically involves subjective interpretation of visual displays. The high-dimensional, multivariate nature of the response, however, can make interpretation difficult. Development of objective predictive algorithms could assist clinical decision making in diagnostic settings, or be automated for use in screening contexts (Myers et al., 2018a; Sanford & Brockett, 2014; Sanford et al., 2009).

An issue with developing such algorithms for use with infants, however, is that there are substantial maturational effects on WAI that need to be controlled for (Aithal et al., 2014b; Hunter et al., 2010; Hunter et al., 2016; Keefe et al., 1993; Keefe et al., 2000; Kei et al., 2013; Myers et al., 2018a, 2019c; Sanford et al., 2009; Shahnaz et al., 2014; Werner et al., 2010). One solution to this problem has been to create age-specific algorithms, such as models that have been developed for neonates (Keefe et al., 2003a; Myers et al., 2018a; Sanford et al., 2009) or different age groups through infancy (Myers et al., 2018b, 2019b). Alternatively, models could be developed with age included as an interaction term with WAI, which would allow interpretation to vary with age (Myers et al., 2019c). Another approach would be to only include frequency regions that are relatively unaffected by age as predictors in a model (Myers et al., 2019c; Sanford & Feeney, 2008). For example, for infants aged 0 to 6 months, Shahnaz et al. (2014) found limited developmental effects on *R* from 600 to 1600 Hz, and Sanford and

Feeney (2008) recommended *R* from 800 to 2000 Hz as a developmentally stable region. For 6- to 18-month infants, Myers et al. (2019c) found significant age effects on *A* from 3000 to 5000 Hz, so a model developed for this age group might include an interaction with age at these frequencies, or exclude them as predictor variables.

Another issue in diagnostic WAI research is that although many predictive models have been developed (Ellison et al., 2012; Keefe et al., 2003a; Keefe et al., 2003b; Myers et al., 2018a, 2018b, 2019b; Piskorski et al., 1999; Sanford et al., 2009), none have been validated in an external sample of subjects. Some studies have estimated the amount of bias (overfitting) in a model with bootstrap resampling, or by developing the model on one ear of subjects and applying it to the opposite ears (Keefe et al., 2003a; Keefe et al., 2003b; Myers et al., 2018a, 2018b, 2019b). Validating a model in a new sample, however, is recommended prior to clinical implementation, to provide a realistic idea of model performance, since results are usually poorer when applied to new subjects (Moons et al., 2012a). This can be due to overfitting, or differences in subject characteristics, environment, or equipment used in the new setting (Myers et al., 2018a; Steyerberg, 2008).

Myers et al. (2018a) developed a model for diagnosing conductive dysfunction in neonates using octave-averaged *A* at 1000 and 2000 Hz, $|Y|$ at 1000 and 2000 Hz, and φ_Y at 1000 and 4000 Hz. The model was developed using results from one ear of subjects, and assessed for bias using the opposite ears, and with bootstrap resampling. Myers et al. (2018b, 2019) developed models using *A* specifically for 6-, and 12-month-old infants, respectively, but it may be useful to have a model that can be used with a broader age range of infants. There were two aims for this chapter: Study 1 aimed to validate the Myers et al. (2018a) model in a new sample of neonates, and Study 2 aimed to develop a predictive model for infants aged approximately 6 to 18 months using *A*, controlling for the effect of age.

6.3 Methods

Seven-hundred and fifty-three infants were recruited at birth and followed through infancy (Myers et al., 2018a, 2018b, 2019b, 2019c). Myers et al. (2018a) developed a diagnostic model using 629 neonates, and the present study applied that model to the remaining 124 subjects. This sample is referred to as the "validation sample" in this report. Characteristics of the validation sample are presented in Table 6.1. Infants were scheduled to be followed up at around 6, 12, and 18 months of age. Of the 753 infants recruited to the study, 357 attended at least one follow up appointment. Two-hundred and seventy-one attended at 6 months, 202 at 12 months, and 126 at 18 months of age. Data collected at the follow-up appointments is referred to as the "development sample" in this study. Characteristics of infants in the

development sample are provided in Table 6.2, and ages of infants in the validation and development samples are presented in Table 6.3.

Table 6.1. Characteristics of the 124 neonates in the validation sample

Characteristic	Value
Gender (count)	1 missing
Female (%)	55 (44)
Male (%)	68 (55)
Ethnicity (count)	0 missing
Caucasian (%)	102 (82)
Asian (%)	12 (10)
Oceanian (%)	7 (6)
African (%)	3 (2)
Gestational age (weeks)	1 missing
Median (IQR)	39 (38 to 39)
Range	36 to 41
Birth type (count)	0 missing
Vaginal (%)	71 (57)
C-section (%)	53 (43)
Birth weight (kg)	2 missing
Median (IQR)	3.4 (3.1 to 3.7)
Range	2 to 4.7

IQR, interquartile range.

Table 6.2. Characteristics of the 357 infants in the development sample

Characteristic	Value
Gender (count)	0 missing
Female (%)	177 (50)
Male (%)	180 (50)
Ethnicity (count)	3 missing
Caucasian (%)	296 (83)
Asian (%)	39 (11)
Oceanian (%)	12 (3)
South American (%)	3 (1)
African (%)	4 (1)

All infants either passed automated auditory brainstem response (ABR) neonatal hearing screening, or had a finding of normal hearing sensitivity upon diagnostic hearing evaluation. Diagnostic audiology defined normal hearing as passing a click-evoked ABR test, and also either tone-burst ABR (1000 and 4000 Hz), or transient evoked otoacoustic emissions.

Table 6.3. Ages of infants in the validation and development samples

	Median	IQR	Range	Missing
Neonates (hrs)	43	27－49	5－81	1
6 months (wks)	28	26－30	23－39	23
12 months (wks)	54	53－57	43－70	3
18 months (wks)	80	79－82	75－100	4

The missing column gives the number of subjects in that age group with missing age data. IQR, interquartile range.

6.3.1 Test procedure

Neonates were tested in the Maternity Ward at the Townsville Hospital, and other age groups in a quiet office at the local paediatric community health centre. The test procedure for the various age groups is described in Myers et al. (2019c). Briefly, tests were performed using Interacoustics Titan devices. Distortion product otoacoustic emissions (DPOAEs), tympanometry and WAI were measured in both ears of an infant if possible. Otoscopy was also performed at 6, 12 and 18 months to ensure that the ear canal was not occluded by wax. DPOAEs were measured for primary tone pairs f_1 (65 dB SPL) and f_2 (55 dB SPL) for f_2 of 2000, 3000, 4000 and 6000 Hz with an f_2 / f_1 ratio of 1.22. Emissions were

classified as present if the signal-to-noise ratio was ≥6 dB and DPOAE levels ≥−10 dB SPL for at least three f_2 frequencies (Gorga et al., 2005).

High-frequency (1000-Hz) tympanometry (HFT) was used for neonates and 6-month infants, and low-frequency (226-Hz) tympanometry (LFT) for the 12- and 18-month groups. For neonates, HFT results were classified according to the criteria described in Myers et al. (2018a). A line was drawn between the positive and negative extremes of a tympanogram, and ears were classified as "peaked" (pass) if a peak extended above the line, otherwise "not peaked" (fail). For the 6-month infants, HFT results were classified in a similar way (drawing a line), but with a third category of "negative peak" for peaked traces with tympanometric peak pressure <−150 daPa. LFT results for the 12- and 18-month groups were classified as "type A" if there was a peak |Y| ≥0.3 mmho between −150 to 50 daPa, "type C" if the peak (≥0.3 mmho) occurred at tympanic peak pressure <−150 daPa, and otherwise "type B" (Myers et al., 2019b; Roush et al., 1995). HFT results for the 6-month group (peaked, negative peak and not peaked) are also referred to as type A, type C and type B, respectively in this report since data from the 6-, 12-, and 18-month groups were analysed together.

A, $|Y|$ and φ_Y were measured at ambient pressure at 1/24 octave frequency resolution. Thirty-two broadband clicks were delivered at 96 dB peSPL and results averaged after removal of artefacts as described by (Liu et al., 2008). A visual display of A was monitored during testing to check for air leaks. The test was stopped and the probe reinserted if A ≥0.7 below 500 Hz for neonates or ≥0.3 below 300 Hz for the other age groups (Aithal et al., 2015; Groon et al., 2015; Keefe et al., 2000).

6.3.2 Reference standards

The validation study (Study 1) used the binary (pass, fail) reference standard described in Myers et al. (2018a). Ears were classified as "pass" if they passed both tests (present DPOAEs and peaked HFT), otherwise "fail".

The development study (Study 2) used the ordinal reference standard described by Myers et al. (2019b) which classified ears as either "normal", or having "mild" or "severe" middle ear dysfunction. Ears were classified as normal if they passed both reference tests (present DPOAEs with type A tympanograms), and severe if they failed both tests (absent DPOAEs with type B tympanograms). The mild category consisted of ears with type C tympanograms (with present/absent DPOAEs), and ears with either present DPOAEs and type B tympanograms, or absent DPOAEs and type A tympanograms. The mild category was created because research has shown an ordinal association in A for infants and children from normal, to negative middle ear pressure/partial middle ear effusion, to severe dysfunction due to middle ear effusion (Beers et al., 2010; Ellison et al., 2012; Hunter et al., 2008b; Myers et al.,

2018b, 2019b; Shaver & Sun, 2013). Therefore, ears with negative middle ear pressure were placed in the mild group (Beers et al., 2010; Hunter et al., 2008b), and also ears that failed only one of the tests (type B or absent DPOAEs), as previous research has found that these ears are likely to have mild dysfunction also (Myers et al., 2018a, 2018b, 2019b).

6.3.3 Statistical modelling

For the validation study (Study 1), WAI results were averaged into octave bandwidths to be applied to the model developed by Myers et al. (2018a). The predictor variables in the model were A at 1000 and 2000 Hz, $|Y|$ at 1000 and 2000 Hz, and φ_Y at 1000 and 4000 Hz.

The development study (Study 2) used A only to develop the model, at 1/2 frequency resolution, in order to limit the number of potential predictors (Myers et al., 2018b). The probability of middle ear dysfunction was modelled using proportional odds ordinal logistic regression. Candidate predictor variables for modelling were A at 1000, 1414, 2000, 2828, 4000, and 5657 Hz, as previous studies have found an ordinal association between A and middle ear dysfunction in infants and children over this frequency range (Beers et al., 2010; Ellison et al., 2012; Hunter et al., 2008b; Myers et al., 2018b, 2019b). Proportional odds logistic regression assumes that the same coefficients can be used to predict each level of the reference standard (mild and severe dysfunction). This assumption was tested by plotting the mean for A variables for each level of the reference standard, with and without assuming proportional odds (Harrell, 2015).

One ear of infants from each test session (6, 12 and 18 months) was randomly selected to develop the model (the "training sample"). The model was assessed for bias (overfitting) with bootstrap resampling, and by applying it to the opposite ears (the "test sample") (Myers et al., 2018a). Infants with results from only one ear were put into the training sample to maximize the size of this group. Bootstrap resampling involved sampling with replacement from the training sample a sample of the same size (a "bootstrapped sample"). A model was fitted to the bootstrapped sample, and then applied to the training sample. This process was repeated 500 times and the difference in performance measures calculated and averaged to estimate the amount of bias in the model (Steyerberg et al., 2001c). Huber-White robust covariance matrix estimates were used to account for correlations in the training sample, since infants potentially had multiple records (i.e., results from 6-, 12- and 18-months) (Harrell, 2015; Myers et al., 2019b). The Huber-White method assumed that results from an infant were independent, rather than results from an individual observation (Hardin, 2005).

Linear models were fitted (with predictor variables assumed to have a linear association with the outcome), and also nonlinear models, with the assumption of linearity relaxed (Harrell, 2015). Nonlin-

earity was implemented with 5-knot restricted cubic splines using the method described by Myers et al. (2018a). Models were also fitted controlling for developmental effects by either including Age (in weeks) as an interaction term with *A* variables, or only using developmentally stable predictors (1000, 1414, 2000 and 5657 Hz) (Myers et al., 2019c). The shrinkage coefficient (γ) was used to evaluate whether a model was too complex, and likely to be overfitting:

$$\gamma = \frac{\text{model } \chi^2 - df}{\text{model } \chi^2}, \tag{6.1}$$

where model χ^2 is the likelihood ratio χ^2 statistic, and *df* the total degrees of freedom from all predictors in the model. A γ of >0.9 was considered acceptable, indicating that model performance in a new sample would not likely be more than 10% worse (Harrell, 2015; Myers et al., 2018a). Akaike's information criterion (AIC), a measure of model fit with a penalty for complexity, was used to select the best-fitting model for further evaluation of performance (Burnham & Anderson, 2002).

Performance was assessed with measures of discrimination and calibration. Discrimination was evaluated using the Somers' D statistic, which was converted to the *c*-index (a.k.a area under the receiver operating characteristic curve) for ease of interpretation: *c*-index = 0.5(Somers' D + 1). Calibration was assessed with calibration curves which plotted actual against predicted probabilities. Calibration evaluates the quality of predictions, which are accurate if they align closely with the frequency of the condition. For example, for infants with predicted probability of 0.3, approximately 3/10 should actually have middle ear dysfunction (Myers et al., 2018a; Steyerberg et al., 2010).

6.3.4 Missing data

The number of ears missing either tympanometry, DPOAE or WAI data in the validation and development samples is provided in Table 6.4. Results were missing due to an infant crying or not tolerating the probe in her ear. Observations with missing data for any of these tests were removed prior to modelling. Infants missing age data in the development sample (see Table 6.3) were imputed with the median age for their age group to avoid having to remove observations with complete reference test and WAI data, missing only age.

Table 6.4. Number of ears with missing data

	Neonates	6 months	12 months	18 months
Tympanometry	1	18	50	37
DPOAEs	0	22	50	40
WAI	10	27	55	39

Number of ears with missing data for tympanometry, DPOAEs and WAI in the validation (neonates) and development (6, 12, 18 months) samples. DPOAEs, distortion product otoacoustic emissions; WAI, wideband acoustic immittance.

The study was approved by the Townsville Health Service District Institutional Ethics Committee, and the University of Queensland Behavioural and Social Science Ethical Review Committee. Analyses was performed using R (R Core Team, 2017) expanded with the *rms* package for regression modelling (Harrell, 2016). This report has been written according to the recommendations of the TRIPOD (transparent reporting of a multivariate prediction model for individual prognosis or diagnosis) statement for the reporting of clinical prediction models (Collins et al., 2015). The data and code for the analyses are available online (https://github.com/Josh-Myers/WAI-All-Ages).

6.4 Results

The validation sample for Study 2 had results from 238 ears after removal of observations with missing HFT, DPOAE or WAI data. Results from both ears of neonates were applied to the model, since there were no assumptions about independence. The development sample had results from 1038 ears after removal of missing data: 508 at 6 months, 330 at 12 months, and 200 at 18 months. The development sample was split into a training sample of 536, and a test sample of 502 observations (the opposite ears). Table 6.5 shows the reference test (tympanometry and DPOAEs) and reference standard results for the validation and development samples after removal of missing data.

Table 6.5. Reference test results for the validation and development samples

Age group	Test	Results
Neonate	Tympanometry (pass, fail)	207, 31
	DPOAEs (pass, fail)	188, 50
	RS (pass, fail)	177, 61
6 months	Tympanometry (type A, C, B)	399, 29, 80
	DPOAEs (pass, fail)	427, 81
	RS (normal, mild, severe)	380, 68, 60
12 months	Tympanometry (type A, C, B)	231, 15, 84
	DPOAEs (pass, fail)	253, 77
	RS (normal, mild, severe)	221, 48, 61
18 months	Tympanometry (type A, C, B)	146, 12, 42
	DPOAEs (pass, fail)	166, 34
	RS (normal, mild, severe)	141, 30, 29

Reference test and reference standard (RS) results for the validation (neonate) and development (6, 12, and 18 months) samples after removal of observations with missing tympanometry, DPOAE or WAI data. DPOAEs, distortion product otoacoustic emissions.

6.4.1 Study 1: The validation study

Mean A, $|Y|$ and φ_Y for the validation sample are compared with the results from Myers et al. (2018a) in Figure 6.1. Mean WAI of the pass group in the validation sample was slightly lower than Myers et al. at frequencies above 2000 Hz for A, and above 1000 Hz for Y, but overall, was similar for both groups between the samples. For both the Myers et al. and validation samples, mean A for the pass group was greater than the fail group across the entire frequency range. Mean $|Y|$ for pass was greater than fail from 226 to 2800 Hz, and mean φ_Y for pass lower than fail from 1200 to 8000 Hz. Median age of subjects in the validation sample was 43 hours (Table 6.3) which was similar to the median age of 42 hours in Myers et al. Appling the validation sample to the Myers et al. model, resulted in a c-index of 0.837 (95% CI 0.773－0.901). The calibration curve is presented in Figure 6.2 which shows predictions were very close to the ideal calibration line across the entire range of probabilities.

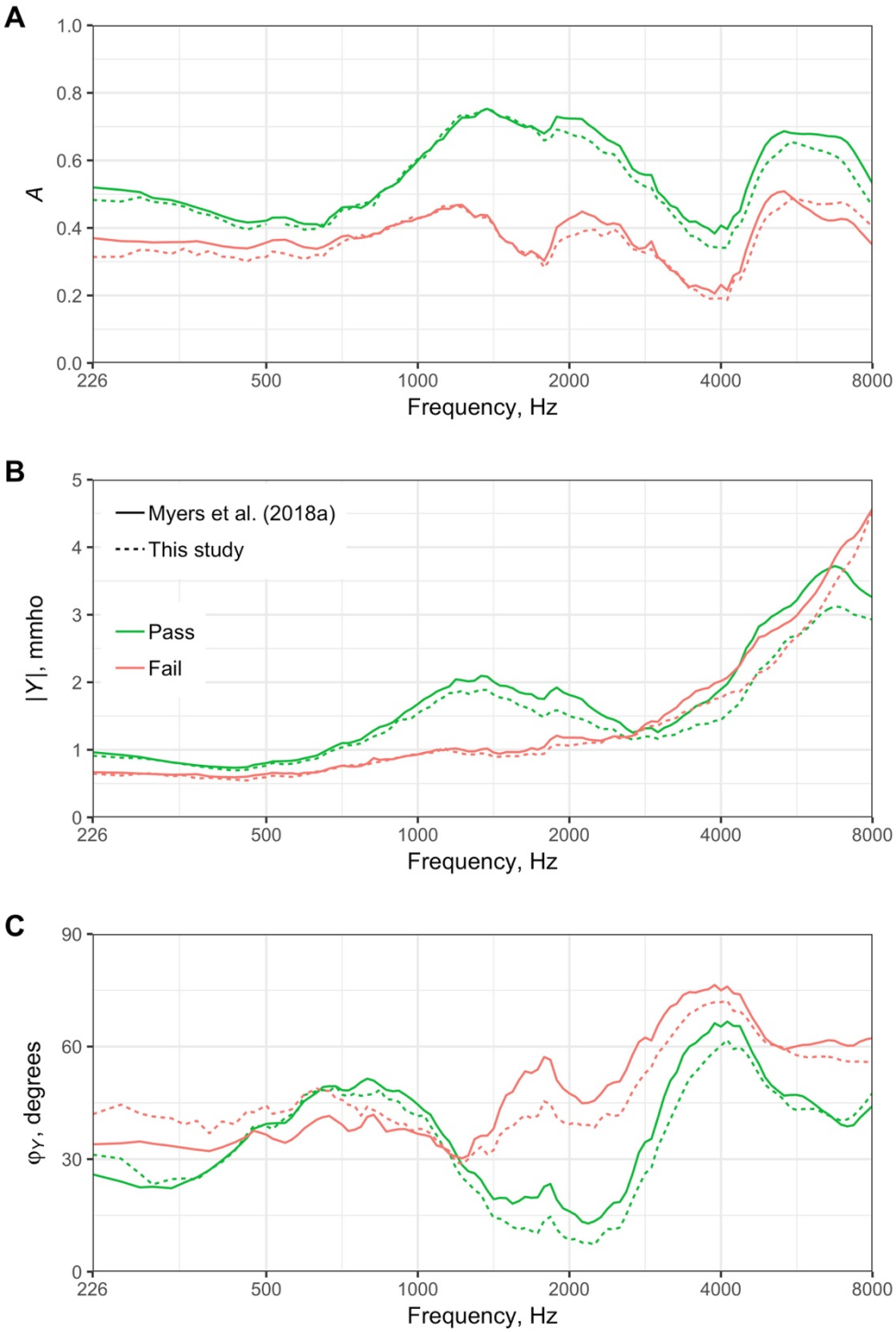


Figure 6-1. Mean WAI compared with results from Myers et al. (2018a).
Mean A, $|Y|$ and φ_Y (A, B and C, respectively) of the validation sample compared with results from Myers et al. (2018a) for neonates that passed and failed the reference standard.

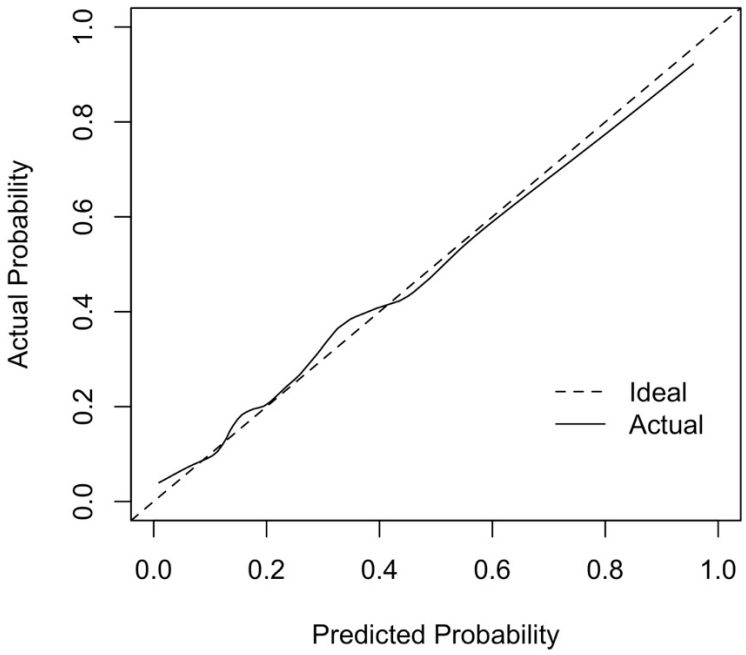


Figure 6-2. Calibration curve for the validation study.
The calibration curve (solid line) from applying the validation sample to the Myers et al. (2018a) model, plotting actual against predicted probability. The dashed line depicts ideal calibration.

6.4.2 Study 2: The model development study

Mean *A* in the development sample for the normal, mild and severe middle ear dysfunction groups, stratified by age group (6, 12 and 18 months), is depicted in Figure 6.3A. There was a clear age effect from around 3000 to 5000 Hz with *A* generally increasing with age for the normal and mild groups, and decreasing with age for the severe middle ear dysfunction group. There appears to be an ordinal association in mean *A* between the groups (normal > mild > severe) above 1000 Hz for 6 months, above 500 Hz for 12 months, and up to 7000 Hz for 18 months. Median *A* stratified by all possible combinations of reference test results is shown in Figure 6.3B. Median *A* for all subgroups of the mild group generally fell between results for the normal (pass both tests), and severe middle ear dysfunction (fail both tests) groups. An exception was that ears that ears that passed DPOAEs with type C tympanograms had higher *A* than ears that passed both tests in the 3000 Hz region. Ears with type A tympanograms that failed DPOAEs had lower *A* compared to normal ears above 1200 Hz, indicating that these were likely to have failed DPOAEs due to conductive, rather than sensory dysfunction. Ears with type A tympanograms and failed DPOAEs had higher *A* than type C and failed DPOAEs from 1000 to 3000 Hz, but were similar from 3000 to 6000 Hz. Ears with type C tympanograms and failed

DPOAEs were similar to type C and pass DPOAEs from 4000 to 6000 Hz, but had lower *A* at around 1500 and 3000 Hz.

To test the assumption of proportional odds, mean *A* of potential predictor variables (1000 to 5657 Hz) for each level of the reference standard is depicted in Figure 6.4, with and without assuming proportional odds. The solid lines connected by circles show the simple stratified means, and the dashed lines the expected values under the assumption of proportional odds. The trend in the solid lines should be monotonic to satisfy the assumption of proportional odds. The stratified means were very close to expected values for 1000, 1400 and 2000 Hz, and were also fairly close for 4000 and 5657 Hz. For 2828 Hz, the stratified mean was further away from expected values, but the trend was still monotonic, indicating that the assumption of proportional odds was satisfied.

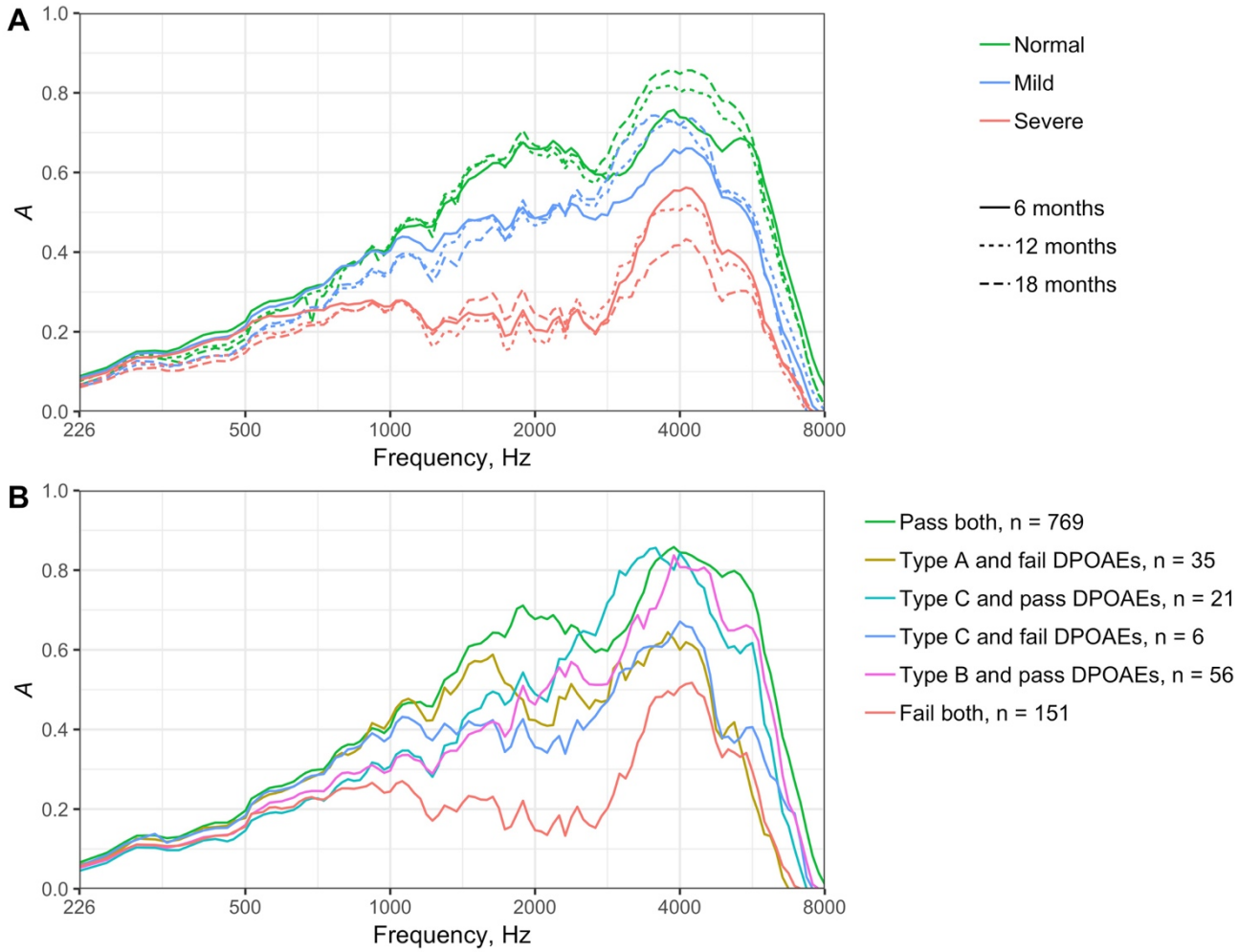


Figure 6-3. Mean absorbance stratified by reference test results.
Mean absorbance (*A*) stratified by both age group and levels of the reference standard (A), and median *A* grouped by all possible combinations of reference test results (B). The mean was chosen for plot A because proportional odds logistic regression models the mean, and the median was chosen for plot B because of the small size of some groups.

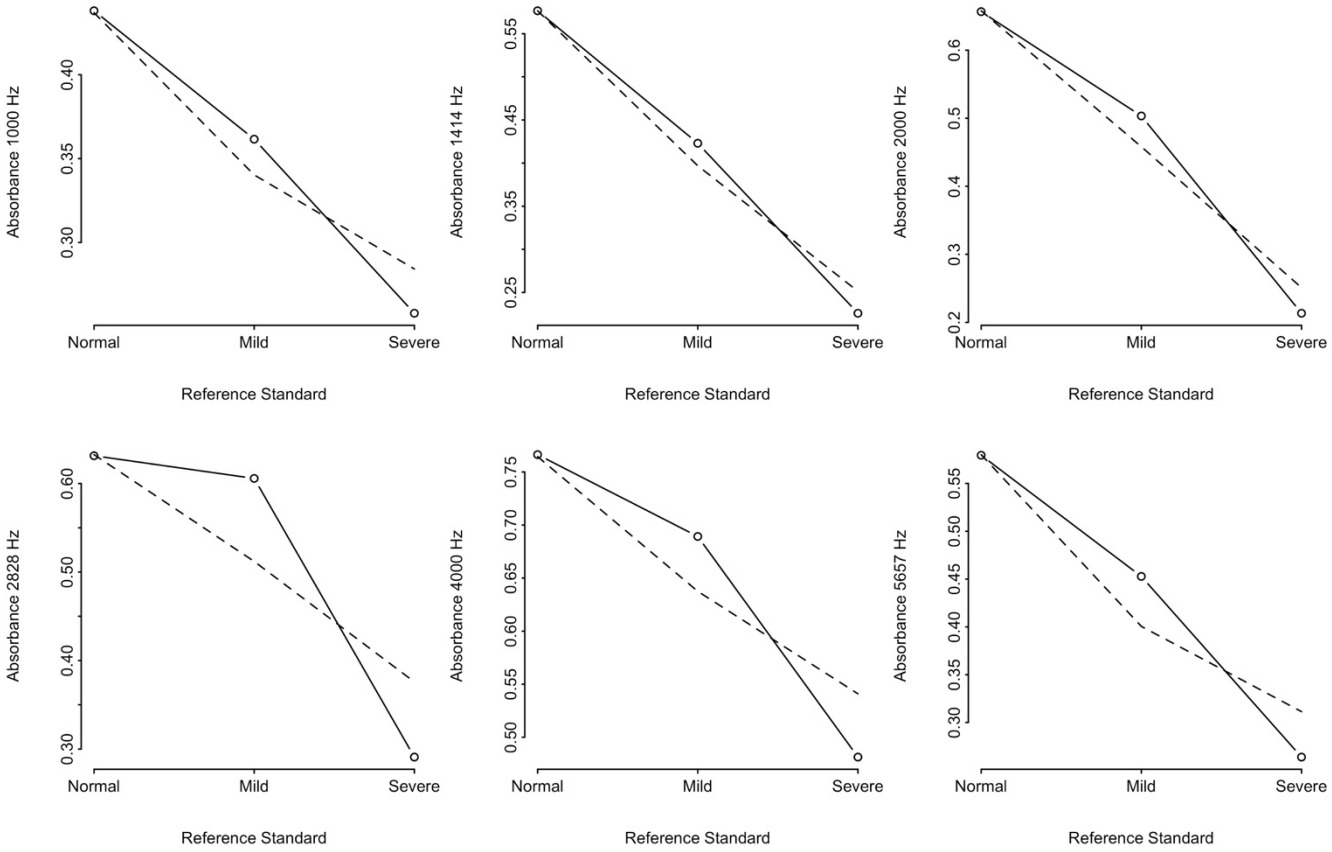


Figure 6-4. Assessing the assumption of proportional odds.
Plots assessing the assumption of proportional odds for candidate *A* predictors in the development study. The circles connected by solid lines show the simple stratified means, and the dashed lines the expected values under the assumption of proportional odds. A monotonic trend in the simple stratified means indicates that the assumption of proportional odds has been met.

The modelling process began by fitting a simple baseline model, Model A, that included all candidate *A* predictors (1000 to 5657 Hz), with predictors assumed to have a linear association with the outcome. Next, Model B used the same predictors, but with the assumption of linearity relaxed using restricted cubic splines. Models C and D included Age as an interaction term with *A* variables: Model C assumed linearity, and Model D allowed nonlinearity. Model E was similar to Model D, but Age interactions were only included for frequencies with the largest developmental effects (2828 and 5657 Hz) (Myers et al., 2019c). Finally, Model F was similar to Model B (nonlinear with no Age interaction), but the predictor variables most affected by age were not included in the model (2828 and 4000 Hz).

LR χ^2, γ and AIC for the models are presented in Table 6.6. Model B had lower AIC than Model A, indicating that allowing nonlinearity improved model fit. Model C also had lower AIC than Model

A, indicating that accounting for age also improved model fit. Model D, the nonlinear model with Age interacting with all *A* variables, had γ of 0.88, indicating that it may be too complex for this dataset (γ < 0.9). Model F had the lowest AIC, indicating that it was the best fitting model, and was therefore taken as the final model for further investigation and evaluation.

Table 6.6. Statistics for the development study models

Model	Predictor Variables	LR χ^2	*df*	γ	AIC
Model A	*A* from 1000 to 5657 Hz	290.11	6	0.98	582.26
Model B	Model A nonlinear	350.42	24	0.93	557.95
Model C	Model A × Age (all variables)	310.06	13	0.96	576.31
Model D	Model B × Age (all variables)	393.41	49	0.88	564.96
Model E	Model B × Age (2828 + 4000 Hz)	371.66	33	0.91	554.71
Model F	Model B – 2828 and 4000 Hz	341.45	16	0.95	550.92

LR χ^2, *df*, γ, and AIC for the development study models. The *p*-values for the LR χ^2 statistics were <0.0001 for all models. The absorbance predictor variables were 1000, 1414, 2000, 2828, 4000, 5657 Hz at 1/2 octave frequency resolution. AIC, Akaike's information criterion; LR χ^2, likelihood ratio chi-squared statistic; γ, shrinkage coefficient; *df*, degrees of freedom.

The importance of predictors in Model F is shown Table 6.7. *A* at 5657 Hz had the highest LR χ^2 (47.19), followed by 1000 Hz (30.08), and then 2000 Hz (18.77). *A* at 1414 Hz was not significant in the model ($p > 0.05$), but it was retained, since it may still be contributing, and removing variables from a model based on the results of statistical tests is not good practice (Gelman & Hill, 2007). Nonlinear contributions were statistically significant, and accounted for a substantial amount of LR χ^2 for the top three predictors, indicating that introducing nonlinearity was important in the model. The *c*-index for Model F was 0.884, and 0.867 after being corrected for bias with bootstrap resampling. Applying the model to the opposite ears resulted in a *c*-index of 0.887. Figure 6.5 shows the calibration curves for Model F for ≥mild and ≥severe middle ear dysfunction (left and right columns, respectively) calculated for the training and test samples (top and bottom rows, respectively). The curves are overall satisfactory, being relatively close to the ideal calibration line for all samples.

Table 6.7. Statistical analysis for the final model

Variable	LR χ^2	*df*	*p*-value
1000 Hz	30.08	4	<0.001
Nonlinear	27.71	3	<0.001
1414 Hz	6.82	4	0.146
Nonlinear	2.94	3	0.401
2000 Hz	18.77	4	<0.001
Nonlinear	8.57	3	0.036
5657 Hz	47.19	4	<0.001
Nonlinear	9.05	3	0.029
TOTAL NONLINEAR	65.27	12	<0.001
TOTAL	173.97	16	<0.001

Statistical analysis for the final model (Model F), showing the LR χ^2 statistics, associated *p*-values, and degrees of freedom (*df*) for each predictor variable (total and nonlinear contributions) in the model. LR χ^2, likelihood ratio chi-squared statistic.

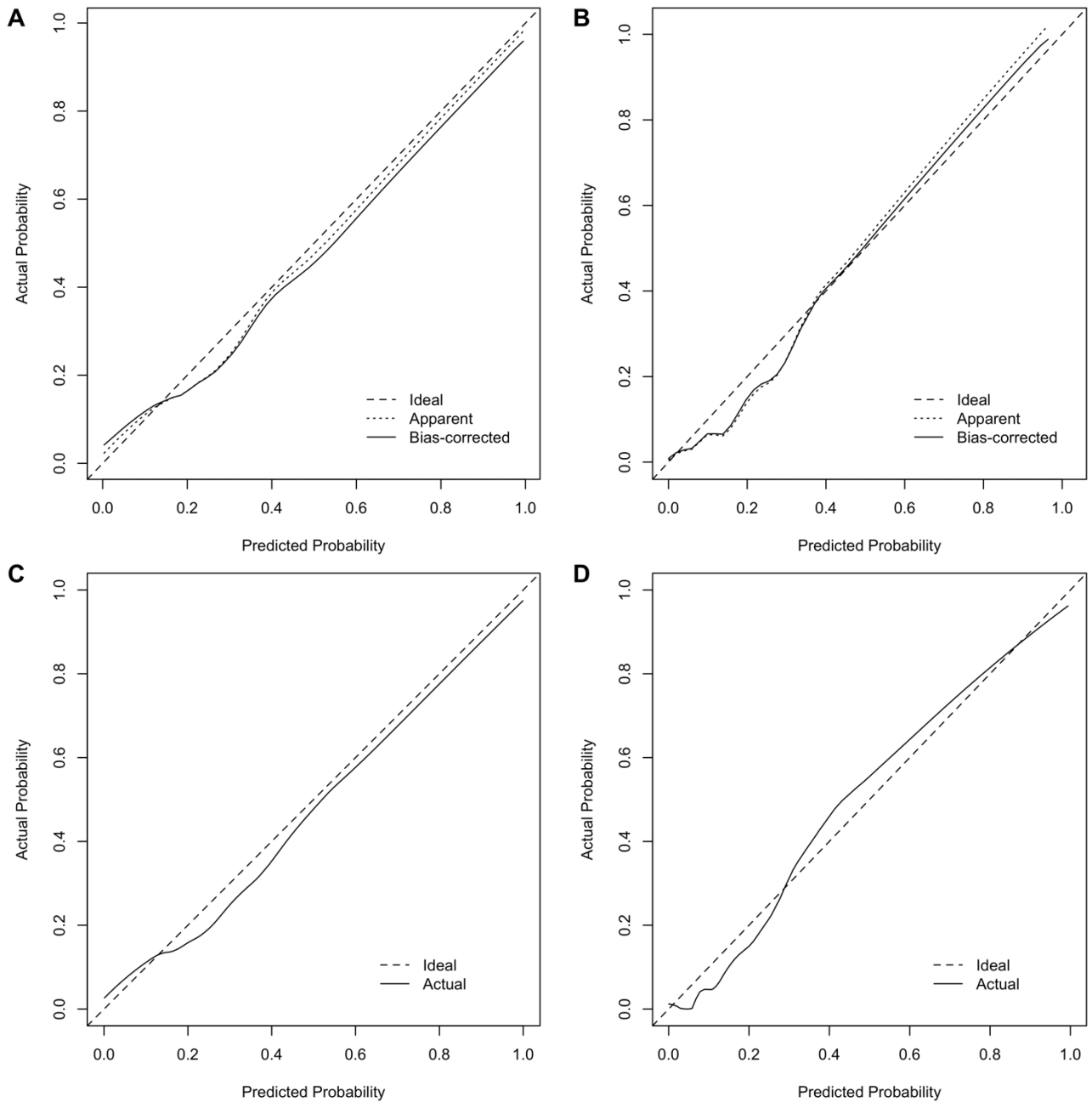


Figure 6-5. Calibration curves for the final model.
Calibration curves for Model F plotting actual versus predicted probability for ≥ mild (left column, A and C) and ≥ severe middle ear dysfunction (right column, B and D). The top row (A and B) depicts the curves for the training (apparent) and bias-corrected (bootstrapped) samples, and the bottom row (C and D), the curves for the test sample (the opposite ears). The dashed lines depict ideal calibration.

The equation for Model F to calculate probabilities is provided in Section 6.6: Appendix. An online application implementing the model is available (http://joshmyers.shinyapps.io/WAIPredictions/) that can make predictions either by uploading a session exported from a Titan device, or by manually entering *A* results.

6.5 Discussion

6.5.1 Study 1: The validation study

WAI results from the validation sample compared well results from Myers et al. (2018a) (Figure 6.1), and the model validated well to the new sample. The calibration curve was very close to the ideal calibration line (Figure 6.2). The *c*-index of 0.837 was lower than the apparent performance of 0.876 reported by Myers et al., but was close to the bias-corrected estimate of 0.845.

6.5.2 Study 2: The model development study

Model F (the best fitting model) may be useful in both screening and diagnostic settings. Probabilities for one or both levels of the reference standard (mild and severe) could be calculated as needed, depending on the context. The cut off for further action would be the point where there is enough concern to warrant further action (Myers et al., 2018a). For example, a screening program wishing to test for severe middle ear dysfunction may only calculate the probability of ≥severe middle ear dysfunction, failing infants with probability >0.5, since middle ear dysfunction is more likely than not at this point. In a diagnostic setting probabilities for both ≥mild and ≥severe dysfunction could be presented along with visual depictions of WAI to aid clinical decision making. Figure 6.6 shows examples of applying the model to two cases, showing *A* results along with the predicted probabilities from the model. In the first example (A) the probability of both ≥mild and ≥severe middle ear dysfunction is high (0.99 and 0.95, respectively), indicating that severe middle ear dysfunction is likely. For the second example (B) mild middle ear dysfunction is likely, since the probability of ≥mild is high (0.79), but ≥severe is not (0.36).

Performance measures for Model F showed accurate discrimination and calibration, and results from bootstrap resampling and the opposite ears indicated that the model was not overfitting, and therefore may perform well when applied to new samples. Accounting for developmental effects and allowing nonlinearity both improved model fit, with Model F using both of these strategies. The *c*-index for Model F of 0.884 (0.867 after being corrected for bias) was not as high as the 0.93 reported by Ellison et al. (2012), but was within the margin of error reported in that study. Results for Study 2 were similar to Beers et al. (2010) with absorbance normal > mild > severe from 1000 to 6000 Hz. Results were also comparable to Hunter et al. (2008b), who found that ears with negative middle ear pressure fell between normal and severe conditions from 1000 to 4000 Hz, but in the present study this trend continued up to 6000 Hz. Factors contributing to this difference could include age differences between the studies, as Hunter, Tubaugh, et al. included subjects from 0 to 47 months, whereas infants

in Study 2 were 6 to 18 months old. Also, Hunter, Tubaugh, et al. used the MEPA (middle ear power analyser) equipment, whereas we recorded absorbance using the Interacoustics Titan WBT (wideband tympanometry) system.

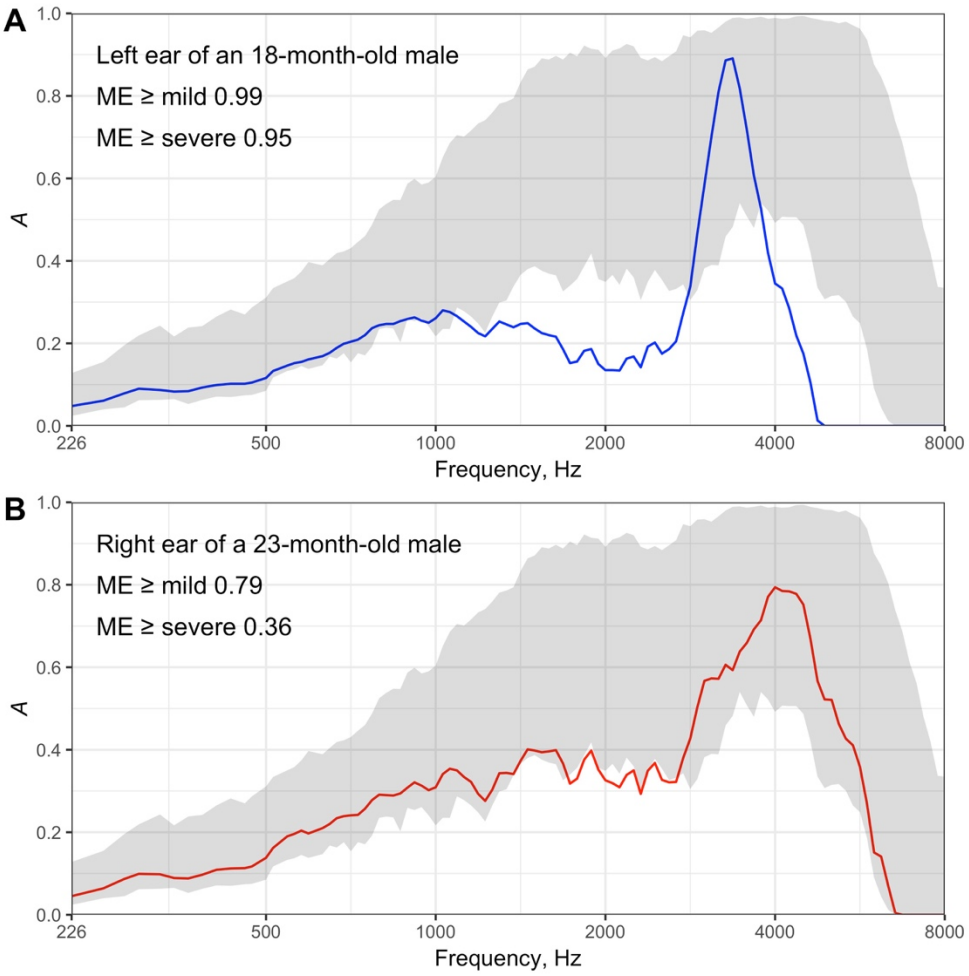


Figure 6-6. Applying the model to individual cases.
Examples of applying the model to 2 cases, showing absorbance (*A*), and corresponding probabilities predicted by the model. The shaded region depicts the 90% range of normal ears from the 18-month group.

6.5.3 Strengths, limitations, and directions for future research

This study was the first to validate a WAI model in a new sample of neonates, and the model developed for 6- to 18-month infants took care to avoid overfitting, and controlled for the effect of age. A limitation of the study, however, was that data collection was not blinded. The same researcher measured both WAI and reference tests, which may have introduced bias. Also, the validation sample was very similar to the sample used by Myers et al. (2018a) to develop the model. The model performed well when applied to the new data, but subjects were of similar age, and were tested by the same researchers using the same equipment in the same environment. Future research could validate the model in a sample tested in a different environment, updating the model if necessary. Another limitation was that

the reference standard used to develop the infant model was not the gold standard for diagnosing middle ear disease in infants. Future research using reference tests such as otomicroscopy and surgical confirmation for middle ear status could create a more stringent reference standard with well-defined pathologies for subgroups, for example, Eustachian tube dysfunction, partial, and complete middle ear effusion.

6.5.4 Summary and conclusions

The model developed by Myers et al. (2018a) validated well to a new sample of neonates and a new model was developed for 6- to 18-month infants that controlled for maturational effects on *A*. The validated and developed models may be clinically useful, and further research validating, updating and assessing their clinical impact in screening and diagnostic settings is warranted.

6.6 Appendix: The equation for Model F

The ordinal logistic regression equation for Model F, the final model in Chapter 6, using absorbance (*A*) at 1000, 1414, 2000, and 5657 Hz expanded with 5-knot restricted cubic splines to calculate the probability (P) of middle ear (ME) dysfunction for levels of the reference standard (*j*) is:

$$\text{Prob}\{\text{ME} \geq j\} = \frac{1}{1+\exp(-\alpha_j - X\beta)}, \quad \text{where}$$

$$\begin{aligned} \hat{\alpha}_{\text{Mild}} &= 4.091712 \\ \hat{\alpha}_{\text{Severe}} &= 2.171674 \end{aligned}$$

$$\begin{aligned} X\hat{\beta} = \\ &+11.64024A_{1000\text{ Hz}} - 275.3909(A_{1000\text{ Hz}} - 0.1921042)_+^3 \\ &+500.0415(A_{1000\text{ Hz}} - 0.3086042)_+^3 + 36.93358(A_{1000\text{ Hz}} - 0.3941667)_+^3 \\ &-309.9347(A_{1000\text{ Hz}} - 0.4775937)_+^3 + 48.35041(A_{1000\text{ Hz}} - 0.6629458)_+^3 \\ &-8.732798A_{1414\text{ Hz}} + 21.97317(A_{1414\text{ Hz}} - 0.1545625)_+^3 \\ &-45.78994(A_{1414\text{ Hz}} - 0.4022917)_+^3 - 10.02401(A_{1414\text{ Hz}} - 0.5262083)_+^3 \\ &+40.44765(A_{1414\text{ Hz}} - 0.6351979)_+^3 - 6.60688(A_{1414\text{ Hz}} - 0.81625)_+^3 \\ &-4.814025A_{2000\text{ Hz}} + 1.510417(A_{2000\text{ Hz}} - 0.10325)_+^3 \\ &-4.627762(A_{2000\text{ Hz}} - 0.4595312)_+^3 + 212.5867(A_{2000\text{ Hz}} - 0.6344583)_+^3 \\ &-361.1814(A_{2000\text{ Hz}} - 0.7311146)_+^3 + 151.7121(A_{2000\text{ Hz}} - 0.8645208)_+^3 \\ &-5.744574A_{5657\text{ Hz}} + 9.477814(A_{5657\text{ Hz}} + 0.02235417)_+^3 \\ &-54.01727(A_{5657\text{ Hz}} - 0.3092812)_+^3 + 215.402(A_{5657\text{ Hz}} - 0.5740833)_+^3 \\ &-324.2183(A_{5657\text{ Hz}} - 0.7481354)_+^3 + 153.3558(A_{5657\text{ Hz}} - 0.8856458)_+^3 \end{aligned}$$

and $(x)_+ = x$ if $x > 0$, 0 otherwise

Note that there are 6 coefficients for each *A* frequency in the model due to being modelled with 5-knot restricted cubic splines. The subscript "+" symbols after a term, signify that if the value of the term is positive, it is cubed, or otherwise it is set to 0. This ensures that the splines meet at the join. Cubing the terms makes the joins smooth.

Chapter 7. General Discussion

7.1 Revisiting the rationale for the study

Early intervention for otitis media in infancy is vital, because onset of the disease at a young age increases risk of recurrent and chronic infections that can impact development. However, early identification of otitis media in infants remains a challenge, as the condition is often asymptomatic, and currently available clinical tools can be inaccurate and difficult to use. Wideband acoustic immittance (WAI) is an innovative, high-resolution middle ear test that is quick and easy to administer, and preliminary studies in neonates have shown high diagnostic accuracy (Aithal et al., 2015; Hunter et al., 2010; Keefe et al., 2003a; Keefe et al., 2003b; Sanford et al., 2009). However, the need has been identified for additional research using a stringent reference standard in a large cohort of neonates. Furthermore, little is known about the diagnostic performance of WAI for infants beyond the neonatal period, with only two studies to date having included infants in this age range. Prieve et al. (2013b) has been the only study of infants entirely outside the newborn period (3 weeks to 9 months), and Ellison et al. (2012) included infants as young as 6 months, but also children up to 7 years of age.

The high-dimensional nature of WAI is both an advantage and disadvantage. Whilst WAI provides insight into middle ear function of the over a wide range of frequencies, each test generates a large volume of data, which clinicians can find difficult to interpret. Research investigating the most effective ways to analyse, and present results is still in its infancy (Hunter et al., 2013). Both univariate and multivariate analytical techniques have been employed in previous research (Prieve, Feeney, et al., 2013). Univariate methods have the advantage of being easier to interpret, but it has been suggested that multivariate techniques may be more accurate, since they are able to utilize more information from the WAI response (Prieve, Feeney, et al., 2013). However, studies using univariate methods have found high diagnostic accuracy with area under the receiver operating characteristic curve (AUC) of up to 0.90, which is as accurate, and even more accurate than the AUC reported in some multivariate studies (Hunter et al., 2010; Prieve et al., 2013a). Furthermore, many studies utilizing multivariate techniques likely had issues with overfitting, as they developed models with many predictors, small sample sizes, and did not use internal validation (Ellison et al., 2012; Keefe et al., 2012; Sanford et al., 2009). Overfitting means that a model has been developed on idiosyncrasies in the data, which characterise the dataset at hand very well but fail to generalise to new samples. This is a serious issue, as clinical implementation of such a model would result in a high number of misdiagnoses (Steyerberg, 2008). Further research is needed into assessing whether there is a multivariate advantage when analysing WAI data, and if so, what are appropriate ways of presenting results to clinicians. Prediction models

may be a useful methodology to help clinical interpretation of results, as they provide an intuitive summary of multivariate data in the form of a probability estimate. Presenting results as a prediction also captures the spectrum of disease (rather than simply classifying as pass/fail), which is useful when diagnosing conditions such as otitis media that occur on a spectrum, where the diagnostic threshold is somewhat arbitrary (Northrop et al., 1986; Palmu & Syrjänen, 2005; Vickers et al., 2008). Furthermore, prediction models can be extended to use an ordinal outcome, rather than a binary (pass/fail) outcome, as has been used previously in diagnostic WAI research. Using an ordinal outcome may be appropriate when analysing WAI data, since previous research has found that absorbance systematically decreases as the severity of middle ear disease increases (Beers et al., 2010; Ellison et al., 2012; Hunter et al., 2008b).

As mentioned, overfitting can be a serious issue when developing multivariate models. It is important to internally validate models at the time of development to assess the degree to which they may be overfitting the data. The authors of the Identification of Neonatal Hearing Impairment studies internally validated their studies using the opposite ears of subjects (Keefe, Gorga, et al., 2003; Keefe, Zhao, et al., 2003). Another approach to internal validation is bootstrap resampling, which estimates the amount of overfitting (bias) in a model using sampling with replacement (Steyerberg, 2008). It is vital that future multivariate WAI research include internal validation as part of development to assess the extent of overfitting in a model (Moons et al., 2012b). External validation assesses model performance in a new sample. This is an important step prior to implementing a model clinically, since model performance is usually poorer when applied to new samples. This can be due to overfitting, or differences in subject characteristics, environmental factors, or equipment used in the new setting (Steyerberg, 2008). No predictive WAI models have been externally validated in a new sample, and research is also needed in this area.

There are substantial maturational effects on WAI through infancy that need to be controlled for when developing diagnostic models (Aithal et al., 2014b; Hunter et al., 2016; Keefe et al., 1993; Shahnaz et al., 2014; Werner et al., 2010). One approach to this problem has been to create age-specific models, such as those developed specifically for use in neonates (Keefe et al., 2003a; Sanford et al., 2009). Alternatively, the effect of age could be controlled for by using developmentally stable regions of WAI as predictors in a model, or by including an interaction term between age and WAI predictors, which would allow interpretation of WAI results to vary with age (Sanford & Feeney, 2008). However, knowledge of maturational effects on WAI is a necessary first step before such strategies can be implemented. Developmental effects on absorbance and reflectance over the first year of life have been investigated (Aithal et al., 2014b; Aithal et al., 2013; Hunter et al., 2018; Hunter et al., 2010; Merchant

et al., 2010; Sanford & Feeney, 2008; Sanford et al., 2009; Shahnaz et al., 2014; Werner et al., 2010), but little is known about the effect of age on these measures over the second year of life. Keefe et al. (1993) measured reflectance in 12- and 24-month-old infants, but did not use any reference standard to assess middle ear status. There is a dearth of evidence about the effect of age for other WAI measures such as admittance through infancy.

There is a need, therefore, to further investigate the diagnostic performance of WAI for identifying conductive conditions in neonates and infants, and also to research suitable methods for interpreting results. Prediction models may be an attractive method for presenting WAI results, since they simplify complex data into a single summary – the probability of disease. There are large developmental effects on WAI through infancy, however, which need to be controlled for when developing such models, either by creating age-specific models or by incorporating age into the modelling process.

7.2 Revisiting the aims of the research

The overall aim of this work was to investigate the diagnostic performance of wideband acoustic immittance (WAI) in infants by developing predictive models. This was achieved by creating age-specific models for neonates, 6- and 12-month infants (Chapters 2, 3 and 4, respectively), investigating developmental effects on WAI through infancy (Chapter 5), then applying this knowledge in a model developed for use in a broader age range of infants (6 to 18 months; Chapter 6). Specifically, the research aimed to:

1. Develop predictive models for specific infant age groups: neonates, 6 months, and 12 months using appropriate internal validation techniques (Chapters 2, 3 and 4, respectively).
2. Investigate strategies for statistical modelling of WAI data (Chapters 2, 3, 4 and 6), including:
    a. Comparing univariate to multivariate methods (Chapters 2, 3 and 4).
    b. Approaches to reducing the large volume of WAI data, such as frequency averaging, predictor selection, and principal component analysis (PCA; Chapters 2, 3 and 4).
    c. Allowing WAI predictors to have a nonlinear association with the outcome (Chapters 2, 4 and 6).
    d. Using an ordinal outcome in predictive WAI models (Chapters 4 and 6).
    e. Whether including demographic information such as age, ethnicity and ear side improves diagnostic performance of WAI models (Chapters 2, 3, 4 and 6).
3. Investigate developmental effects of WAI through infancy, and establish normative data for various age groups (Chapter 5).

4. Develop a predictive WAI model for use in a broader age range through infancy (6 to 18 months), controlling for developmental effects (Chapter 6, Study 1).
5. Externally validate the neonate model (Chapter 2) in a new sample to assess generalizability of the model (Chapter 6, Study 2).

The main findings of the research with respect to each of these aims are discussed below.

7.3 Discussion of main findings

7.3.1 Development of predictive models for specific age groups

The first aim of the research was to develop predictive models for specific infant age groups utilising appropriate internal validation techniques. Models were developed for neonates, 6-, and 12-month infants in Chapters 2, 3 and 4, respectively. All models showed that WAI has powerful ability to predict middle ear dysfunction in infants, with *c*-index/AUC results above 0.8 after internal validation for all models (0.85, 0.85, and 0.91, for the neonate, 6-, and 12-month models, respectively). Although AUCs of the developed models were not as high as some previous reports, they were overall comparable (Beers et al., 2010; Ellison et al., 2012; Keefe et al., 2003a; Keefe et al., 2003b; Sanford et al., 2009). Furthermore, many studies reporting very high AUC in the literature (>0.90), have had serious methodological issues, such as selection bias and/or likely overfitting due to including many predictors in a model compared to the sample size (Beers et al., 2010; Ellison et al., 2012; Keefe et al., 2012).

The only previous WAI studies to include internal validation have been the Identification of Neonatal Hearing Impairment studies, which internally validated their models using the opposite ears of subjects (one ear of each subject was used to develop the models and the other for validation). They found that the models validated well with a difference in AUC between development and validation samples of 0.03 to 0.04 (Keefe, Gorga, et al., 2003; Keefe, Zhao, et al., 2003). No previous diagnostic WAI studies in infants outside the neonatal period have implemented internal validation in the modelling process. We internally validated the neonate, 6-, and 12-month models with bootstrap resampling, and the neonate and 6-month models with the opposite ears as well. Results were comparable with the Identification of Neonatal Hearing Impairment studies (Keefe, Gorga, et al., 2003; Keefe, Zhao, et al., 2003), showing maximum difference in *c*-index/AUC between development and validation samples of 0.03 for neonates (bootstrapped), 0.04 for 6 months (opposite ears), and 0.01 for 12-month infants (bootstrapped).

7.3.2 Strategies for modelling WAI data

The second aim of the study was to investigate various strategies for modelling WAI data. These included comparing multivariate and univariate models, data reduction strategies, allowing nonlinearity, using an ordinal (rather than binary) outcome, and including subject demographic information in models to assess whether this improved model fit.

7.3.2.1 Comparing univariate and multivariate modelling approaches

Multivariate models were shown to consistently outperform univariate, indicating that the extra complexity of having multiple predictors in a model was worthwhile (Chapters 2, 3 and 4). Akaike's information criterion (AIC) was always lower for multivariate models compared to univariate: 474.96 compared to 491.60 for neonates, 155.03 compared to 168.73 for 6 months, and 353.61 compared to 433.08 for 12 months. This is important, since previous studies reporting a multivariate advantage have had methodological issues, such as likely overfitting, and basing conclusions on improvements in the AUC (Keefe et al., 2012; Piskorski et al., 1999; Prieve et al., 2013a). Using the AUC to demonstrate multivariate superiority may be misleading, because it does not correct for overfitting, so a more complex model (with more predictors) may have higher AUC simply because it is overfitting the data. By using AIC to compare models, we showed that multivariate models outperformed univariate, even after controlling for model complexity.

7.3.2.2 Data reduction

Reducing the number of variables is an important aspect of modelling WAI data that has not previously received a lot of attention in the literature. The easiest way to decrease the number of variables for modelling is through simple frequency binning (averaging). In the 6-month model we compared various frequency resolutions (1, 1/2, 1/3, 1/6, 1/12 and 1/24 octave), and found that 1/2 octave frequency resolution produced the best-fitting model (based on the AIC). We also explored other methods of data reduction including predictor selection and PCA. For the neonate model (Chapter 2), we found that a priori predictor selection based on previous research yielded a slightly better model than taking an iterative approach that used the best univariate frequencies as model variables.

We also found PCA to be an effective for data reduction technique when developing the 6-month model (Chapter 3). PCA converts the original WAI frequency variables into new variables called principal components (PCs) that are ordered so that most of the information (variation) in the data is contained in the first few PCs (Jolliffe, 2002). All original WAI variables contribute to each PC, but not equally, and there are as many PCs as original variables. In a multivariate model, a subset of the PCs

can be used as predictors instead of the original variables. This retains most of the information from the originals, but fewer variables need to be included in the model, reducing risk of overfitting (Harrell, 2015). The advantage of using PCA over predictor selection for data reduction is that the model utilizes information from the entire WAI response, rather than specific regions only.

We found that using PCA in conjunction with frequency averaging for the 6-month model resulted in PCs that were readily interpretable – it was clear which frequency regions were contributing the most to each PC, and these corresponded well with known diagnostically important regions in the literature. The PCs were interpretable in this way because we did not scale variables to have a variance of 1. Scaling was not necessary because we only used absorbance frequencies as predictors in the 6-month model, which were already on the same scale. If using absorbance as well as other WAI measures (e.g., admittance) as predictors, however, it would be important to scale the WAI results before performing PCA. Doing so might reduce interpretability, however, because scaling assumes all variables are equally important, so variables (frequencies) that are apparently contributing a lot to a PC may not reflect a clinician's intuition about what the most important variables in the WAI response should be. This would not affect the accuracy of the model, only interpretability. A possible solution might be to do PCA separately for each WAI measure, and then use a subset of each as predictors in a model. For example, if using absorbance and admittance magnitude as predictors, one could do a PCA (without scaling) for absorbance (PCA_A) and another for admittance magnitude ($PCA_{|Y|}$). If using the first two PCs from each of the PCAs the model would have four predictors: $PC_{A1} + PC_{A2} + PC_{|Y|1} + PC_{|Y|2}$. This approach would eliminate the need for scaling, and the resulting PCs from each WAI measure would be readily interpretable.

7.3.2.3 Allowing nonlinearity

An assumption of simple regression is that the relationship between predictors and the reference standard behaves in a linear fashion over the entire range of predictor values. This is not necessarily the case, however, and allowing the relationship to be modelled as nonlinear may result in a better fitting model, since truly linear relationships are rare in biological data (Harrell, 2015). Even if nonlinearity does not improve model fit, there is no harm in allowing it, since nonlinearity is not enforced, only allowed if it is in the data. We found that modelling WAI as nonlinear with regression (restricted cubic) splines improved model fit (based on AIC) for the neonate model (Chapter 2) and the 6- to 18-month model (Chapter 6), but not the 12-month model (Chapter 4). This may have been due to sample size, as the neonate and 6- to 18-month models were developed using larger samples than the 12-month model. The neonate model was developed using results from 612 infants, and the 6- to 18-month model on

results of 1038 observations from 357 infants. The 12-month model, however, was developed using results from 358 observations from 186 infants. The studies with larger samples may have been able to make better use of the additional information (i.e., the extra model parameters introduced with nonlinearity).

7.3.2.4 Using an ordinal outcome

This was the first research to use an ordinal outcome to model WAI data (Chapters 4 and 6), rather than a binary (pass/fail) reference standard. Using an ordinal outcome to model WAI may be more appropriate, given that otitis media is a disease that occurs on a spectrum, and WAI has been shown to behave in an ordinal manner with the degree of severity of disease in children (Beers et al., 2010; Ellison et al., 2012; Hunter et al., 2008b). Using an ordinal outcome minimizes information loss, and may reduce the temptation for researchers to discard mild, or uncertain cases (i.e., not clearly normal or abnormal), which has been an issue with previous WAI research (Aithal et al., 2015; Beers et al., 2010; Ellison et al., 2012). The ordinal models we developed in Chapters 4 and 6 had powerful diagnostic performance, with a bias-corrected *c*-index of 0.914 and 0.867, respectively. The extra information provided by the models may be particularly useful in diagnostic settings.

7.3.2.5 Using demographic information

Including demographic variables (such as age) that are associated with the outcome as covariates in the model can help improve model fit (Harrell, 2015). In the case of middle ear dysfunction, age is the covariate that is most strongly associated with otitis media, since it is a disease of infancy. However, we did not find that controlling for demographic variables such as age, gender, ear side or ethnicity to improve model fit in our models (neonates, 6 months, 12 months, and 6 to 18 months), consistent with previous research (Beers et al., 2010; Shahnaz et al., 2013). This might be because these models were developed for specific age groups already. Including age as a covariate might be important, however, if developing a model for use in a wider age group, for example, from infancy through childhood, since infants will have higher baseline risk than older children.

7.3.3 Developmental effects on WAI

The third aim of the study was to investigate developmental effects of WAI through infancy, and establish normative data for various age groups. We found large developmental effects on WAI through infancy, consistent with previous reports (Chapter 5). The biggest changes occurred between birth and 6 months, and there were statistically significant differences between each age group (neo-

nates, 6, 12, and 18 months) in the region of 3000 to 4000 Hz. For absorbance, we recommend separate reference data be used at birth, 6 months, and 12–18 months, and for admittance (magnitude and phase, or conductance and susceptance), we advise using separate normative regions for each age group (neonates, 6, 12, and 18 months).

There were also substantial ethnicity effects for certain interactions with age and frequency for all WAI measures. For example, for absorbance at 6000 Hz at 12 months, there was over 10% difference between Caucasians and non-Caucasians, and also for admittance magnitude at 4000 to 6000 Hz at 18 months of age. This indicates that having separate ethnicity reference data or diagnostic criteria may be potentially worthwhile for certain age groups. However, previous studies in infants and school-aged children have not found using ethnic-specific normative data to improve diagnostic performance (Beers et al., 2010; Shahnaz et al., 2013), and, as mentioned, in this study including demographic information such as age and ethnicity did not improve performance of the models (Chapters 2, 3, 4 and 6).

7.3.4 Development of a model for a broader age range

The fourth aim of the research was to develop a predictive WAI model for use in a broader age range through infancy (6 to 18 months), controlling for developmental effects. An issue with developing predictive WAI models for infants is that the substantial age developmental effects on WAI need to be controlled for (Aithal et al., 2014b; Hunter et al., 2010; Hunter et al., 2016; Keefe et al., 1993; Keefe et al., 2000; Kei et al., 2013; Sanford et al., 2009; Shahnaz et al., 2014; Werner et al., 2010). We overcame this problem in Chapters 2, 3, and 4 by creating age-specific models for neonates, 6-, and 12-month infants, respectively. An alternative approach is to develop models for use over a broader age range while controlling for the effect of age on the WAI response. We suggested that this could be achieved by either including age as an interaction term with WAI variables (frequencies), or by only using developmentally stable regions of the response as predictor variables in a model (Sanford & Brockett, 2014). In Chapter 6 (Study 1) we fitted models using both of these strategies, and found that for 6- to 18-month infants, using developmentally stable predictors (1000, 1414, 2000, and 5657 Hz) for this age group produced the best-fitting model (based on the AIC). Although this strategy resulted in the best fitting model in this case, including age as an interaction term may be a better strategy in other scenarios. For example, if developing a model for use over a broad age range (e.g., neonate through childhood), developmental changes may affect much, if not all of the WAI response. In this case, controlling for age with interactions may be a better approach, since there may not be a lot of the response that is developmentally stable over such a wide age range.

7.3.5 External validation

The fifth aim of this work was to externally validate the neonate model in a new sample to assess generalizability. This was an important part of the research, since no previous WAI models have been externally validated in new subjects. Some studies have estimated the degree that the model may be overfitting with internal validation, using the opposite ears as a validation sample (Keefe, Gorga, et al., 2003; Keefe, Zhao, et al., 2003), but assessing model performance in a new sample is advised before implementing a model clinically, to give a realistic idea of how well the model is likely to generalize, since performance is usually poorer when used on new subjects (Moons et al., 2012a).

In Chapter 6 (Study 2) we applied the neonatal model developed in Chapter 2 to a new sample of 124 neonates. The model generalized well both in terms of discrimination and calibration. When applied to the new sample, the AUC was 0.837 (95% CI 0.773－0.901), and calibration was excellent, with predictions aligning closely to observed frequencies of conductive conditions. Although the AUC was lower than the apparent performance of 0.876 at model development (Chapter 2) it was close to the bias-corrected (bootstrapped) estimate of 0.845, indicating that bootstrap resampling provided a reasonable estimate of future model performance. The fact that the model validated well in a new sample indicates that the strategies used to develop the neonate model, including limiting the number of predictors, predictor selection based on prior research, and allowing nonlinearity minimized overfitting, were effective.

7.4 Strengths, limitations and directions for future research

Strengths of this research included the relatively large sample sizes used to develop the models and investigate developmental effects on WAI. We employed careful modelling strategies and internal validation with the aim of developing models likely to generalize well to new samples, and introduced advanced statistical modelling techniques into the WAI literature, such as using regression splines, shrinkage, AIC for model selection, and bootstrap resampling for internal validation of model performance. We used sophisticated methods to take into consideration correlations in the data, for example, using the Huber-White method for robust covariance matrix estimates in Chapters 4 and 6, and multilevel hierarchical models in Chapter 5. We chose appropriate statistical techniques for analysing age and demographic effects on WAI data, such Beta generalized mixed models for absorbance, which is a proportion, and analysing admittance in rectangular, rather than polar form, since the polar form is circular in nature, therefore not normally distributed.

Although we utilized sophisticated statistical and modelling techniques, we sought to develop interpretable, transparent and accessible models. We presented results in a manner that interested clinicians should be able to understand, and have provided the equations for all of the models in the Appendices (III, V, VI and VII). Implementing the models in an online application (https://joshmyers.shinyapps.io/WAIPredictions/), makes the models accessible to clinicians and researchers.

This was the first research to use an ordinal outcome to develop predictive WAI models. We used this strategy for the 12-month, and 6- to 18-month models developed in Chapters 4 and 6. This approach was able to better capture the spectrum of middle ear disease compared to using a binary (pass/fail) outcome, and eliminated the problem of how to treat mild cases, as it can be difficult to know whether these should be classified as normal or diseased if using a dichotomous outcome (Aithal et al., 2015; Beers et al., 2010; Ellison et al., 2012). This difficulty can lead to a temptation to discard data that does not neatly fit into one of the assigned categories (normal or diseased), which has been an issue with some previous WAI studies in infants and children. Ellison et al. (2012) only included normal ears (judged by an otolaryngologist with pneumatic otoscopy) or ears with confirmed middle ear effusion on surgery. Ears that were found to be dry on surgery, and abnormal ears not severe enough for surgery were not included in the study. Similarly, Beer et al. had three groups, normal, negative middle ear pressure, and middle ear effusion, but only calculated AUC using the normal and middle ear effusion groups. Aithal et al. used a test battery as the reference standard, but only included ears that either passed or failed all tests in the test battery (e.g., both tympanometry and distortion product otoacoustic emissions). This exclusion of certain groups of subjects in these studies meant that the performance of WAI was assessed on a subset of subjects that had been selected based on the results of the reference tests. By removing mild or uncertain diagnoses, the test is only being assessed on easy-to-diagnose cases, so test performance measures (e.g., AUC) may be artificially inflated (overly optimistic; Bossuyt et al., 2003). From a clinical point of view, it is understandable why these cases were excluded in each of these studies, but the resulting selection bias is a problem when attempting to interpret the performance measures. In all of these studies, the problem could have been addressed by using an ordinal reference standard with a third "mild" or "uncertain" group.

The longitudinal design of the developmental study in Chapter 5 was a strength, as observed developmental effects were likely due to real changes, rather that differences between subjects. A limitation, however, was that no data were collected between birth and 6 months of age, where some of the largest developmental effects on WAI occur. Furthermore, no predictive models were developed for infants in this age range in this study (1 to 5 months), which is an age group often seen for diagnostic

audiological follow up after the newborn hearing screen. Future research developing predictive models for use in this age group should be a priority.

This research was the first to develop a model controlling for age effects over a broader age range of infants (Chapter 6, Study 1). The methods investigated, including using interactions with age, and/or developmentally stable regions of WAI, could be used to develop other age-independent models (e.g., models for use in an even wider range of paediatric subjects). However, although we developed a model for use in 6- to 18-month infants, this range is still relatively narrow compared to the age range of patients typically seen in a paediatric audiology, or doctor's clinic. Future research could develop models for use over a larger age range of paediatric patients. For example, with a large enough sample, a model could potentially be developed that could be used in any paediatric patient (birth to 18 years). Such a model may include age as an interaction term with WAI variables and also age as a predictor (term) in the model, since infants have higher baseline risk of otitis media. However, neonates may still need their own model, since they are so different from other age groups and developmental changes in WAI are much more rapid at this age. Whether or not this is the case remains a question for future research.

This study was the first to externally validate a WAI model in a new sample (Chapter 6, Study 2). A limitation of the external validation study, however, was that although the model generalized well, subjects in the validation sample were of a similar age, and tested in the same environment by the same researchers used to develop the model. Future research could validate this model (Section 2.6: Appendix B) in a sample tested in a different environment, updating it as necessary. Further research could also externally validate the infant models (the equations in Sections 3.7, 4.6 and 6.6), and studies assessing the clinical impact the models would also be valuable. Research into clinical impact is important to assess whether implementing a model improves clinical decision making and patient outcomes (Kappen et al., 2018).

A limitation of this research was that data collection was not blinded. When testing a neonate or infant, both the reference standard tests and WAI measurements were performed by the same research audiologist. Even though interpretation of the reference tests was objective, this may have introduced bias. Another limitation was that the reference standard used to develop the infant models was not the gold standard for diagnosing middle ear disease in infants (Chapters 3, 4 and 6). Future research using reference tests such as otomicroscopy and surgical confirmation of middle ear status could create a more stringent reference standard with well-defined pathologies for an ordinal reference standard such as Eustachian tube dysfunction, partial, and complete middle ear effusion.

Further research is also needed investigating the diagnostic performance of WAI in other paediatric middle ear conditions. This research investigated developing predictive models for otitis media and Eustachian tube dysfunction, but there is evidence that WAI is sensitive to a number of other conditions including conductive hearing loss, ossicular chain discontinuity, otosclerosis, perforated eardrum, hypermobile tympanic membrane, and superior semicircular canal dehiscence (Feeney, Grant, & Marryott, 2003; Feeney, Grant, & Mills, 2009; Keefe et al., 2012; Nakajima et al., 2012; Nakajima, Rosowski, Shahnaz, & Voss, 2013; Piskorski et al., 1999; Shahnaz, Longridge, & Bell, 2009; Voss et al., 2012). Predictive models could potentially be developed for specific conditions, or multiple conditions in a single model. For example, a model could be developed using a nominal (unordered) outcome with multiple levels such as middle ear effusion, Eustachian tube dysfunction, perforated eardrum and hypermobile tympanic membrane.

This work used traditional statistical risk modelling to develop predictive models, but there have been recent advances in the field of machine learning and artificial intelligence, in particular deep learning (models made up of layers of artificial neural networks), that have shown state-of-the-art predictive performance for image and speech recognition problems (Marcus, 2018). A difficulty in applying deep learning techniques (and machine learning methods in general) to health and medical problems, however, is the large amount of data needed to develop these models. Large WAI studies have hundreds, perhaps thousands of subjects, but a deep learning model is ideally fitted using tens of thousands of samples (Gulshan et al., 2016). Transfer learning, however, makes it possible to repurpose a deep learning model for a problem that may be quite different than the initial task it was designed for. In such cases, it is conceivable to fit a model with hundreds, perhaps thousands of observations, rather than needing tens of thousands. For example, if we think of predicting WAI results as an image classification problem, we could use a deep learning model that has been trained for an image classification task on millions of images, and repurpose (fine-tune) it using a sample of WAI "images". This repurposing works because the initial layers of a deep learning model are not specific to the dataset used to create the model, being for more general tasks, such as identifying edges and contrasts (Yosinski, Clune, Bengio, & Lipson, 2014). Figure 7-1 shows how absorbance results might be presented for use in an image classification model. In this case the image (pixels) would be the predictor in the model rather than absorbance frequencies. Whether such a model would perform better than traditional modelling methods remains to be seen.

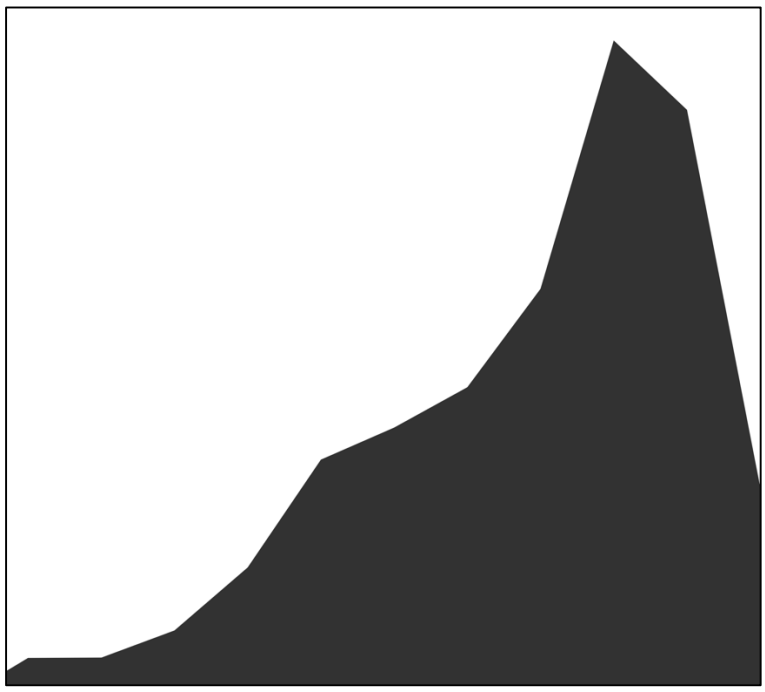


Figure 7-1. Potential way of preparing absorbance results as an image for use in a deep learning model.
The *x*-axis is frequency and *y*-axis absorbance. Absorbance is depicted at 1/2 octave frequency resolution, although this need not be the case. Results have been saved as a square image with axes and white space around the edges removed. The black border would be unnecessary.

7.5 Conclusion

This work developed predictive models for diagnosing conductive dysfunction in neonates, 6-, and 12-month infants, investigated developmental effects on WAI, and used this knowledge to create a model that controlled for the effect of age for use in 6- to 18-month infants.

The models developed for specific age groups (neonates, 6, and 12 months) accurately identified conductive conditions in infants. Models were carefully fitted and internally validated to increase the likelihood that they will generalize to new samples. Multivariate methods consistently outperformed univariate, and frequency binning (averaging), PCA, and selecting predictors based on prior research all worked well as data reduction techniques. Data mining approaches do have merit, however, and may be especially useful for informing subsequent studies. This may be especially important for studies of pressurized WAI, where little is known about diagnostically important features of the wideband tympanogram. If using PCA for data reduction for multilevel reference standard model (e.g., ordinal or multinomial), it would be advisable to limit the range of WAI frequencies used for PCA to those that show a separation between the various groups. For example, if using an ordinal model, it would be preferable to limit the WAI predictors in the PCA to those that show an ordinal association with the response (the reference standard levels; Harrell, 2015).

Allowing WAI to have a nonlinear association with the outcome did not always improve model fit, but did in two out of three models (Chapters 2 and 6), notably the models with larger sample sizes, which may have been able to make better use of the additional information (model parameters). We recommend trying this approach in future research, especially if there are enough degrees of freedom, since allowing nonlinearity is generally preferable to enforcing linearity (Harrell, 2015).

This research was the first to utilise an ordinal outcome when modelling WAI data, and we recommend this approach for future research investigating conditions such as otitis media that have an ordinal association with WAI. Having a "mild" or "uncertain" category can avoid the problem of what to do with borderline or uncertain cases, and removes the temptation to omit these observations, which can introduce selection bias.

We found statistically significant ear-side and ethnicity effects, but including demographic information as predictors did not improve model performance. However, if developing a model for use in a larger age range, for example, a model encompassing infancy and childhood, it may be important to include age as a demographic predictor, since infants will have higher baseline risk of otitis media (i.e., it is more common in infants).

We investigated developmental effects on the WAI response, and established normative data for various infant age groups. We also found statistically significant interactions for certain combinations of age, ethnicity and frequency. For example, 12-month infants had over 10% difference between Caucasians and non-Caucasians for absorbance at 6000 Hz. However, previous studies in infants and school-aged children have not found using ethnic-specific normative data to improve diagnostic performance (Beers et al., 2010; Shahnaz et al., 2013), and in this study including demographic information such as age and ethnicity as predictors did not improve model performance (Chapters 2, 3, 4, and 6). The model controlling for the effect of age (Chapter 6, Study 1) performed very well in terms of discrimination (bias-corrected *c*-index = 0.867) and calibration.

The neonate model developed in Chapter 2 validated well to a new sample (Chapter 6, Study 2). This was an important finding, as it was the first WAI model to be validated in an external sample. The developed models may be clinically useful, future research further validating the models and investigating clinical impact is warranted. A priority for future research should be to develop models for other age groups, particularly for infants aged 1 to 5 months. With a large and diverse enough sample could potentially develop a model encompassing most of childhood, although it may be advisable to have a separate model for neonates and possibly young infants, since there are such strong developmental effects for these age groups.

In conclusion, the developed models accurately identified conductive conditions in infants. The models were carefully fitted and internally validated, to increase the likelihood that they will generalize to new samples. The neonate model did effectively generalize to a new sample, which indicates that the strategies employed to minimize overfitting were effective. There were large developmental effects on WAI measurements, and this knowledge was used to develop a model that controlled for maturational effects through infancy. The models have potential applications in both screening and diagnostic settings. In a screening context, predictions could be used to set a referral threshold sensitive to the costs associated with true, and false positive referrals, that is intuitive and easy to apply. In a diagnostic setting, predicted probabilities could be used in conjunction with graphical depictions of WAI for individualized diagnoses of conductive dysfunction. Further research validating, updating, and assessing the clinical impact of the models is warranted.

List of References

Aarhus, L., Tambs, K., Kvestad, E., & Engdahl, B. (2014). Childhood otitis media: A cohort study with 30-year follow-up of hearing (the HUNT study). *Ear and Hearing, 36*(3), 302-308.

Agresti, A. (2013). *Categorical Data Analysis* (3rd ed.). Hoboken, NJ: John Wiley & Sons.

Aithal, S., Aithal, V., Kei, J., Anderson, S., & Liebenberg, S. (2018). Eustachian tube dysfunction and wideband absorbance measurements at tympanometric peak pressure and 0 daPa. *Journal of the American Academy of Audiology*.

Aithal, S., Aithal, V., Kei, J., & Driscoll, C. (2012). Conductive hearing loss and middle ear pathology in young infants referred through a newborn universal hearing screening program in Australia. *Journal of the American Academy of Audiology, 23*(9), 673-685.

Aithal, S., Kei, J., Aithal, V., Manuel, A., Myers, J., Driscoll, C., & Khan, A. (2017). Normative study of wideband acoustic immittance measures in newborn infants. *Journal of Speech Language and Hearing Research, 60*(5), 1417-1426.

Aithal, S., Kei, J., & Driscoll, C. (2014a). Wideband absorbance in Australian Aboriginal and Caucasian neonates. *Journal of the American Academy of Audiology, 25*(5), 482-494.

Aithal, S., Kei, J., & Driscoll, C. (2014b). Wideband absorbance in young infants (0-6 months): A cross-sectional study. *Journal of the American Academy of Audiology, 25*(5), 471-481.

Aithal, S., Kei, J., Driscoll, C., & Khan, A. (2013). Normative wideband reflectance measures in healthy neonates. *International Journal of Pediatric Otorhinolaryngology, 77*(1), 29-35.

Aithal, S., Kei, J., Driscoll, C., Khan, A., & Swanston, A. (2015). Wideband absorbance outcomes in newborns: A comparison with high-frequency tympanometry, automated brainstem response, and transient evoked and distortion product otoacoustic emissions. *Ear and Hearing, 36*(5), e237-e250.

Aithal, S., Yonovitz, A., & Aithal, V. (2008). Perceptual consequences of conductive hearing loss: Speech perception in Indigenous students learning English as a 'school' language. *The Australian and New Zealand Journal of Audiology, 30*(1), 1-18.

Akdogan, O., & Özkan, S. (2006). Otoacoustic emissions in children with otitis media with effusion. *International Journal of Pediatric Otorhinolaryngology, 70*(11), 1941-1944.

Akinpelu, O. V., Funnell, W. R. J., & Daniel, S. J. (2015). Detection of otoacoustic emissions in chinchilla when the middle ear contains amniotic fluid. *The Laryngoscope, 125*(4), E138-E142.

Alaerts, J., Luts, H., & Wouters, J. (2007). Evaluation of middle ear function in young children: Clinical guidelines for the use of 226- and 1,000-Hz tympanometry. *Otology & Neurotology, 28*(6), 727-732.

Alho, O.-P., Koivu, M., Sorri, M., & Rantakallio, P. (1991). The occurrence of acute otitis media in infants. A life-table analysis. *International Journal of Pediatric Otorhinolaryngology, 21*(1), 7-14.

Allen, E. K., Manichaikul, A., & Sale, M. M. (2014). Genetic contributors to otitis media: Agnostic discovery approaches. *Current Allergy and Asthma Reports, 14*(2), 1-6.

Allen, J. B. (1986). Measurement of eardrum acoustic impedance. In J. B. Allen, J. L. Hall, A. Hubbard, S. T. Neely, & A. Tubis (Eds.), *Peripheral Auditory Mechanisms* (pp. 44-51). Berlin, Germany: Springer-Verlag.

Allen, J. B., Jeng, P. S., & Levitt, H. (2005). Evaluation of human middle ear function via an acoustic power assessment. *Journal of Rehabilitation Research and Development, 42*(4), 63-78.

Alonzo, T. A., & Pepe, M. S. (1999). Using a combination of reference tests to assess the accuracy of a new diagnostic test. *Statistics in Medicine, 18*(22), 2987-3003.

Amedee, R. G. (1995). The effects of chronic otitis media with effusion on the measurement of transiently evoked otoacoustic emissions. *Laryngoscope, 105*(6), 589-595.

Bagherzadeh-Khiabani, F., Ramezankhani, A., Azizi, F., Hadaegh, F., Steyerberg, E. W., & Khalili, D. (2016). A tutorial on variable selection for clinical prediction models: Feature selection methods in data mining could improve the results. *Journal of Clinical Epidemiology, 71*, 76-85.

Baldwin, M. (2006). Choice of probe tone and classification of trace patterns in tympanometry undertaken in early infancy. *International Journal of Audiology, 45*(7), 417-427.

Balkany, T. J., Berman, S. A., Simmons, M. A., & Jafek, B. W. (1978). Middle ear effusions in neonates. *The Laryngoscope, 88*(3), 398-405.

Baraibar, R. (1997). Incidence and risk factors of acute otitis media in children. *Clinical Microbiology and Infection, 3*, 3S13-13S22.

Barber, C., Ille, S., Vergison, A., & Coates, H. (2014). Acute otitis media in young children-What do parents say? *International Journal of Pediatric Otorhinolaryngology, 78*(2), 300-306.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Beers, A. N., Shahnaz, N., Westerberg, B. D., & Kozak, F. K. (2010). Wideband reflectance in normal Caucasian and Chinese school-aged children and in children with otitis media with effusion. *Ear and Hearing, 31*(2), 221-233.

Benade, A. H. (1968). On the propagation of sound waves in a cylindrical conduit. *The Journal of the Acoustical Society of America, 44*(2), 616-623.

Bennett, K., Haggard, M., Silva, P., & Stewart, I. (2001). Behaviour and developmental effects of otitis media with effusion into the teens. *Archives of Disease in Childhood, 85*(2), 91-95.

Berkun, Y., Nir-Paz, R., Ami, A. B., Klar, A., Deutsch, E., & Hurvitz, H. (2008). Acute otitis media in the first two months of life: Characteristics and diagnostic difficulties. *Archives of Disease in Childhood, 93*(8), 690-694.

Blankenship, C. M., Hunter, L. L., Keefe, D. H., Feeney, M. P., Brown, D. K., McCune, A., . . . Lin, L. (2018). Optimizing clinical interpretation of distortion product otoacoustic emissions in infants. *Ear and Hearing, 39*(6), 1075-1090.

Bluestone, C. D. (1998). Epidemiology and pathogenesis of chronic suppurative otitis media: Implications for prevention and treatment. *International Journal of Pediatric Otorhinolaryngology, 42*(3), 207-223.

Bluestone, C. D., & Klein, J. O. (2007). *Otitis Media in Infants and Children*. Hamilton, Canada: BC Decker Inc.

Bossuyt, P. M., Reitsma, J. B., Bruns, D. E., Gatsonis, C. A., Glasziou, P. P., Irwig, L. M., . . . Lijmer, J. G. (2003). The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Annals of Internal Medicine, 138*(1), W1-12.

Box, G. E., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 211-252.

Brennan-Jones, C. G., Whitehouse, A. J., Park, J., Hegarty, M., Jacques, A., Eikelboom, R. H., . . . Jamieson, S. E. (2014). Prevalence and risk factors for parent-reported recurrent otitis media during early childhood in the Western Australian Pregnancy Cohort (Raine) Study. *Journal of Paediatrics and Child Health, 249*(8), 1022-1025.

Briggs, W. M., & Zaretzki, R. (2008). The skill plot: A graphical technique for evaluating continuous diagnostic tests. *Biometrics, 64*(1), 250-256; discussion, 256-261.

Brooks, M. E., Kristensen, K., van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., . . . Bolker, B. M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *R Journal, 9*(2), 378-400.

Burnham, K. P., & Anderson, D. R. (2002). *Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach* (2nd ed.). New York, NY: Springer.

Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of Statistical Software, 45*(3).

Calandruccio, L., Fitzgerald, T. S., & Prieve, B. A. (2006). Normative multifrequency tympanometry in infants and toddlers. *Journal of the American Academy of Audiology, 17*(7), 470-480.

Cameron, S., Dillon, H., Glyde, H., Kanthan, S., & Kania, A. (2014). Prevalence and remediation of spatial processing disorder (SPD) in Indigenous children in regional Australia. *International Journal of Audiology, 53*(5), 326-335.

Cantekin, E., Bluestone, C. D., Fria, T., Stool, S., Beery, Q., & Sabo, D. (1979). Identification of otitis media with effusion in children. *The Annals of Otology, Rhinology & Laryngology. Supplement, 89*(3 Pt 2), 190-195.

Casselbrant, M. L., Furman, J. M., Mandel, E. M., Fall, P. A., Kurs-Lasky, M., & Rockette, H. E. (2000). Past history of otitis media and balance in four-year-old children. *The Laryngoscope, 110*(5), 773-778.

Casselbrant, M. L., & Mandel, E. M. (2014). Acute otitis media and otitis media with effusion. In M. M. Lesperance & P. W. Flint (Eds.), *Cummings Pediatric Otolaryngology* (pp. 209-227). Philadelphia, PA: Elsevier Health Sciences.

Casselbrant, M. L., Mandel, E. M., Kurs-Lasky, M., Rockette, H. E., & Bluestone, C. D. (1995). Otitis media in a population of Black American and White American infants, 0-2 years of age. *International Journal of Pediatric Otorhinolaryngology, 33*(1), 1-16.

Cavanaugh, R. M. (1987). Pneumatic otoscopy in healthy full-term infants. *Pediatrics, 79*(4), 520-523.

Cavanaugh, R. M. (1989). Pediatricians and the pneumatic otoscope: Are we playing it by ear? *Pediatrics, 84*(2), 362-364.

Chang, K. W., Vohr, B. R., Norton, S. J., & Lekas, M. D. (1993). External and middle ear status related to evoked otoacoustic emission in neonates. *Archives of Otolaryngology - Head and Neck Surgery, 119*(3), 276-282.

Chianese, J., Hoberman, A., Paradise, J. L., Colborn, D. K., Kearney, D., Rockette, H. E., & Kurs-Lasky, M. (2007). Spectral gradient acoustic reflectometry compared with tympanometry in diagnosing middle ear effusion in children aged 6 to 24 months. *Archives of Pediatrics and Adolescent Medicine, 161*(9), 884-888.

Choi, S. S., Pafitis, I. A., Herer, G. R., Zalzal, G. H., & Patel, K. M. (1999). Clinical applications of transiently evoked otoacoustic emissions in the pediatric population. *Annals of Otology, Rhinology and Laryngology, 108*(2), 132-138.

Chonmaitree, T., Revai, K., Grady, J. J., Clos, A., Patel, J. A., Nair, S., . . . Henrickson, K. J. (2008). Viral upper respiratory tract infection and otitis media complication in young children. *Clinical Infectious Diseases, 46*(6), 815-823.

Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. (2015). Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. *Annals of Internal Medicine, 162*(1), 55-63.

Corbeel, L. (2007). What is new in otitis media? *European Journal of Pediatrics, 166*(6), 511-519.

Cunningham, M., Guardiani, E., Kim, H. J., & Brook, I. (2012). Otitis media. *Future Microbiology, 7*(6), 733-753.

da Costa, S. S., Rosito, L. P. S., & Dornelles, C. (2009). Sensorineural hearing loss in patients with chronic otitis media. *European Archives of Oto-Rhino-Laryngology, 266*(2), 221-224.

Daly, K. A., & Giebink, G. S. (2000). Clinical epidemiology of otitis media. *The Pediatric Infectious Disease Journal, 19*(5), S31-S36.

Daly, K. A., Hoffman, H. J., Kvaerner, K. J., Kvestad, E., Casselbrant, M. L., Homoe, P., & Rovers, M. M. (2010). Epidemiology, natural history, and risk factors: Panel report from the Ninth International Research Conference on Otitis Media. *International Journal of Pediatric Otorhinolaryngology, 74*(3), 231-240.

Damoiseaux, R. A., Rovers, M. M., Van Balen, F. A., Hoes, A. W., & de Melker, R. A. (2006). Long-term prognosis of acute otitis media in infancy: Determinants of recurrent acute otitis media and persistent middle ear effusion. *Family Practice, 23*(1), 40-45.

Dawes, P. J., & Welch, D. (2010). Childhood hearing and its relationship with tinnitus at thirty-two years of age. *Annals of Otology, Rhinology and Laryngology, 119*(10), 672.

De Lyra-Silva, K. A., Sanches, S. G. G., Neves-Lobo, I. F., Ibidi, S. M., & Carvallo, R. M. M. (2015). Middle ear muscle reflex measurement in neonates: Comparison between 1000Hz and 226Hz probe tones. *International Journal of Pediatric Otorhinolaryngology, 79*(9), 1510-1515.

De Sa, D. (1973). Infection and amniotic aspiration of middle ear in stillbirths and neonatal deaths. *Archives of Disease in Childhood, 48*(11), 872.

DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics, 44*(3), 837-845.

Diefendorf, A. O. (2015). Assessment of hearing loss in children. In J. Katz, M. Chasin, K. English, L. J. Hood, & K. L. Tillery (Eds.), *Handbook of Clinical Audiology* (pp. 459-476). Philadelphia, PA: Wolters Kluwer.

Doyle, K. J., Burggraaff, B., Fujikawa, S., Kim, J., & Macarthur, C. J. (1997). Neonatal hearing screening with otoscopy, auditory brain stem response, and otoacoustic emissions. *Otolaryngology--Head and Neck Surgery, 116*(6), 597-603.

Doyle, K. J., Kong, Y. Y., Strobel, K., Dallaire, P., & Ray, R. M. (2004). Neonatal middle ear effusion predicts chronic otitis media with effusion. *Otology & Neurotology, 25*(3), 318-322.

Doyle, K. J., Rodgers, P., Fujikawa, S., & Newman, E. (2000). External and middle ear effects on infant hearing screening test results. *Otolaryngology--Head and Neck Surgery, 122*(4), 477-481.

Driscoll, C., Kei, J., & McPherson, B. (2001). Outcomes of transient evoked otoacoustic emission testing in 6-year-old school children: A comparison with pure tone screening and tympanometry. *International Journal of Pediatric Otorhinolaryngology, 57*(1), 67-76.

Eavey, R. D. (1993). Abnormalities of the neonatal ear: otoscopic observations, histologic observations, and a model for contamination of the middle ear by cellular contents of amniotic fluid. *The Laryngoscope, 103*(1 Pt 2 Suppl 58), 1-31.

Ellison, J. C., Gorga, M., Cohn, E., Fitzpatrick, D., Sanford, C. A., & Keefe, D. H. (2012). Wideband acoustic transfer functions predict middle-ear effusion. *Laryngoscope, 122*(4), 887-894.

Engel, J. A., Anteunis, L., Volovics, A., Hendriks, J., & Marres, E. (1999). Prevalence rates of otitis media with effusion from 0 to 2 years of age: Healthy-born versus high-risk-born infants. *International Journal of Pediatric Otorhinolaryngology, 47*(3), 243-251.

Feeney, M. P., Grant, I. L., & Marryott, L. P. (2003). Wideband energy reflectance measurements in adults with middle-ear disorders. *Journal of Speech, Language, and Hearing Research*.

Feeney, M. P., Grant, I. L., & Mills, D. M. (2009). Wideband energy reflectance measurements of ossicular chain discontinuity and repair in human temporal bone. *Ear and Hearing, 30*(4), 391-400.

Feeney, M. P., Hunter, L. L., Kei, J., Lilly, D. J., Margolis, R. H., Nakajima, H. H., . . . Voss, S. E. (2013). Consensus statement: Eriksholm Workshop on wideband absorbance measures of the middle ear. *Ear and Hearing, 34*, 78s-79s.

Finitzo, T., Friel-Patti, S., Chinn, K., & Brown, O. (1992). Tympanometry and otoscopy prior to myringotomy: Issues in diagnosis of otitis media. *International Journal of Pediatric Otorhinolaryngology, 24*(2), 101-110.

Fortanier, A. C., Venekamp, R. P., Boonacker, C. W., Hak, E., Schilder, A. G., Sanders, E. A., & Damoiseaux, R. A. (2014). Pneumococcal conjugate vaccines for preventing otitis media. *Cochrane Database of Systematic Reviews, 2014*(4), 1-55.

Freyss, G., Narcy, P., Manac'h, Y., & Toupet, M. (1979). Acoustic reflex as a predictor of middle ear effusion. *The Annals of Otology, Rhinology & Laryngology. Supplement, 89*(3 Pt 2), 196-199.

Fria, T. J., Cantekin, E. I., & Eichler, J. A. (1985). Hearing acuity of children with otitis media with effusion. *Archives of Otolaryngology, 111*(1), 10-16.

Froom, J., Culpepper, L., Grob, P., Bartelds, A., Bowers, P., Bridges-Webb, C., . . . Somaini, B. (1990). Diagnosis and antibiotic treatment of acute otitis media: Report from International Primary Care Network. *BMJ, 300*(6724), 582-586.

Gelman, A., & Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models.* Cambridge, England: Cambridge University Press.

Goossens, H., Ferech, M., Vander Stichele, R., Elseviers, M., & ESAC Project Group. (2005). Outpatient antibiotic use in Europe and association with resistance: A cross-national database study. *The Lancet, 365*(9459), 579-587.

Gorga, M. P., Dierking, D. M., Johnson, T. A., Beauchaine, K. L., Garner, C. A., & Neely, S. T. (2005). A validation and potential clinical application of multivariate analyses of DPOAE data. *Ear and Hearing, 26*(6), 593.

Gould, J. M., & Matz, P. S. (2010). Otitis media. *Pediatrics in review/American Academy of Pediatrics, 31*(3), 102-116.

Gouma, P., Mallis, A., Daniilidis, V., Gouveris, H., Armenakis, N., & Naxakis, S. (2011). Behavioral trends in young children with conductive hearing loss: A case-control study. *European Archives of Oto-Rhino-Laryngology, 268*(1), 63-66.

Gravel, J. S., & Wallace, I. F. (1992). Listening and language at 4 years of age: effects of early otitis media. *Journal of Speech and Hearing Research, 35*(3), 588-595.

Graydon, K., Rance, G., Dowell, R., & Van Dun, B. (2017). Consequences of Early Conductive Hearing Loss on Long-Term Binaural Processing. *Ear and Hearing, 38*(5), 621-627.

Grindler, D. J., Blank, S. J., Schulz, K. A., Witsell, D. L., & Lieu, J. E. (2014). Impact of otitis media severity on children's quality of life. *Otolaryngology-Head and Neck Surgery*, 0194599814525576.

Groon, K. A., Rasetshwane, D. M., Kopun, J. G., Gorga, M. P., & Neely, S. T. (2015). Air-leak effects on ear-canal acoustic absorbance. *Ear and Hearing, 36*(1), 155-163.

Guggenheim, P., Clements, L., & Schlesinger, A. (1956). The significance and fate of the mesenchyme of the middle ear. *The Laryngoscope, 66*(10), 1303-1326.

Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., . . . Cuadros, J. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA, 316*(22), 2402-2410.

Gunasekera, H., Knox, S., Morris, P. S., Britt, H., McIntyre, P., & Craig, J. C. (2007). The spectrum and management of otitis media in Australian Indigenous and nonIndigenous children: A national study. *The Pediatric Infectious Disease Journal, 26*(8), 689-692.

Hall, J. W., 3rd, Grose, J. H., & Mendoza, L. L. (1995). Masker interaural phase and the MLD: effects of conductive hearing loss. *Hearing Research, 84*(1-2), 91-98.

Hardin, J. (2005). Small-sample adjustments to the sandwich estimate of variance. Retrieved from https://www.stata.com/support/faqs/statistics/sandwich-estimate-of-variance/

Harmes, K. M., Blackwood, R. A., Burrows, H. L., Cooke, J. M., Harrison, R., & Passamani, P. P. (2013). Otitis media: Diagnosis and treatment. *American Family Physician, 88*(7), 435-440.

Harrell, F. E., Jr. (2015). *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis* (2nd ed.). New York, NY: Springer.

Harrell, F. E., Jr. (2016). rms: Regression Modeling Strategies [R package]. (Version 5.1-1). Retrieved from https://CRAN.R-project.org/package=rms

Harrell, F. E., Jr., Lee, K., & Mark, D. (1996). Multivariable prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in Medicine, 15*(4), 361-387.

Harrell, F. E., Jr., & Slaughter, J. C. (2016). *Biostatistics for biomedical research*.

Harsten, G., Prellner, K., Heldrup, J., Kalm, O., & Kornfält, R. (1989). Recurrent acute otitis media: A prospective study of children during the first three years of life. *Acta Oto-Laryngologica, 107*(1-2), 111-119.

Hayden, G. F. (1981). Acute suppurative otitis media in children: Diversity of clinical diagnostic criteria. *Clinical Pediatrics, 20*(2), 99-104.

Hoberman, A., Paradise, J. L., Rockette, H. E., Shaikh, N., Wald, E. R., Kearney, D. H., . . . Barbadora, K. A. (2011). Treatment of acute otitis media in children under 2 years of age. *New England Journal of Medicine, 364*(2), 105-115.

Hoberman, A., Ruohola, A., Shaikh, N., Tähtinen, P. A., & Paradise, J. L. (2013). Acute otitis media in children younger than 2 years. *JAMA Pediatrics, 167*(12), 1171-1172.

Hoffman, H. J., Daly, K. A., Bainbridge, K. E., Casselbrant, M. L., Homøe, P., Kvestad, E., . . . Vernacchio, L. (2013). Panel 1: Epidemiology, natural history, and risk factors. *Otolaryngology-Head and Neck Surgery, 148*(4 suppl), E1-E25.

Hoffmann, A., Deuster, D., Rosslau, K., Knief, A., am Zehnhoff-Dinnesen, A., & Schmidt, C. M. (2013). Feasibility of 1000 Hz tympanometry in infants: Tympanometric trace classification and choice of probe tone in relation to age. *International Journal of Pediatric Otorhinolaryngology, 77*(7), 1198-1203.

Hogan, A., Phillips, R. L., Howard, D., & Yiengprugsawan, V. (2014). Psychosocial outcomes of children with ear infections and hearing problems: A longitudinal study. *BMC Pediatrics, 14*(1), 65.

Holte, L., Cavanaugh, R. M., & Margolis, R. H. (1990). Ear canal wall mobility and tympanometric shape in young infants. *The Journal of Pediatrics, 117*(1), 77-80.

Holte, L., Margolish, R. H., & Cavanaugh, R. M. (1991). Developmental changes in multifrequency tympanograms. *Audiology, 30*(1), 1-24.

Homøe, P., Christensen, R. B., & Bretlau, P. (1999). Acute otitis media and age at onset among children in Greenland. *Acta Oto-Laryngologica, 119*(1), 65-71.

Howie, V. M., Ploussard, J. H., & Sloyer, J. (1975). The otitis-prone condition. *American Journal of Diseases of Children, 129*(6), 676-678.

Huang, G. T., Rosowski, J. J., & Peake, W. T. (2000). Relating middle-ear acoustic performance to body size in the cat family: Measurements and models. *Journal of Comparative Physiology. A: Sensory, Neural, and Behavioral Physiology, 186*(5), 447-465.

Hudde, H. (1983). Measurement of the eardrum impedance of human ears. *The Journal of the Acoustical Society of America, 73*(1), 242-247.

Hunter, L. L., Bagger-Sjöbäck, D., & Lundberg, M. (2008a). Wideband reflectance associated with otitis media in infants and children with cleft palate. *International Journal of Audiology, 47*(S1), S57-S61.

Hunter, L. L., Blankenship, C. M., Keefe, D. H., Feeney, M. P., Brown, D. K., McCune, A., . . . hearing. (2018). Longitudinal development of distortion product otoacoustic emissions in infants with normal hearing. *Ear and Hearing, 39*(5), 863-873.

Hunter, L. L., Feeney, M. P., Miller, J. A. L., Jeng, P. S., & Bohning, S. (2010). Wideband reflectance in newborns: Normative regions and relationship to hearing screening results. *Ear and Hearing, 31*(5), 599.

Hunter, L. L., Keefe, D. H., Feeney, M. P., Fitzpatrick, D. F., & Lin, L. (2016). Longitudinal development of wideband reflectance tympanometry in normal and at-risk infants. *Hearing Research, 340*, 3-14.

Hunter, L. L., Prieve, B. A., Kei, J., & Sanford, C. A. (2013). Pediatric applications of wideband acoustic immittance measures. *Ear and Hearing, 34*, 36s-42s.

Hunter, L. L., & Shahnaz, N. (2014). *Acoustic Immittance Measures: Basic and Advanced Practics*. San Diego, CA: Plural Publishing.

Hunter, L. L., Tubaugh, L., Jackson, A., & Propes, S. (2008b). Wideband middle ear power measurement in infants and children. *Journal of the American Academy of Audiology, 19*(4), 309-324.

Iacovidou, N., Falaena, V., Alexaki, A., & Nika, A. (2010). Spontaneous perforation of the tympanic membrane in the first 10 days of life. *American Journal of Perinatology, 27*(8), 663-665.

Jacob-Corteletti, L. C. B., Duarte, J. L., Zucki, F., Mariotto, L. D. F., Lauris, J. R. P., & Alvarenga, K. d. F. (2015). Acoustic reflex on newborns: The influence of the 226 and 1,000 Hz probes. *CoDAS, 27*(3), 223-229.

Jaffe, B. F., Hurtado, F., & Hurtado, E. (1970). Tympanic membrane mobility in the newborn (with seven months' follow-up). *The Laryngoscope, 80*(1), 36-48.

Jerger, J. (1970). Clinical experience with impedance audiometry. *Archives of Otolaryngology, 92*(4), 311-324.

Joglekar, S., Morita, N., Cureoglu, S., Schachern, P. A., Deroee, A. F., Tsuprun, V., . . . Juhn, S. K. (2010). Cochlear pathology in human temporal bones with otitis media. *Acta Oto-Laryngologica, 130*(4), 472-476.

Joint Committee on Infant Hearing. (2007). Year 2007 position statement: Principles and guidelines for early hearing detection and intervention programs. *Pediatrics, 120*(4), 898-921.

Jolliffe, I. T. (2002). *Principal Component Analysis* (2nd ed.). New York, NY: Springer.

Jung, T. T., Alper, C. M., Hellstrom, S. O., Hunter, L. L., Casselbrant, M. L., Groth, A., . . . Nittrouer, S. (2013). Panel 8 Complications and sequelae. *Otolaryngology-Head and Neck Surgery, 148*(4 suppl), E122-E143.

Kappen, T. H., van Klei, W. A., van Wolfswinkel, L., Kalkman, C. J., Vergouwe, Y., & Moons, K. G. (2018). Evaluating the impact of prediction models: Lessons learned, challenges, and recommendations. *Diagnostic and Prognostic Research, 2*, 1-11.

Karma, P. H., Penttilä, M. A., Sipilä, M. M., & Kataja, M. J. (1989a). Otoscopic diagnosis of middle ear effusion in acute and non-acute otitis media. I. The value of different otoscopic findings. *International Journal of Pediatric Otorhinolaryngology, 17*(1), 37-49.

Karma, P. H., Perälä, M., & Kuusela, A. (1989b). Morbidity of very young infants with and without acute otitis media. *Acta Oto-Laryngologica, 107*(5-6), 460-466.

Keefe, D. H., Bulen, J. C., Arehart, K. H., & Burns, E. M. (1993). Ear-canal impedance and reflection coefficient in human infants and adults. *Journal of the Acoustical Society of America, 94*(5), 2617-2638.

Keefe, D. H., Bulen, J. C., Campbell, S. L., & Burns, E. M. (1994). Pressure transfer function and absorption cross section from the diffuse field to the human infant ear canal. *The Journal of the Acoustical Society of America, 95*(1), 355-371.

Keefe, D. H., & Feeney, M. P. (2009). Principles of acoustic immittance and acoustic transfer functions. In J. Katz, L. Medwetsky, R. Burkard, & L. J. Hood (Eds.), *Handbook of Clinical Audiology* (pp. 125-156). Baltimore, MD: Lippincott Williams & Wilkins.

Keefe, D. H., Fitzpatrick, D., Liu, Y. W., Sanford, C. A., & Gorga, M. P. (2010). Wideband acoustic-reflex test in a test battery to predict middle-ear dysfunction. *Hearing Research, 263*(1-2), 52-65.

Keefe, D. H., Folsom, R. C., Gorga, M. P., Vohr, B. R., Bulen, J. C., & Norton, S. J. (2000). Identification of neonatal hearing impairment: Ear-canal measurements of acoustic admittance and reflectance in neonates. *Ear and Hearing, 21*(5), 443-461.

Keefe, D. H., Gorga, M. P., Neely, S. T., Zhao, F., & Vohr, B. R. (2003a). Ear-canal acoustic admittance and reflectance measurements in human neonates. II. Predictions of middle-ear dysfunction and sensorineural hearing loss. *Journal of the Acoustical Society of America, 113*(1), 407-422.

Keefe, D. H., Hunter, L. L., Feeney, M. P., & Fitzpatrick, D. F. (2015). Procedures for ambient-pressure and tympanometric tests of aural acoustic reflectance and admittance in human infants and adults. *Journal of the Acoustical Society of America, 138*(6), 3625-3653.

Keefe, D. H., & Levi, E. (1996). Maturation of the middle and external ears: Acoustic power-based responses and reflectance tympanometry. *Ear and Hearing, 17*(5), 361-373.

Keefe, D. H., Ling, R., & Bulen, J. C. (1992). Method to measure acoustic impedance and reflection coefficient. *The Journal of the Acoustical Society of America, 91*(1), 470-485.

Keefe, D. H., Sanford, C. A., Ellison, J. C., Fitzpatrick, D. F., & Gorga, M. P. (2012). Wideband aural acoustic absorbance predicts conductive hearing loss in children. *International Journal of Audiology, 51*(12), 880-891.

Keefe, D. H., & Simmons, J. L. (2003). Energy transmittance predicts conductive hearing loss in older children and adults. *Journal of the Acoustical Society of America, 114*(6), 3217-3238.

Keefe, D. H., Zhao, F., Neely, S. T., Gorga, M. P., & Vohr, B. R. (2003b). Ear-canal acoustic admittance and reflectance effects in human neonates. I. Predictions of otoacoustic emission and auditory brainstem responses. *Journal of the Acoustical Society of America, 113*(1), 389-406.

Kei, J., Allison-Levick, J., Dockray, J., Harrys, R., Kirkegard, C., Wong, J., . . . Tudehope, D. (2003). High-frequency (1000 Hz) tympanometry in normal neonates. *Journal of the American Academy of Audiology, 14*(1), 20-28.

Kei, J., Brazel, B., Crebbin, K., Richards, A., & Willeston, N. (2007). High frequency distortion product otoacoustic emissions in children with and without middle ear dysfunction. *International Journal of Pediatric Otorhinolaryngology, 71*(1), 125-133.

Kei, J., & Mazlan, R. (2012a). Acoustic stapedial reflexes: Clinical applications. In J. Kei & F. Zhao (Eds.), *Assessing Middle Ear Function in Infants* (pp. 39-68). Abingdon, United Kingdom: Plural Publishing.

Kei, J., & Mazlan, R. (2012b). High-frequency (1000 Hz) tympanometry: Clinical applications. In J. Kei & F. Zhao (Eds.), *Assessing Middle Ear Function in Infants* (pp. 69-86). Abingdon, United Kingdom: Plural Publishing.

Kei, J., Mazlan, R., Hickson, L., Gavranich, J., & Linning, R. (2007). Measuring middle ear admittance in newborns using 1000 Hz tympanometry: A comparison of methodologies. *Journal of the American Academy of Audiology, 18*(9), 739-748.

Kei, J., Sanford, C. A., Prieve, B. A., & Hunter, L. L. (2013). Wideband acoustic immittance measures: Developmental characteristics (0 to 12 Months). *Ear and Hearing, 34*, 17s-26s.

Kei, J., & Zhao, F. (2012). *Assessing middle ear function in infants*. Abingdon, England: Plural Publishing.

Kenny, S. (2011). *Clinical application of the interacoustics REFLWIN system wideband reflectance machine in the assessment of the eustachian tube* (Unpublished master's thesis). University of British Columbia, Vancouver, Canada.

Keogh, T., Kei, J., Driscoll, C., Cahill, L., Hoffmann, A., Wilce, E., . . . Marinac, J. (2005). Measuring the ability of school children with a history of otitis media to understand everyday speech. *Journal of the American Academy of Audiology, 16*(5), 301-311.

Kero, P., & Piekkala, P. (1987). Factors affecting the occurrence of acute otitis media during the first year of life. *Acta Paediatrica, 76*(4), 618-623.

Klein, J. O. (2000). The burden of otitis media. *Vaccine, 19*, S2-S8.

Koike, K. J., & Wetmore, S. J. (1999). Interactive effects of the middle ear pathology and the associated hearing loss on transient-evoked otoacoustic emission measures. *Otolaryngology--Head and Neck Surgery, 121*(3), 238-244.

Koivunen, P., Uhari, M., Laitakari, K., Alho, O.-P., & Luotonen, J. (2000). Otoacoustic emissions and tympanometry in children with otitis media. *Ear and Hearing, 21*(3), 212-217.

Kok, M., Van Zanten, G., & Brocuar, M. (1992). Growth of evoked otoacoustic emissions during the first days postpartum: A preliminary report. *Audiology, 31*(3), 140-149.

Kong, K., & Coates, H. L. (2009). Natural history, definitions, risk factors and burden of otitis media. *Medical Journal of Australia, 191*(9), S39.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTEST package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13).

Kværner, K. J., Nafstad, P., Hagen, J. A., Mair, I. W., & Jaakkola, J. J. (1997). Recurrent acute otitis media: The significance of age at onset. *Acta Oto-Laryngologica, 117*(4), 578-584.

Leach, A. J. (1999). Otitis media in Australian Aboriginal children: An overview. *International Journal of Pediatric Otorhinolaryngology, 49*, S173-S178.

Lee, D. H. (2010). How to improve the accuracy of diagnosing otitis media with effusion in a pediatric population. *International Journal of Pediatric Otorhinolaryngology, 74*(2), 151-153.

Lee, H., Kim, J., & Nguyen, V. (2013). Ear infections: Otitis externa and otitis media. *Primary Care: Clinics in Office Practice, 40*(3), 671-686.

Lehmann, D., Weeks, S., Jacoby, P., Elsbury, D., Finucane, J., Stokes, A., . . . the Kalgoorlie Otitis Media Research Project Team. (2008). Absent otoacoustic emissions predict otitis media in young Aboriginal children: A birth cohort study in Aboriginal and non-Aboriginal children in an arid zone of Western Australia. *BMC Pediatrics, 8*(1), 32.

Lenth, R. (2018). emmeans: Estimated Marginal Means, aka Least-Squares Means [R package] (Version 1.2.1). Retrieved from https://CRAN.R-project.org/package=emmeans

Lidén, G. (1969). The scope and application of current audiometric tests. *The Journal of Laryngology & Otology, 83*(06), 507-520.

Lieberthal, A. S., Carroll, A. E., Chonmaitree, T., Ganiats, T. G., Hoberman, A., Jackson, M. A., . . . Sevilla, X. D. (2013). The diagnosis and management of acute otitis media. *Pediatrics, 131*(3), e964-e999.

Lilly, D. J., & Margolis, R. H. (2013). Wideband acoustic immittance measurements of the middle ear: Introduction and some historical antecedents. *Ear and Hearing, 34*, 4s-8s.

Liu, Y. W., Sanford, C. A., Ellison, J. C., Fitzpatrick, D. F., Gorga, M. P., & Keefe, D. H. (2008). Wideband absorbance tympanometry using pressure sweeps: System development and results on adults with normal hearing. *Journal of the Acoustical Society of America, 124*(6), 3708-3719.

MacIntyre, E. A., Karr, C. J., Koehoorn, M., Demers, P., Tamburic, L., Lencar, C., & Brauer, M. (2010). Otitis media incidence and risk factors in a population-based birth cohort. *Paediatrics & Child Health, 15*(7), 437.

Madell, J. R. (2014). Using visual reinforcement audiomety to evaluate hearing in infants frum 5 to 36 months. In J. R. Madell & C. Flexer (Eds.), *Pediatric Audiology: Diagnosis, Technology and Management* (pp. 79-88). New York, NY: Thieme.

Mahadevan, M., Navarro-Locsin, G., Tan, H., Yamanaka, N., Sonsuwan, N., Wang, P.-C., . . . Vijayasekaran, S. (2012). A review of the burden of disease due to otitis media in the Asia-Pacific. *International Journal of Pediatric Otorhinolaryngology, 76*(5), 623-635.

Marchant, C. D., McMillan, P. M., Shurin, P. A., Johnson, C. E., Turczyk, V. A., Feinstein, J. C., & Panek, D. M. (1986). Objective diagnosis of otitis media in early infancy by tympanometry and ipsilateral acoustic reflex thresholds. *Journal of Pediatrics, 109*(4), 590-595.

Marchant, C. D., Shurin, P. A., Turczyk, V. A., Wasikowski, D. E., Tutihasi, M. A., & Kinney, S. E. (1984a). Course and outcome of otitis media in early infancy: A prospective study. *The Journal of pediatrics, 104*(6), 826-831.

Marchant, C. D., Shurin, P. A., Turczyk, V. A., Wasikowski, D. E., Tutihasi, M. A., & Kinney, S. E. (1984b). Course and outcome of otitis media in early infancy: A prospective study. *Journal of Pediatrics, 104*(6), 826-831.

Marcus, G. (2018). Deep learning: A critical appraisal. *arXiv* 2-27.

Margolis, R. H., Bass-Ringdahl, S., Hanks, W. D., Holte, L., & Zapala, D. A. (2003). Tympanometry in newborn infants—1 kHz norms. *Journal of the American Academy of Audiology, 14*(7), 383-392.

Margolis, R. H., Rykken, J. R., Hunter, L. L., & Giebink, G. S. (1993). Effects of otitis media on extended high-frequency hearing in children. *Annals of Otology, Rhinology and Laryngology, 102*(1), 1-5.

Margolis, R. H., Saly, G. L., & Hunter, L. L. (2000). High-frequency hearing loss and wideband middle ear impedance in children with otitis media histories. *Ear and Hearing, 21*(3), 206-211.

Marom, T., Tan, A., Wilkinson, G. S., Pierson, K. S., Freeman, J. L., & Chonmaitree, T. (2014). Trends in otitis media-related health care use in the United States, 2001-2011. *JAMA pediatrics, 168*(1), 68-75.

Mazlan, R., & Kei, J. (2012). Measuring middle ear function in infants: An overview. In J. Kei & F. Zhao (Eds.), *Assessing Middle Ear Function in Infants* (pp. 69-86). Abingdon, United Kingdom: Plural Publishing.

Mazlan, R., Kei, J., & Hickson, L. (2009a). Test-retest reliability of the acoustic stapedial reflex test in healthy neonates. *Ear and Hearing, 30*(3), 295-301.

Mazlan, R., Kei, J., Hickson, L., Khan, A., Gavranich, J., & Linning, R. (2009b). High frequency (1000 Hz) tympanometry findings in newborns: Normative data using a component compensated admittance approach.

Merchant, G. R., Horton, N. J., & Voss, S. E. (2010). Normative reflectance and transmittance measurements on healthy newborn and 1-month-old infants. *Ear and Hearing, 31*(6), 746-754.

Metz, C. E. (1978). Basic principles of ROC analysis. *Seminars in Nuclear Medicine, 8*(4), 283-298.

Monasta, L., Ronfani, L., Marchetti, F., Montico, M., Brumatti, L. V., Bavcar, A., . . . Tamburlini, G. (2012). Burden of disease caused by otitis media: Systematic review and global estimates. *PloS One, 7*(4), e36226.

Moons, K. G., Donders, A. R., Steyerberg, E. W., & Harrell, F. E. (2004). Penalized maximum likelihood estimation to directly adjust diagnostic and prognostic prediction models for overoptimism: A clinical example. *Journal of Clinical Epidemiology, 57*(12), 1262-1270.

Moons, K. G., Kengne, A. P., Grobbee, D. E., Royston, P., Vergouwe, Y., Altman, D. G., & Woodward, M. (2012a). Risk prediction models: II. External validation, model updating, and impact assessment. *Heart, 98*(9), 691.

Moons, K. G., Kengne, A. P., Woodward, M., Royston, P., Vergouwe, Y., Altman, D. G., & Grobbee, D. E. (2012b). Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio) marker. *Heart, 98*(9), 683-690.

Moore, D. R., Hartley, D. E., & Hogan, S. C. (2003). Effects of otitis media with effusion (OME) on central auditory function. *International Journal of Pediatric Otorhinolaryngology, 67 Suppl 1*, S63-67.

Morris, P. S., Leach, A. J., Silberberg, P., Mellon, G., Wilson, C., Hamilton, E., & Beissbarth, J. (2005). Otitis media in young Aboriginal children from remote communities in Northern and Central Australia: A cross-sectional survey. *BMC Pediatrics, 5*(1), 27.

Morris, P. S., Richmond, P., Lehmann, D., Leach, A. J., Gunasekera, H., & Coates, H. L. (2009). New horizons: Otitis media research in Australia. *Medical Journal of Australia, 191*(9), S73.

Myers, J., Kei, J., Aithal, S., Aithal, V., Driscoll, C., Khan, A., . . . Malicka, A. N. (2018a). Development of a diagnostic prediction model for conductive conditions in neonates using wideband acoustic immittance. *Ear and Hearing, 39*(6), 1116-1135.

Myers, J., Kei, J., Aithal, S., Aithal, V., Driscoll, C., Khan, A., . . . Malicka, A. N. (2018b). Diagnosing middle ear pathology in 6- to 9-month-old infants using wideband absorbance: A risk prediction model. *Journal of Speech Language and Hearing Research, 61*(9), 2386-2404.

Myers, J., Kei, J., Aithal, S., Aithal, V., Driscoll, C., Khan, A., . . . Malicka, A. N. (2019a). Diagnosing conductive dysfunction in infants using wideband acoustic immittance: Validation and development of predictive models. *Journal of Speech Language and Hearing Research, 62*(9), 3607-3619.

Myers, J., Kei, J., Aithal, S., Aithal, V., Driscoll, C., Khan, A., . . . Malicka, A. N. (2019b). Diagnosing middle ear dysfunction in 10- to 16-month old infants using wideband absorbance: An ordinal prediction model. *Journal of Speech Language and Hearing Research, 62*(8), 2906-2917.

Myers, J., Kei, J., Aithal, S., Aithal, V., Driscoll, C., Khan, A., . . . Malicka, A. N. (2019c). Longitudinal development of wideband absorbance and admittance through infancy. *Journal of Speech Language and Hearing Research, 62*(7), 2535-2552.

Naaktgeboren, C. A., Bertens, L. C., van Smeden, M., de Groot, J. A., Moons, K. G., & Reitsma, J. B. (2013). Value of composite reference standards in diagnostic research. *BMJ, 347*, f5605.

Naeve, S. L., Margolis, R. H., Levine, S. C., & Fournier, E. M. (1992). Effect of ear-canal air pressure on evoked otoacoustic emissions. *The Journal of the Acoustical Society of America, 91*(4), 2091-2095.

Nakajima, H. H., Pisano, D. V., Roosli, C., Hamade, M. A., Merchant, G. R., Mahfoud, L., . . . Merchant, S. N. (2012). Comparison of ear-canal reflectance and umbo velocity in patients with conductive hearing loss: a preliminary study. *Ear and Hearing, 33*(1), 35.

Nakajima, H. H., Rosowski, J. J., Shahnaz, N., & Voss, S. E. (2013). Assessment of ear disorders using power reflectance. *Ear and Hearing, 34*(7 0 1), 48s.

Nicholls, K. (2016). *The Healthy Hearing Program Audiology Diagnostic Assessment Protocol*. Children's Health Queensland. Retrieved from https://www.childrens.health.qld.gov.au/wp-content/uploads/PDF/healthy-hearing/hh-audiology-protocol.pdf

Northern, J. L., & Downs, M. P. (2014). *Hearing in Children*. San Diego, CA: Plural Publishing.

Northrop, C., Piza, J., & Eavey, R. D. (1986). Histological observations of amniotic fluid cellular content in the ear of neonates and infants. *International Journal of Pediatric Otorhinolaryngology, 11*(2), 113-127.

Nozza, R. J., Bluestone, C. D., Kardatzke, D., & Bachman, R. (1992). Towards the validation of aural acoustic immittance measures for diagnosis of middle ear effusion in children. *Ear and Hearing, 13*(6), 442-453.

Nozza, R. J., Bluestone, C. D., Kardatzke, D., & Bachman, R. (1994). Identification of middle ear effusion by aural acoustic admittance and otoscopy. *Ear and Hearing, 15*(4), 310-323.

Olmsted, R. W., Alvarez, M. C., Moroney, J. D., & Eversden, M. (1964). The pattern of hearing following acute otitis media. *The Journal of Pediatrics, 65*(2), 252-255.

Osterhammel, P. A., Nielsen, L. H., & Rasmussen, A. N. (1993). Distortion product otoacoustic emissions: The influence of the middle ear transmission. *Scandinavian Audiology, 22*(2), 111-116.

Palmu, A., Puhakka, H., Rahko, T., & Takala, A. K. (1999). Diagnostic value of tympanometry in infants in clinical practice. *International Journal of Pediatric Otorhinolaryngology, 49*(3), 207-213.

Palmu, A., & Syrjänen, R. (2005). Diagnostic value of tympanometry using subject-specific normative values. *International Journal of Pediatric Otorhinolaryngology, 69*(7), 965-971.

Palva, T., Johnsson, L.-G., & Ramsay, H. (2000). Attic aeration in temporal bones from children with recurring otitis media: Tympanostomy tubes did not cure disease in Prussak's space. *Otology & Neurotology, 21*(4), 485-493.

Palva, T., Northrop, C., & Ramsay, H. (2001). Foreign body neonatal otitis media in infants. *Otology & Neurotology, 22*(4), 433-443.

Paradise, J. L., Feldman, H. M., Colborn, D. K., Campbell, T. F., Dollaghan, C. A., Rockette, H. E., . . . Smith, C. G. (1999). Parental stress and parent-rated child behavior in relation to otitis media in the first three years of life. *Pediatrics, 104*(6), 1264-1273.

Paradise, J. L., Rockette, H. E., Colborn, D. K., Bernard, B. S., Smith, C. G., Kurs-Lasky, M., & Janosky, J. E. (1997). Otitis media in 2253 Pittsburgh-area infants: Prevalence and risk factors during the first two years of life. *Pediatrics, 99*(3), 318-333.

Paradise, J. L., Smith, C. G., & Bluestone, C. D. (1976). Tympanometric detection of middle ear effusion in infants and young children. *Pediatrics, 58*(2), 198-210.

Pelton, S. I., & Leibovitz, E. (2009). Recent advances in otitis media. *The Pediatric Infectious Disease Journal, 28*(10), S133-S137.

Penido, N. D. O., Borin, A., Iha, L. C., Suguri, V. M., Onishi, E., Fukuda, Y., & Cruz, O. L. M. (2005). Intracranial complications of otitis media: 15 years of experience in 33 patients. *Otolaryngology-head and neck surgery, 132*(1), 37-42.

Pepe, M. S. (2003). *The Statistical Evaluation of Medical Tests for Classification and Prediction*. Oxford, England: Oxford University Press.

Pereira, P. K. S., Azevedo, M. F. d., & Testa, J. R. (2010). Conductive impairment in newborn who failed the newborn hearing screening. *Brazilian Journal of Otorhinolaryngology, 76*(3), 347-354.

Pichichero, M. E. (2013). Otitis media. *Pediatric Clinics of North America, 60*(2), 391-407.

Pichichero, M. E., & Poole, M. D. (2001). Assessing diagnostic accuracy and tympanocentesis skills in the management of otitis media. *Archives of Pediatrics and Adolescent Medicine, 155*(10), 1137-1142.

Pillsbury, H. C., Grose, J. H., & Hall, J. W., 3rd. (1991). Otitis media with effusion in children. Binaural hearing before and after corrective surgery. *Archives of Otolaryngology - Head and Neck Surgery, 117*(7), 718-723.

Piskorski, P., Keefe, D. H., Simmons, J. L., & Gorga, M. P. (1999). Prediction of conductive hearing loss based on acoustic ear-canal response using a multivariate clinical decision theory. *Journal of the Acoustical Society of America, 105*(3), 1749-1764.

Pitaro, J., Al Masaoudi, L., Motallebzadeh, H., Funnell, W. R. J., & Daniel, S. J. (2016). Wideband reflectance measurements in newborns: Relationship to otoscopic findings. *International Journal of Pediatric Otorhinolaryngology, 86*, 156-160.

Piza, J., Northrop, C., & Eavey, R. D. (1998). Embryonic middle ear mesenchyme disappears by redistribution. *The Laryngoscope, 108*(9), 1378-1381.

Plinkert, P., Bootz, F., & Vossieck, T. (1994). Influence of static middle ear pressure on transiently evoked otoacoustic emissions and distortion products. *European Archives of Oto-Rhino-Laryngology, 251*(2), 95-99.

Plinkert, P., & Plok, M. (1994). Influence of eustachian tube dysfunction on transiently evoked and distortion product otoacoustic emissions. *HNO, 42*(7), 434-434.

Prieve, B. A., Calandruccio, L., Fitzgerald, T., Mazevski, A., & Georgantas, L. M. (2008). Changes in transient-evoked otoacoustic emission levels with negative tympanometric peak pressure in infants and toddlers. *Ear and Hearing, 29*(4), 533-542.

Prieve, B. A., Feeney, M. P., Stenfelt, S., & Shahnaz, N. (2013a). Prediction of conductive hearing loss using wideband acoustic immittance. *Ear and Hearing, 34*, 54s-59s.

Prieve, B. A., & Fitzgerald, T. (2015). Otoacoustic emissions. In J. Katz, M. Chasin, K. English, L. J. Hood, & K. L. Tillery (Eds.), *Handbook of Clinical Audiology* (pp. 357-379). Philadelphia, PA: Wolters Kluwer.

Prieve, B. A., Vander Werff, K. R., Preston, J. L., & Georgantas, L. (2013b). Identification of conductive hearing loss in young infants using tympanometry and wideband reflectance. *Ear and Hearing, 34*(2), 168-178.

Priner, R., Freeman, S., Perez, R., & Sohmer, H. (2003). The neonate has a temporary conductive hearing loss due to fluid in the middle ear. *Audiology & Neurotology, 8*(2), 100-110.

Pukander, J., Luotonen, J., Sipilau, M., Timonen, M., & Karma, P. (1982). Incidence of acute otitis media. *Acta Oto-Laryngologica, 93*(1-6), 447-453.

Qureishi, A., Lee, Y., Belfield, K., Birchall, J. P., & Daniel, M. (2014). Update on otitis media-prevention and treatment. *Infection and drug resistance, 7*, 15.

R Core Team. (2017). R: A Language and Environment for Statistical Computing [Computer software]. (Version 3.6.0). Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Roberts, D. G., Johnson, C. E., Carlin, S. A., Turczyk, V., Karnuta, M. A., & Yaffee, K. (1995). Resolution of middle ear effusion in newborns. *Archives of Pediatrics and Adolescent Medicine, 149*(8), 873-877.

Roberts, J. E., Hunter, L., Gravel, J., Rosenfeld, R., Berman, S., Haggard, M., . . . Wallace, I. (2004a). Otitis media, hearing loss, and language learning: Controversies and current research. *Journal of Developmental and Behavioral Pediatrics, 25*(2), 110-122.

Roberts, J. E., Rosenfeld, R. M., & Zeisel, S. A. (2004b). Otitis media and speech and language: A meta-analysis of prospective studies. *Pediatrics, 113*(3), e238-e248.

Robinson, S. R., Thompson, S., & Allen, J. B. (2016). Effects of negative middle ear pressure on wideband acoustic immittance in normal-hearing adults. *Ear and Hearing, 37*(4), 452-464.

Rogers, D. J., Boseley, M. E., Adams, M. T., Makowski, R. L., & Hohman, M. H. (2010). Prospective comparison of handheld pneumatic otoscopy, binocular microscopy, and tympanometry in identifying middle ear effusions in children. *International Journal of Pediatric Otorhinolaryngology, 74*(10), 1140-1143.

Rosenfeld, R. M., Culpepper, L., Doyle, K. J., Grundfast, K. M., Hoberman, A., Kenna, M. A., . . . American Academy of Otolaryngology-Head and Neck Surgery. (2004). Clinical practice guideline: Otitis media with effusion. *Otolaryngology-Head and Neck Surgery, 130*(5), S95-S118.

Rosowski, J. J., Stenfelt, S., & Lilly, D. (2013). An overview of wideband immittance measurements techniques and terminology: You say absorbance, I say reflectance. *Ear and Hearing, 34*, 9s-16s.

Roush, J., Bryant, K., Mundy, M., Zeisel, S., & Roberts, J. (1995). Developmental changes in static admittance and tympanometric width in infants and toddlers. *Journal of the American Academy of Audiology, 6*(4), 334-338.

Rovers, M. M. (2008). The burden of otitis media. *Vaccine, 26*, G2-G4.

Rovers, M. M., Glasziou, P., Appelman, C. L., Burke, P., McCormick, D. P., Damoiseaux, R. A., . . . Hoes, A. W. (2007). Predictors of pain and/or fever at 3 to 7 days for children with acute otitis media not treated initially with antibiotics: A meta-analysis of individual patient data. *Pediatrics, 119*(3), 579-585.

Ruben, R. J. (1997). A time frame of critical/sensitive periods of language development. *Acta Oto-Laryngologica, 117*(2), 202-205.

Rutjes, A. W., Reitsma, J. B., Coomarasamy, A., Khan, K. S., & Bossuyt, P. M. (2007). Evaluation of diagnostic tests when there is no gold standard. A review of methods. *Health Technology Assessment, 11*(50), 1-50.

Sakran, W., Makary, H., Colodner, R., Ashkenazi, D., Rakover, Y., Halevy, R., & Koren, A. (2006). Acute otitis media in infants less than three months of age: Clinical presentation, etiology and concomitant diseases. *International Journal of Pediatric Otorhinolaryngology, 70*(4), 613-617.

Sanford, C. A., & Brockett, J. E. (2014). Characteristics of wideband acoustic immittance in patients with middle-ear dysfunction. *Journal of the American Academy of Audiology, 25*(5), 425-440.

Sanford, C. A., & Feeney, M. P. (2008). Effects of maturation on tympanometric wideband acoustic transfer functions in human infants. *The Journal of the Acoustical Society of America, 124*(4), 2106-2122.

Sanford, C. A., Keefe, D. H., Liu, Y.-W., Fitzpatrick, D., McCreery, R. W., Lewis, D. E., & Gorga, M. P. (2009). Sound-conduction effects on distortion-product otoacoustic emission screening outcomes in newborn infants: Test performance of wideband acoustic transfer functions and 1-kHz tympanometry. *Ear and Hearing, 30*(6), 635-652.

Sanford, C. A., Keefe, D. H., Liu, Y.-W., Fitzpatrick, D., McCreery, R. W., Lewis, D. E., & Gorga, M. P. (2014). Sound-conduction effects on distortion-product otoacoustic emission screening outcomes in newborn infants: Test performance of wideband acoustic transfer functions and 1-kHz tympanometry: Erratum. *Ear and Hearing, 35*(2), 288.

Sassen, M. L., Aarem, A., & Grote, J. J. (1994). Validity of tympanometry in the diagnosis of middle ear effusion. *Clinical Otolaryngology and Allied Sciences, 19*(3), 185-189.

Schmidt, R. L., & Factor, R. E. (2013). Understanding sources of bias in diagnostic accuracy studies. *Archives of Pathology and Laboratory Medicine, 137*(4), 558-565.

Serbetcioglu, B., Ugurtay, O., Kirkim, G., & Mutlu, B. (2008). No association between hearing loss due to bilateral otitis media with effusion and Denver-II test results in preschool children. *International Journal of Pediatric Otorhinolaryngology, 72*(2), 215-222.

Shahnaz, N. (2008). Wideband reflectance in neonatal intensive care units. *Journal of the American Academy of Audiology, 19*(5), 419.

Shahnaz, N., & Bork, K. (2006). Wideband reflectance norms for Caucasian and Chinese young adults. *Ear and Hearing, 27*(6), 774-788.

Shahnaz, N., Cai, A., & Qi, L. (2014). Understanding the developmental course of the acoustic properties of the human outer and middle ear over the first 6 months of life by using a longitudinal analysis of power reflectance at ambient pressure. *Journal of the American Academy of Audiology, 25*(5), 495-511.

Shahnaz, N., Feeney, M. P., & Schairer, K. S. (2013). Wideband acoustic immittance normative data: Ethnicity, gender, aging, and instrumentation. *Ear and Hearing, 34*, 27s-35s.

Shahnaz, N., Longridge, N., & Bell, D. (2009). Wideband energy reflectance patterns in preoperative and post-operative otosclerotic ears. *International Journal of Audiology, 48*(5), 240-247.

Shanks, J., & Shohet, J. (2009). Tympanometry in clinical practice. In J. Katz, L. Medwetsky, R. Burkard, & L. J. Hood (Eds.), *Handbook of Clinical Audiology* (pp. 157-188). Baltimore, Maryland: Lippincott Williams & Wilkins.

Shaver, M. D., & Sun, X.-M. (2013). Wideband energy reflectance measurements: Effects of negative middle ear pressure and application of a pressure compensation procedure. *The Journal of the Acoustical Society of America, 134*(1), 332-341.

Shaw, J. (2009). *Comparison of wideband energy reflectance and tympanometric measures obtained with Reflwin Interacoustics, Mimosa Acoustics and GSI Tympstar systems.*

Shurin, P. A., Pelton, S. I., Donner, A., & Klein, J. O. (1979). Persistence of middle-ear effusion after acute otitis media in children. *New England Journal of Medicine, 300*(20), 1121-1123.

Smith, C. G., Paradise, J. L., Sabo, D. L., Rockette, H. E., Kurs-Lasky, M., Bernard, B. S., & Colborn, D. K. (2006). Tympanometric findings and the probability of middle-ear effusion in 3686 infants and young children. *Pediatrics, 118*(1), 1-13.

Smolkin, T., Mick, O., Dabbah, M., Blazer, S., Grakovsky, G., Gabay, N., . . . Makhoul, I. R. (2012). Birth by cesarean delivery and failure on first otoacoustic emissions hearing test. *Pediatrics, 130*(1), e95-100.

Sorrento, A., & Pichichero, M. E. (2001). Assessing diagnostic accuracy and tympanocentesis skills by nurse practitioners in management of otitis media. *Journal of the American Academy of Nurse Practitioners, 13*(11), 524-529.

Sox, H. C., Higgins, M. C., & Owens, D. K. (2013). *Medical decision making* (2nd ed.). Hoboken, NJ: John Wiley & Sons.

Steyerberg, E. W. (2008). *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. New York, NY: Springer.

Steyerberg, E. W., Eijkemans, M. J., Harrell, F. E., Jr., & Habbema, J. D. (2001a). Prognostic modeling with logistic regression analysis: In search of a sensible strategy in small data sets. *Medical Decision Making, 21*(1), 45-56.

Steyerberg, E. W., Eijkemans, M. J. C., & Habbema, J. D. F. (2001b). Application of shrinkage techniques in logistic regression analysis: a case study. *Statistica Neerlandica, 55*(1), 76-88.

Steyerberg, E. W., Harrell, F. E., Borsboom, G. J., Eijkemans, M., Vergouwe, Y., & Habbema, J. D. F. (2001c). Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis. *Journal of Clinical Epidemiology, 54*(8), 774-781.

Steyerberg, E. W., Van Calster, B., & Pencina, M. J. (2011). Performance measures for prediction models and markers: Evaluation of predictions and classifications. *Revista Española de Cardiología (English Edition), 64*(9), 788-794.

Steyerberg, E. W., Vickers, A. J., Cook, N. R., Gerds, T., Gonen, M., Obuchowski, N., . . . Kattan, M. W. (2010). Assessing the performance of prediction models: A framework for some traditional and novel measures. *Epidemiology 21*(1), 128.

Stinson, M. R., Shaw, E. A. G., & Lawton, B. W. (1982). Estimation of acoustical energy reflectance at the eardrum from measurements of pressure distribution in the human ear canal. *The Journal of the Acoustical Society of America, 72*(3), 766-773.

Stuart, A., Yang, E., & Green, W. (1994). Neonatal auditory brainstem response thresholds to air-and bone-conducted clicks: 0 to 96 hours postpartum. *JOURNAL-AMERICAN ACADEMY OF AUDIOLOGY, 5*, 163-163.

Swanepoel, D. W., Werner, S., Hugo, R., Louw, B., Owen, R., & Swanepoel, A. (2007). High frequency immittance for neonates: A normative study. *Acta Oto-Laryngologica, 127*(1), 49-56.

Syggelou, A., Fanos, V., & Iacovidou, N. (2011). Acute otitis media in neonatal life: A review. *Journal of Chemotherapy, 23*(3), 123-126.

Tähtinen, P. A., Laine, M. K., Huovinen, P., Jalava, J., Ruuskanen, O., & Ruohola, A. (2011). A placebo-controlled trial of antimicrobial treatment for acute otitis media. *New England Journal of Medicine, 364*(2), 116-126.

Takata, G. S., Chan, L. S., Morphew, T., Mangione-Smith, R., Morton, S. C., & Shekelle, P. (2003). Evidence assessment of the accuracy of methods of diagnosing middle ear effusion in children with otitis media with effusion. *Pediatrics, 112*(6), 1379-1387.

Tas, A., Yagiz, R., Uzun, C., Adali, M. K., Koten, M., Tas, M., & Karasalihoglu, A. R. (2004). Effect of middle ear effusion on distortion product otoacoustic emission. *International Journal of Pediatric Otorhinolaryngology, 68*(4), 437-440.

Taylor, S., Marchisio, P., Vergison, A., Harriague, J., Hausdorff, W. P., & Haggard, M. (2012). Impact of pneumococcal conjugate vaccination on otitis media: A systematic review. *Clinical Infectious Diseases, 54*(12), 1765-1773.

Teele, D. W., Klein, J. O., Rosner, B., & the Greater Boston Otitis Media Study Group. (1989). Epidemiology of otitis media during the first seven years of life in children in greater Boston: A prospective, cohort study. *Journal of Infectious Diseases, 160*(1), 83-94.

Thakur, J., Chauhan, I., Mohindroo, N., Sharma, D., Azad, R., & Vasanthalakshmi, M. (2013). Otoacoustic emissions in otitis media with effusion: Do they carry any clinical significance? *Indian Journal of Otolaryngology and Head & Neck Surgery, 65*(1), 29-33.

Thomas, N. M., & Brook, I. (2014). Otitis media: An update on current pharmacotherapy and future perspectives. *Expert Opinion on Pharmacotherapy, 15*(8), 1069-1083.

Thompson, D. C., McPhillips, H., Davis, R. L., Lieu, T. L., Homer, C. J., & Helfand, M. (2001). Universal newborn hearing screening: Summary of evidence. *JAMA, 286*(16), 2000-2010.

Thornton, A. R., Kimm, L., Kennedy, C. R., & Cafarelli-Dees, D. (1993). External- and middle-ear factors affecting evoked otoacoustic emissions in neonates. *British Journal of Audiology, 27*(5), 319-327.

Todberg, T., Koch, A., Andersson, M., Olsen, S. F., Lous, J., & Homøe, P. (2014). Incidence of otitis media in a contemporary Danish national birth cohort. *PloS One, 9*(12), e111732.

Tomlin, D., & Rance, G. (2014). Long-term hearing deficits after childhood middle ear disease. *Ear and Hearing, 35*(6), e233-e242.

Topolska, M., Hassman, E., & Baczek, M. (2000). The effects of chronic otitis media with effusion on the measurement of distortion products of otoacoustic emissions: Presurgical and postsurgical examination. *Clinical Otolaryngology and Allied Sciences, 25*(4), 315-320.

Trine, M. B., Hirsch, J. E., & Margolis, R. H. (1993). The effect of middle ear pressure on transient evoked otoacoustic emissions. *Ear and Hearing, 14*(6), 401-407.

Tucci, D. L., Cant, N. B., & Durham, D. (1999). Conductive hearing loss results in a decrease in central auditory system activity in the young gerbil. *The Laryngoscope, 109*(9), 1359-1371.

Turner, D., Leibovitz, E., Aran, A., Piglansky, L., Raiz, S., Leiberman, A., & Dagan, R. (2002). Acute otitis media in infants younger than two months of age: Microbiology, clinical presentation and therapeutic approach. *The Pediatric Infectious Disease Journal, 21*(7), 669-674.

Turner, R. G., Frazer, G. J., & Shepard, N. T. (1984). Formulating and evaluating audiological test protocols. *Ear and Hearing, 5*(6), 321-330.

Ueda, H., Nakata, S., & Hoshino, M. (1998). Effects of effusion in the middle ear and perforation of the tympanic membrane on otoacoustic emissions in guinea pigs. *Hearing Research, 122*(1), 41-46.

Van Calster, B., Vickers, A. J., Pencina, M. J., Baker, S. G., Timmerman, D., & Steyerberg, E. W. (2013). Evaluation of markers and risk prediction models: Overview of relationships between NRI and decision-analytic measures. *Medical Decision Making, 33*(4), 490-501.

Van Camp, K. J., Margolis, R., Wilson, R., Creten, W., & Shanks, J. (1986). Principles of tympanometry. *ASHA Monographs*(24), 1.

Vander Werff, K. R., Prieve, B. A., & Georgantas, L. M. (2007). Test-retest reliability of wideband reflectance measures in infants under screening and diagnostic test conditions. *Ear and Hearing, 28*(5), 669-681.

Vander Werff, K. R., Prieve, B. A., & Georgantas, L. M. (2009). Infant air and bone conduction tone burst auditory brain stem responses for classification of hearing loss and the relationship to behavioral thresholds. *Ear and Hearing, 30*(3), 350-368.

Venekamp, R. P., Sanders, S., Glasziou, P., del Mar, C., & Rovers, M. (2015). Antibiotics for acute otitis media in children (Review). *Cochrane Database of Systematic Reviews, 2015*(6), 1-83.

Vickers, A. J. (2008). Decision analysis for the evaluation of diagnostic tests, prediction models and molecular markers. *The American Statistician, 62*(4), 314-320.

Vickers, A. J. (2011). Prediction models in cancer care. *CA: A Cancer Journal for Clinicians, 61*(5), 315-326.

Vickers, A. J., Basch, E., & Kattan, M. W. (2008). Against diagnosis. *Annals of Internal Medicine, 149*(3), 200-203.

Villa, P. C., & Zanchetta, S. (2014). Auditory temporal abilities in children with history of recurrent otitis media in the first years of life and persistent in preschool and school ages. *CoDAS, 26*(6), 494-502.

Voss, S. E., Herrmann, B. S., Horton, N. J., Amadei, E. A., & Kujawa, S. G. (2016). Reflectance measures from infant ears with normal hearing and transient conductive hearing loss. *Ear and Hearing, 37*(5), 560-571.

Voss, S. E., Horton, N. J., Woodbury, R. R., & Sheffield, K. N. (2008). Sources of variability in reflectance measurements on normal cadaver ears. *Ear and Hearing, 29*(4), 651-665.

Voss, S. E., Merchant, G. R., & Horton, N. J. (2012). Effects of middle-ear disorders on power reflectance measured in cadaveric ear canals. *Ear and Hearing, 33*(2), 195.

Voss, S. E., Stenfelt, S., Neely, S. T., & Rosowski, J. J. (2013). Factors that introduce intrasubject variability into ear-canal absorbance measurements. *Ear and Hearing, 34*, 60s-64s.

Weatherby, L., & Bennett, M. (1980). The neonatal acoustic reflex. *Scandinavian Audiology, 9*(2), 103-110.

Webster, D. B. (1984). A critical period during postnatal auditory development of mice. *International Journal of Pediatric Otorhinolaryngology, 6*, 107-118.

Werner, L. A., Levi, E. C., & Keefe, D. H. (2010). Ear-canal wideband acoustic transfer functions of adults and two- to nine-month-old infants. *Ear and Hearing, 31*(5), 587-598.

Werner, Y. L., & Igic, P. G. (2002). The middle ear of gekkonoid lizards: Interspecific variation of structure in relation to body size and to auditory sensitivity. *Hearing Research, 167*(1-2), 33-45.

Werner, Y. L., Montgomery, L. G., Safford, S. D., Igic, P. G., & Saunders, J. C. (1998). How body size affects middle-ear structure and function and auditory sensitivity in gekkonoid lizards. *Journal of Experimental Biology, 201*(4), 487-502.

Whitton, J. P., & Polley, D. B. (2011). Evaluating the perceptual and pathophysiological consequences of auditory deprivation in early postnatal life: A comparison of basic and clinical studies. *Journal of the Association for Research in Otolaryngology, 12*(5), 535-547.

Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis*. New York: Springer-Verlag.

Williams, C. J., & Jacobs, A. M. (2009). The impact of otitis media on cognitive and educational outcomes. *Medical Journal of Australia, 191*(9), S69.

Wilson, W. (2012). Anatomy and physiology of the outer and middle ear in young infants. In J. Kei & F. Zhao (Eds.), *Assessing Middle Ear Function in Infants* (pp. 1-16). Abingdon, United Kingdom: Plural Publishing.

Winskel, H. (2006). The effects of an early history of otitis media on children's language and literacy skill development. *British Journal of Educational Psychology, 76*(4), 727-744.

Wolff, D. (1934). XV. Significant anatomic features of the auditory mechanism with special reference to the late fetus. *Annals of Otology, Rhinology and Laryngology, 43*(1), 193-247.

Wright, C. G. (1997). Development of the human external ear. *Journal of the American Academy of Audiology, 8*, 379-382.

Wright, P. F., McConnell, K. B., Thompson, J. M., Vaughn, W. K., & Sell, S. H. (1985). A longitudinal study of the detection of otitis media in the first two years of life. *International Journal of Pediatric Otorhinolaryngology, 10*(3), 245-252.

Yeo, S. W., Park, S.-N., Park, Y. S., & Suh, B. D. (2002). Effect of middle-ear effusion on otoacoustic emissions. *The Journal of Laryngology & Otology, 116*(10), 794-799.

Yiengprugsawan, V., Hogan, A., & Strazdins, L. (2013). Longitudinal analysis of ear infection and hearing impairment: Findings from 6-year prospective cohorts of Australian children. *BMC Pediatrics, 13*(1), 28.

Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). *How transferable are features in deep neural networks?* Paper presented at the Advances in Neural Information Processing Systems.

Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer, 3*(1), 32-35.

Young, D. E., Ten Cate, W. J. F., Ahmad, Z., & Morton, R. P. (2009). The accuracy of otomicroscopy for the diagnosis of paediatric middle ear effusions. *International Journal of Pediatric Otorhinolaryngology, 73*(6), 825-828.

Zhao, F., Wada, H., Koike, T., Ohyama, K., Kawase, T., & Stephens, D. (2003). Transient evoked otoacoustic emissions in patients with middle ear disorders. *International Journal of Audiology, 42*(3), 117-131.

Zhiqi, L., Kun, Y., & Zhiwu, H. (2010). Tympanometry in infants with middle ear effusion having been identified using spiral computerized tomography. *American Journal of Otolaryngology, 31*(2), 96-103.

Zielhuis, G. A., Rach, G. H., Bosch, A., & Broek, P. (1990). The prevalence of otitis media with effusion: A critical review of the literature. *Clinical Otolaryngology and Allied Sciences, 15*(3), 283-288.

Zumach, A., Chenault, M. N., Anteunis, L. J., & Gerrits, E. (2010). Speech perception after early-life otitis media with fluctuating hearing loss. *Audiology and Neurotology, 16*(5), 304-314.

Zumach, A., Gerrits, E., Chenault, M., & Anteunis, L. (2008). Otitis media and speech-in-noise recognition in school-aged children. *Audiology and Neurotology, 14*(2), 121-129.

Appendix I: Townsville Hospital and Health Service Ethics Approval

Queensland Government

TOWNSVILLE HEALTH SERVICE DISTRICT

Enquires to: Cass Donovan - Medical Administration
Telephone: 07 4796 1140
Facsimile: 07 4796 1051
Email: cass_donovan@health.qld.gov.au
File Number: Ethics – Protocol_ 09/30
Our Reference: dbs/ethics/Protocol/2009/0930_1

Queensland Health

14 May 2009

Mrs Sreedevi Aithal
Senior Paediatric Audiologist
Department of Audiology
The Townsville Hospital
IMB 79

Dear Mrs Aithal,

HREC Reference number: **HREC/09/QTHS/30**
Project title: Identification of middle ear pathology in infants
Protocol number: **09/30**

Thank you for your application for approval for the above study.

At their meeting held on 4 June 2009, the Townsville Health Service Human Research Ethics Committee gave ethical approval for the commencement of this study at the Townsville site subject to the following amendments are made and submitted to the Chairperson for approval.

The Committee brings the researcher's attention to Section 1.4 (f) of the *National Statement of Ethical Conduct in Research Involving Humans* which states:

> *1.4 (f) there is fair access to the benefits of research.*

The Committee brings the researcher attention to Section 20, question 18 of the Application Form. The Committee suggests "yes, individual participant's results be recorded with their personal records.

The Committee requested the following amendments to the Participant Information and Consent Form:

- *Inclusion of the Townsville Health Service District Human Research Ethics Committee contact details, as follows: Chairperson, Townsville Health Service District Human Research Ethics Committee, PO Box 670, Townsville Qld 4810; Phone: (07) 4796 1140.*

The Committee suggests the researcher include the preference "do not wish to answer" option in the Parent questionnaire at 12 Months / 24 Months.

The Committee requests the researcher to obtain a letter of support from Dr Kei.

The Committee requires the researcher to appoint a local monitor with knowledge of the area of proposed research but not directly involved with the research proposal, for the regular review of the conduct and progress of the research project.

You and your research team are informed that Townsville Health Service District HREC approval at the Townsville site is subject to continuing conditions which require all researchers to comply with certain requirements throughout the duration of the research.

Specifically, these requirements are:

1. The Principal Investigator must complete an interim report at **six months; annual report; and final report**. Such reports are to address matters including:

a. The progress of the research or the outcome in the case of completed research;
b. The maintenance and security of data and records;
c. Demonstrated compliance with the approved protocol; and demonstrated compliance with any specific conditions of approval.

The initial report **MUST** be completed at six months after the approval date of this letter and every twelve months thereafter until the completion of the study.

2. The Principal Investigator will immediately report anything which might warrant review of ethical approval of the project in the specific format, including:

a. Unforeseen events that might affect continued ethical acceptability of the project. Serious Adverse Events must be notified to the Committee as soon as possible. In addition the Investigator must provide a summary of the adverse events, in the specified format, including a comment as to suspected causality and whether changes are required to the Patient Information and Consent Form. In the case of Serious Adverse Events occurring at the local site, a full report is required from the Principal Investigator, including duration of the treatment and outcome of event.

3. Proposed changes to the research protocol, conduct of the research, or lengths of HREC approval will be provided to the HREC for review. The amended documents must be accompanied by a letter, signed by the Principal Investigator, providing a brief description of the changes, the rationale for them and their implications for the ongoing conduct of the study. All amendments should be made on the Queensland Health online NEAF application and must contain revised version numbers, version dates and page numbers. Changes must be highlighted using Microsoft Word "Track Changes" or similar. Please contact the HREC Coordinator if assistance is required.
4. The HREC will be notified, giving reasons, if the project discontinued at a site before the expected date of completion.
5. The Principal Investigator will provide an annual report to the HREC and at completion of the study in the specified format.
6. The District administration and the Human Research Ethics Committee may inquire into the conduct of any research or purported research, whether approved or not and regardless of the source of funding, being conducted on hospital premises or claiming any association with the Hospital; or which the Committee has approved if conducted outside Townsville Hospital Health Service District.

As indicated in your protocol submission, this study is scheduled to close on *30/06/2012*. The study cannot continue after this date unless approval has been given to do so. If you require an extension to this date you must send in a written request to us with an explanation of the need for the extension

Should you have any queries about the HREC's consideration of your project please contact HREC Coordinator, Cass Donovan on ☎ (07) 4796 1140. The HREC terms Reference, Standards Operating Procedures, membership and standard forms are available from the **http://www.health.qld.gov.au/cpic/ethics/reagu_homepage.asp**.

You are reminded that this letter constitutes ethical approval only. You must not commence this research project at a site until a separate authorization form, from the District CEO or delegate of that site has been obtained.

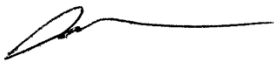
A copy of this approval must be submitted to the District Research Governance officer / Delegated Personnel with a completed Site Specific Assessment (SSA) Forms for authorization from the District CEO or delegate to conduct this research at the Townsville Hospital Health Service District.

Once authorization to conduct the research has been granted, please complete the Commencement Form and return to the office of the Human Research Ethics Committee.

The HREC wishes you every success in your research.

Should you require any additional information, please contact the HREC Coordinator, Cass Donovan on ☎ (07) 4796 1140.

Yours sincerely



Dr Isaac Seidl
Chairperson
Townsville Health Service District
Human Research Ethics Committee

HREC/09/QTHS/30_8
Human Research Ethics Committee
Townsville Hospital and Health Service
07 4433 1440

Queensland Government

Townsville
Hospital and Health Service

4TH July 2014

Sreedevi Aithal
Audiology Department
Level 2, Acute Block
Surgical Clincs
The Townsville Hospital
IMB 79

Dear Sreedevi

HREC Reference number: **HREC/09/QTHS/30**
Project title Identification of middle ear pathology in infants
Protocol number: SSA/09/QTHS/63

Thank you for submitting an amendment for the above mentioned study on 17/06/2014. The correspondence was reviewed out of session by the Human Research Ethics Committee Chairperson on 04/07/2014.

The amended documents reviewed and approved at the meeting were:

Document	Version	Date
Notification of amendment: *Addition of Research Assistant – Joshua Myers* *Curriculum Vitae – Joshua Myers*		16.06.2014

The Townsville Hospital and Health Service HREC is constituted and operates in accordance with the National Health and Medical Research Council's "*National Statement on Ethical Conduct in Human Research (2007), NHMRC* and *Universities Australia Australian Code for the Responsible Conduct of Research (2007)*, *Values and Ethics: Guidelines for Ethical Conduct in Aboriginal and Torres Strait Islander Health Research (2003)* and the *"CPMP/ICH Note for Guidance on Good Clinical Practice"*.

A copy of this letter must be forwarded to the Townsville Hospital and Health Service Research Governance Office/r.

It should be noted that all requirements of the original approval still apply.

Yours sincerely,

A/Prof Andrew Crowden
Chairperson
Townsville Hospital and Health Service
Human Research Ethics Committee

Townsville Hospital and Health Service
Human Research Ethics Committee
Telephone +617 4433 1440
Email TSV-Ethics-Committee@health.qld.gov.au

HREC/09/QTHS/30_13
Human Research Ethics Committee
Research Support Unit
07 4433 1440

Queensland Government

Townsville
Hospital and Health Service

16th July 2015

Sreedevi Aithal
IMB 79 Audiology Department
The Townsville Hospital

Dear Sreedevi

HREC Reference number: HREC/09/QTHS/30
Project title Identification of middle ear pathology in infants

Thank you for submitting an amendment for the above mentioned study on 7/07/2015.
The correspondence was reviewed out of session by the Chairperson on 16/07/2015.

The documents reviewed and approved at the meeting were:

Document	Version	Date
Notification of amendment: Single questionnaire separated into two questionnaires based on evidence from literature		23.06.15
Parent Questionnaire at 12 Months	4.0	June 2015
Parent Questionnaire at 24 Months	1.0	June 2015

The Townsville Hospital and Health Service HREC is constituted and operates in accordance with the National Health and Medical Research Council's "*National Statement on Ethical Conduct in Human Research (2007), NHMRC* and *Universities Australia Australian Code for the Responsible Conduct of Research (2007)*, *Values and Ethics: Guidelines for Ethical Conduct in Aboriginal and Torres Strait Islander Health Research (2003)* and the *"CPMP/ICH Note for Guidance on Good Clinical Practice"*.

A copy of this letter with a copy of the amended documents listed above must be forwarded to the Research Governance Office/r at each site the research is approved for.

It should be noted that all requirements of the original approval still apply.

Kind regards,

Dr Nikola Stepanov
Chairperson
Townsville Hospital and Health Service
Human Research Ethics Committee

Townsville Hospital and Health Service
Human Research Ethics Committee
Telephone +617 4433 1440
Email TSV-Ethics-Committee@health.qld.gov.au

HREC/09/QTHS/30_15
Human Research Ethics Committee
Townsville Hospital and Health Service

Queensland Government

Townsville
Hospital and Health Service

29th September 2016

Sreedevi Aithal
IMB 79 Audiology Department
The Townsville Hospital

SSA/09/QTHS/63

Dear Sreedevi

HREC Reference number: HREC/09/QTHS/30
Project title Identification of middle ear pathology in infants

Thank you for submitting an amendment for the above mentioned study on 28/09/2016.
The correspondence was reviewed out of session by the Chairperson on 28/09/2016.

The documents reviewed and approved at the meeting were:

Document	Version	Date
Notification of amendment: Request for extension		15.06.16

Please note the following key dates for this study:
HREC approval expiry: 05/07/2018
Annual report due: 05/07/2017

The Townsville Hospital and Health Service HREC is constituted according to the National Health and Medical Research Council's '*National Statement on Ethical Conduct in Human Research*' (NHMRC, 2007). The Townsville Hospital and Health Service HREC operates in accordance with the '*Values and Ethics: Guidelines for Ethical Conduct in Aboriginal and Torres Strait Islander Health Research*' (NHMRC, 2003); and the '*National Statement on Ethical Conduct in Human Research*' (NHMRC, 2007).

Please notify the Research Governance Office/r at each site of the amendment, and provide a copy of this letter with a copy of the supporting documents as listed above.

It should be noted that all requirements of the original approval still apply.

Kind Regards,

A/Prof Nikola Stepanov PhD (Melb)
Chairperson
Townsville Hospital and Health Service
Human Research Ethics Committee

Reviewed and acknowledged
by Sue Jenkins-Marsh
THHS Research Governance Officer

Signed: .....
Date: 10/10/2016

Townsville Hospital and Health Service
Human Research Ethics Committee
Telephone +617 4433 1440
Email TSV-Ethics-Committee@health.qld.gov.au

HREC/09/QTHS/30_16
Human Research Ethics Committee
Townsville Hospital and Health Service

Queensland Government

Townsville
Hospital and Health Service

21st June 2017

Sreedevi Aithal
IMB 79 Audiology Department
The Townsville Hospital

Dear Sreedevi

HREC Reference number: HREC/09/QTHS/30
Project title Identification of middle ear pathology in infants

Thank you for submitting an amendment for the above mentioned study on 18/05/2017.
The correspondence was reviewed by the Deputy Chairperson on 20/06/2017.

The amended documents reviewed and approved at the meeting were:

Document	Version	Date
Notification of amendment: Notice that Joshua Myers and Alehandrea Manuel are now full time research assistants, and Stacey Myers and Karen Nielsen have been appointed as administration officers to the project. Joshua Myers is now a PhD candidate for the project.		16.05.17
Amendment letter from PI		16.05.17
Amendment letter from PI		08.05.17
Curriculum Vitae – S.Aithal		
Curriculum Vitae – J.Kei		
Curriculum Vitae – A.Malicka		
Curriculum Vitae – S.Cheng		
Curriculum Vitae – K.Nielsen		
PhD Plan – Joseph Myers		
Adult Consent Form	3.0	16.05.17
Information Brochure	3.0	16.05.17
Parent Consent Form	3.0	16.05.17
Participant Information Sheet – Adults	3.0	16.05.17
Participant Information Sheet – Children	3.0	16.05.17

The Townsville Hospital and Health Service HREC is constituted according to the National Health and Medical Research Council's '*National Statement on Ethical Conduct in Human Research*' (NHMRC, 2007). The Townsville Hospital and Health Service HREC operates in accordance with the *'Values and Ethics: Guidelines for Ethical Conduct in Aboriginal and Torres Strait Islander Health Research'* (NHMRC, 2003); and the '*National Statement on Ethical Conduct in Human Research*' (NHMRC, 2007).

Townsville Hospital and Health Service
Human Research Ethics Committee
Telephone +617 4433 1440
Email TSV-Ethics-Committee@health.qld.gov.au

Appendix II: University of Queensland Ethics Approval

THE UNIVERSITY OF QUEENSLAND
Institutional Approval Form For Experiments On Humans Including Behavioural Research

Chief Investigator:	Mrs Sreedevi Aithal, Mr Venkatesh Aithal
Project Title:	Identification Of Middle Ear Pathology In Infants
Supervisor:	Dr Joseph Kei, Dr Carlie Driscoll
Co-Investigator(s)	Dr Joseph Kei, Dr Andrew Swanston, Katrina Roberts
Department(s):	Division of Audiology, School of Health and Rehabilitation Sciences
Project Number:	2010000842
Granting Agency/Degree:	QLD Health
Duration:	31st July 2013

Comments:

Expedited review on the basis of approval from the Townsville HSD HREC, dated 27/07/2009.

Name of responsible Committee:-
Behavioural & Social Sciences Ethical Review Committee
This project complies with the provisions contained in the *National Statement on Ethical Conduct in Human Research* and complies with the regulations governing experimentation on humans.

Name of Ethics Committee representative:-
Dr Jack Broerse
Chairperson
Behavioural & Social Sciences Ethical Review Committee

Date 06/07/10 Signature