Improving access to hearing services using automated audiometry and asynchronous telehealth

Christopher Gerard Brennan-Jones
BSc(Hons)

This thesis is presented for the degree of Doctor of Philosophy of The University of Western Australia

Ear Sciences Centre, School of Surgery, Faculty of Medicine, Dentistry and Health Sciences, The University of Western Australia

2016
Abstract

There is a global shortage of audiological services. In developed economies such as Australia, one in six people will have a hearing loss. A disabling hearing loss can have significant, negative consequences for interpersonal communication, psychosocial well-being, quality of life and economic independence irrespective of the age at which it develops.

Telehealth models of care utilising automated audiometers are a potential means to increase access to hearing services in both urban and rural regions but the evidence for this approach has been limited. The aim of this thesis was to examine the need, accuracy and application of automated audiometry for use in tele-audiology service models through a series of related experiments.

A range of methodological approaches were used to examine these questions utilising data from a mobile hearing screening service, a tertiary otolaryngology department, a meta-analysis of international research studies in the field, and a cohort of audiologists. These studies i) identified a lower rate of hearing aid uptake in rural compared to urban populations; ii) validated the use of automated audiometry in tertiary care; iii) showed that automated audiometry is highly sensitive for detecting hearing loss and identified some
limitations in its use for complex hearing pathologies; and iv) demonstrated the feasibility of incorporating automated audiometry into an asynchronous telehealth service.

This series of studies has significantly progressed the evidence-base and supports the use of automated audiometry in clinical practice and within asynchronous telehealth applications models. This approach has the potential to considerably increase efficiency and access to diagnostic audiology in underserved areas in Australia and globally.
Acknowledgements

To my supervisors, Professor Rob Eikelboom, Professor De Wet Swanepoel and Professor Marcus Atlas, thank you for all the help and guidance you have given me. You have provided wisdom, encouragement and criticism in exactly the right quantities. Rob, thanks for steering me through the last three years, celebrating the triumphs and teaching me to learn from the setbacks, and all the opportunities you have provided along the way.

I would also like to thank Professor Jonathan Foster, Dr Bill Whitmer, Dr Owen Brimijoin, Professor Michael Akeroyd, Dr Jo White, Dr Alan Archer-Boyd, Bec Bennett, Dr Dunay Taljaard, and Professor Andrew Whitehouse – all of you have given me wonderful opportunities to broaden my knowledge and experience along the way. Your help, kindness and enthusiasm has motivated me to do more.

To my Mum & Dad, my Auntie Mag and Nigel & Nancy – thank you for supporting and believing in me, without all of your help I wouldn’t be writing this. To my grandparents, Glyn, Iris, Gerard and Betty, thank you for instilling a pride in education that has extended across four generations.

Most of all, thank you to my wife and partner Sophie and our son Will, who have provided unwavering support throughout my candidature and who provide so much richness to life.
Statement of candidate contribution

This thesis contains published work that has been co-authored. The bibliographical details of the work and where it appears in the thesis are outlined below.

In regard to Regulation 41.2 from the Rules Governing Research Higher Degrees of the Graduate Research School of The University of Western Australia, the candidate was the primary contributor to the design, analysis, interpretation and preparation and revision of manuscripts for each of the studies reported in this thesis.

Chapter 3


The study design was formulated by Chris Brennan-Jones, Prof Eikelboom and Dr Taljaard. Data was collected by volunteers from the Lions Hearing Foundation of Western Australia. Statistical analysis was conducted by Chris Brennan-Jones and Prof Eikelboom. Writing of the manuscript was performed by Chris
Brennan-Jones and all co-authors contributed to editing the manuscript.

Chapter 4

The study design was formulated by Prof Eikelboom, Prof Swanepoel and Chris Brennan-Jones. Data was collected by Chris Brennan-Jones and Prof Eikelboom. Statistical analysis was conducted by Chris Brennan-Jones. Writing of the manuscript was performed by Chris Brennan-Jones and all co-authors contributed to editing the manuscript.

Chapter 5
The study design was formulated by Chris Brennan-Jones, Prof Eikelboom and Prof Swanepoel. Data for this study was collected by Chris Brennan-Jones. Statistical analysis was conducted by Chris Brennan-Jones. Writing of the manuscript was performed by Chris Brennan-Jones and all co-authors contributed to editing the manuscript.

Chapter 6

The study design was formulated by Chris Brennan-Jones, Prof Eikelboom and Prof Swanepoel. Data was collected by Chris Brennan-Jones. Statistical analysis was conducted by Chris Brennan-Jones. Writing of the manuscript was performed by Chris Brennan-Jones and all co-authors contributed to editing the manuscript.

Chapter 7
Brennan-Jones, C.G., Eikelboom, R.H., Bennett, R.J., Tao, K.F.M., Swanepoel, D. Asynchronous interpretation of manual and

The study design was formulated by Chris Brennan-Jones and Prof Eikelboom. Data was collected by Chris Brennan-Jones. Statistical analysis was conducted by Chris Brennan-Jones. Writing of the manuscript was performed by Chris Brennan-Jones and all co-authors contributed to editing the manuscript.

Christopher G. Brennan-Jones (Candidate)

Professor Robert H. Eikelboom (Principal Supervisor)

Professor Marcus D. Atlas (Co-ordinating Supervisor)
# Table of contents

Abstract ..................................................................................................................... i

Acknowledgements ............................................................................................... iii

Statement of candidate contribution ................................................................... iv

Abbreviations ......................................................................................................... xi

1. General Introduction ........................................................................................ 1
   1.1. Preamble ................................................................................................. 2
   1.2. Research aims ........................................................................................ 3
   1.3. Organisation of this thesis ........................................................................ 3
   1.4. Publications arising from this thesis ......................................................... 4
   1.5. Conference presentations arising from this thesis .................................. 5

2. Background & Rationale ............................................................................... 11
   2.1. Hearing and hearing loss ...................................................................... 12
   2.2. Access to audiological services ............................................................ 22
   2.3. Tele-audiology ....................................................................................... 25
   2.4. Early history of audiometry .................................................................... 29
   2.5. The need for further research ................................................................ 34
   2.6. References ............................................................................................. 41

3. Self-reported hearing loss and manual audiometry: a rural versus urban comparison .................................................................................................... 47
   Foreword to Chapter Three ........................................................................... 48
   Abstract ............................................................................................................ 49
   Introduction ...................................................................................................... 51
   Methods ............................................................................................................ 54
   Results ............................................................................................................. 56
   Discussion ........................................................................................................ 60
   Conclusion ........................................................................................................ 67
   References ........................................................................................................ 69
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>4FA</td>
<td>Four frequency average</td>
</tr>
<tr>
<td>ABC</td>
<td>Australian Broadcasting Corporation</td>
</tr>
<tr>
<td>AIHW</td>
<td>Australian Institute of Health and Welfare</td>
</tr>
<tr>
<td>ANOVA</td>
<td>Analysis of variance</td>
</tr>
<tr>
<td>ARIA</td>
<td>Accessibility/Remoteness Index of Australia</td>
</tr>
<tr>
<td>ASHA</td>
<td>American Speech-Language and Hearing Association</td>
</tr>
<tr>
<td>BAA</td>
<td>British Academy of Audiology</td>
</tr>
<tr>
<td>CI</td>
<td>Confidence interval</td>
</tr>
<tr>
<td>dB</td>
<td>Decibel</td>
</tr>
<tr>
<td>dBA</td>
<td>Decibel (A-weighted)</td>
</tr>
<tr>
<td>dBHL</td>
<td>Decibel (Hearing level)</td>
</tr>
<tr>
<td>GDP</td>
<td>Gross Domestic Product</td>
</tr>
<tr>
<td>HHIE-S</td>
<td>Hearing Handicap Inventory for the Elderly, Screening</td>
</tr>
<tr>
<td>HL</td>
<td>Hearing loss</td>
</tr>
<tr>
<td>HIV</td>
<td>Human immunodeficiency virus</td>
</tr>
<tr>
<td>Hz</td>
<td>Hertz</td>
</tr>
<tr>
<td>IEC</td>
<td>International Electrotechnical Commission</td>
</tr>
<tr>
<td>ISA</td>
<td>International Society of Audiology</td>
</tr>
<tr>
<td>ISO</td>
<td>International Organisation for Standardisation</td>
</tr>
<tr>
<td>OR</td>
<td>Odds ratio</td>
</tr>
<tr>
<td>QUADAS</td>
<td>Quality assessment of diagnostic accuracy studies</td>
</tr>
<tr>
<td>SNHL</td>
<td>Sensorineural hearing loss</td>
</tr>
<tr>
<td>SPSS</td>
<td>Statistical Package for the Social Sciences</td>
</tr>
<tr>
<td>UWA</td>
<td>University of Western Australia</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organisation</td>
</tr>
</tbody>
</table>
CHAPTER ONE
General Introduction
1.1 Preamble

The aim of the thesis is to present the findings of a series of experiments designed to examine the accuracy and potential applications of automated audiometry as a means to increase the efficiency and access to audiological services. Over 1.2 billion people worldwide suffer from some form of hearing loss and over 360 million people are estimated to have a disabling hearing loss (WHO, 2013). The demand for audiological services is high in both developing and developed countries and there is not sufficient capacity in audiology educational programs to meet this service need (Goulios & Patuzzi, 2008). Pure-tone audiometry, the primary method used to test hearing in humans, provides the key information on which diagnostic and treatment decisions are made. Audiometry is an essential first step along the pathway from identification to treatment for hearing loss. Automated audiometry may have the potential to facilitate improved identification of hearing loss and access to further treatment.
1.2 Research Aims

The first aim of this research was to establish the accuracy of automated audiometry and investigate the potential applications of this technology to telehealth models of care. Secondary objectives were to examine potential alternatives such as self-report measures for identifying hearing loss and to examine the specific need to improve access to audiological services in Australia. This work therefore sits in a multidisciplinary area between audiology and telehealth and aims to combine the various clinical and technological aspects of these disciplines to contribute new knowledge on methods for improving access to audiological services.

1.3 Organisation of this Thesis

This thesis is organised as a series of publications and every attempt to minimise repetition has been made. Chapter Two provides a literature review that introduces the reader to the topics of audiology, telehealth and automated audiometry and has not been submitted for publication. Chapter Three has been published and examines the accuracy of possible alternatives to audiometry and reports the discrepancies in the prevalence of hearing loss and intervention rates between rural and urban areas of Western Australia. Chapter Four is published and provides the main findings from an investigation of
automated audiometry accuracy in a clinically heterogeneous population. Chapter Five has been published and provides a pilot examination of the diagnostic accuracy of automated audiometry using clinical protocols in an asynchronous telehealth model. Chapter Six has been submitted for publication and provides an individual patient data meta-analysis of automated audiometry accuracy which builds on the work reported in Chapters Four and Five. Chapter Seven has been submitted for publication and provides the first report of agreement and reliability for remote interpretations of audiometry in an asynchronous telehealth context.

Chapter Eight provides a discussion of the findings from the research conducted as part of this project. The publications that have arisen from this thesis are included as appendices at the end of this thesis.

1.4 Publications arising from this thesis


1.5 Conference presentations arising from this thesis


diagnostic test accuracy. A Life Worth Hearing: 4th Annual Hearing Research Symposium, 27 June, Perth, Australia. (Regional)

Brennan-Jones, C.G., Taljaard, D., Safstrom, S., Bennett, R.J., Eikelboom, R.H. 2014. Sensitivity and specificity of self-reported hearing difficulty in adults over 60 years of age. XXXII World Congress of Audiology, 7 May, Brisbane, Australia (International)


1.6 Known media coverage of this thesis

A number of the studies contained within this thesis have generated local and national media attention. Below is some of the known media coverage specific to the research contained in this thesis.

- Invited opinion piece for ScienceNetworkWA on Telehealth applications in Western Australia, 2nd April 2015.

- ABC local radio “Mid West Mornings” 23rd November 2015. Expert interview about hearing loss in rural areas based on the findings from Chapter Three of this thesis.

- ABC local radio (Great Southern), 26th November 2015. Expert interview about hearing loss in rural areas based on my recent publication of rates of hearing loss in rural Western Australia.
- Interview by *ScienceNetworkWA* on the findings from Chapter Three of this thesis relating to hearing loss in rural Western Australia (15th November, 2015).

- Interview by *Stirling Times* regarding the findings from my doctoral research (30th July, 2016).

1.7 Awards and recognition during candidature

Much of the work conducted during this thesis has been of international interest in its field and has directly informed clinical practice, resulting in translational outcomes for audiology services in both metropolitan Perth and remote Aboriginal communities in Western Australia. The influence of both empirical findings and translational benefits to the community of this work have been recognised by a number of personal awards:

- Winner, Travel Scholarship, World Congress of Audiology, 2014
- Finalist, WA Science Awards – Student Scientist of the Year, 2015
- Winner, Best Clinical Science Publication, School of Surgery, 2015
- Finalist, WA Young Achiever of the Year, 2016
- Winner, WA Science Awards – Student Scientist of the Year, 2016
1.8 Other publications arising during candidature


1.9 Other conference presentations arising during candidature


1.10 References

CHAPTER TWO
Background & Rationale
2.1 Hearing and hearing loss

The human body has five senses "which serve as receivers of stimulation from outside the body...the ear is the end-organ for hearing" (Myklebust, 1971) p.11. Sound waves travelling through the air are gathered in the pinna, the outer part of the ear, and travel through the auditory canal and pass through the tympanic membrane to what is commonly known as the middle ear. Sound waves set up vibrations of the tympanic membrane which separates the outer and middle ear (see Figure 2.1). These vibrations are transformed via three small bones that make up the ossicular chain, the malleus, incus and stapes, so as to permit vibration of the fluid which fills the inner ear (the cochlea). Inside the cochlea are thousands of cilia, also known as cochlear hair cells. The cilia move to and fro in response to movements in the inner ear fluid which has been vibrated by incoming sound. Movement of the cilia discharges an electrical activity in the neurons that form the eighth cranial nerve, which connects the receptor surface of the cochlea with the central nervous system. Through developmental learning processes, differing forms and sequences of sound ultimately become associated with different events, objects and meanings. A person's ability to understand this variety of events, objects and meanings produced by sound is usually called hearing (Schubert, 1980, p. 48). If any part of this auditory pathway is compromised then a person's
hearing sensitivity can be reduced. This reduced hearing sensitivity is often termed hearing loss or hearing impairment; the potential causes are numerous, the onset can be sudden or gradual and the reduction in sensitivity can be mild to profound.

A disabling hearing loss can have devastating consequences for interpersonal communication, psychosocial well-being, quality of life and economic independence irrespective of the age at which it develops (Kotby et al., 2008; Mason & Mason, 2007). Disabling hearing impairment can lead to embarrassment, loneliness, social isolation and stigmatization, prejudice, abuse, psychiatric disturbance, depression, difficulties in relationships with partners and children, restricted career choices, occupational stress and relatively low earnings (Mohr et al., 2000; Ruben, 2000).

Hearing loss is an important public and global health concern (Davis et al., 2016). It is projected to be within the top 15 leading causes of burden of disease by 2030 (Mathers & Loncar, 2006; Davis et al., 2016). As well as hearing loss directly affecting communication and quality-of-life, a number of population-based studies suggest that hearing loss is associated with more rapid cognitive and physical aging (Lin & Ferrucci, 2012; Lin et al., 2013). The majority of people suffering hearing loss live in low-income countries where access to
audiological testing and hearing aids are severely limited (Olusanya, 2007).

Figure 2.1: Cross-sectional illustration of the auditory system (reproduced from Grey, 1911).

2.1.1. Definition

Hearing is normally assessed by audiometry whereby the sensitivity to sound is measured at selected frequencies. Hearing has traditionally been considered normal if the hearing sensitivity of one or more frequencies tested is below 20dB. Various classification systems for hearing loss have been proposed. The World Health Organisation considers a person to have a hearing loss if the average of their hearing thresholds is ≥26 dB (see Table 2.1.1).
Persons are considered to have a disabling hearing loss if the average of their hearing thresholds are ≥41 dB. This definition, whilst practical in terms of epidemiological research, is often viewed to be overly prescriptive. In 2001, the WHO published the *International classification of functioning, disability and health*, in which all impairments were assessed in relation to activity limitations and participation restrictions (WHO, 2001). This *Classification* has since provided a uniform framework for evaluating and comparing diverse body dysfunctions. It recognizes the role of contextual factors – such as environmental noise – in exacerbating functional deficits in people with “mild” or “slight” hearing impairments. It also treats “disabling hearing impairment” – or hearing disability – as a complex phenomenon that embraces bodily functions and structures as well as factors related to activity, participation and context. However, in experimental hearing research studies the use of WHO categories of hearing loss to quantify hearing loss is still widespread. Although a number of new methods of identifying hearing impairment have been proposed there is a lack of general agreement regarding the best way to quantify hearing loss (Clark, 1981; Margolis & Saly, 2007; Bucks et al., 2016).
Table 2.1.1: WHO classifications of hearing impairment (WHO, 2008)

<table>
<thead>
<tr>
<th>Grade of hearing impairment</th>
<th>Audiometric thresholds (average of 0.5, 1, 2 &amp; 4 kHz in the better ear)</th>
<th>Description of impairment</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 (no impairment)</td>
<td>25 dBHL or less</td>
<td>None or very slight hearing problems. Able to hear whispers</td>
</tr>
<tr>
<td>1 (mild/slight impairment)</td>
<td>26-40 dBHL</td>
<td>Able to hear and repeat words spoken in normal voice at 1 metre</td>
</tr>
<tr>
<td>2 (moderate/disabling impairment)</td>
<td>41-60 dBHL</td>
<td>Able to hear and repeat words using raised voice at 1 metre</td>
</tr>
<tr>
<td>3 (severe impairment)</td>
<td>61-80 dBHL</td>
<td>Able to hear some words when shouted into better ear</td>
</tr>
<tr>
<td>4 (profound impairment)</td>
<td>81 dBHL or greater</td>
<td>Unable to hear and understand even a shouted voice</td>
</tr>
</tbody>
</table>

2.1.2. Global prevalence of hearing loss

There is an upward trend in the global prevalence of hearing loss (Olusanya et al., 2014). In 1985, 42 million people, less than 1% of the global population, were thought to have a disabling hearing loss (Smith, 2003). By 1995, this had doubled to 120 million, 2.1% of the world’s population, and by 2015, 1.2 billion people were estimated to have a disabling hearing loss (Salmon et al., 2015).

Several factors have contributed to the upward trend seen in estimates of the global prevalence of disabling hearing impairment. One is the increasing prevalence of presbyacusis as mean life expectancy increases in many countries; another is improvement in and access to technology for the early detection and diagnosis of
hearing impairment (WHO, 2013). A third reason is the widespread use of ototoxic medications for treating neonatal infections, ear infections, malaria, cancer, human immunodeficiency virus (HIV) infection and drug resistant tuberculosis (Arslan et al., 1999). Rubella, mumps and measles remain significant causes of hearing impairment in regions with inadequate vaccine coverage (WHO, 2013). Furthermore, rapid and uncontrolled urbanization in many emerging economies – coupled with a common lack of enforceable regulations on environmental and occupational noise – constitutes a growing source of noise-induced hearing impairment (Basner et al., 2014). It should be noted that the estimates of the prevalence of hearing loss remain crude because many countries are unable to conduct relevant population-based surveys using standardised protocols and classification methods (Smith, 2003).

2.1.3. Australian prevalence and burden

In 2006 a commissioned report was published by Access Economics which examined the economic impact and cost of hearing loss in Australia (Access Economics, 2006). The report used prevalence-based costing to estimate the financial costs and impact on well-being caused by hearing loss. The report estimated that one in six Australians are affected by hearing loss and that prevalence rates for hearing loss are associated with increasing age, rising from less than
1% for people aged younger than 15 years to three in every four people aged over 70 years. With an ageing population, hearing loss was projected to increase to 1 in every 4 Australians by 2050.

In 2005, the real financial cost of hearing loss was $11.75 billion or 1.4% of GDP. This figure represents an average cost of $3,314 per person per annum for each of the 3.55 million Australians who have hearing loss or $578 for every Australian (Access Economics, 2006). These costs however, do not include the financial and well-being implications associated with recurrent otitis media, which affects approximately one in four Australian children and can have lasting effects on language development, behaviour, educational achievement and employment prospects (Access Economics, 2006; Brennan-Jones et al., 2014).

The figure of $11.75 billion also does not take into account the net cost of the loss of wellbeing, or disease burden, associated with hearing loss. Accounting for the disease burden of hearing loss adds a further $11.3 billion to the financial burden of hearing loss in Australia, with the largest financial cost component being productivity loss, which accounts for well over half (57%) of all financial costs ($6.7 billion).
Nearly half the people with hearing loss were of working age (15-64 years), and there are an estimated 158,876 people who were not employed in 2005 due to hearing loss. Other hidden costs associated with hearing loss are the cost of informal carers, which was estimated to cost $3.2 billion.

2.1.4. Common causes of hearing loss

Hearing loss in adults is commonly caused by the ageing process and excessive noise exposures resulting from occupational or recreational noise (Cruickshanks et al., 1998). Thus, as the population ages, there will be increasing numbers of people with hearing loss. Some of the causal factors associated with hearing loss, such as ototoxic substances (i.e. chemicals that damage or destroy the cochlear hair cells), are not as yet well understood, limiting prevention efforts in this area. However, some conditions, such as noise occupational or recreational noise induced hearing loss, are preventable.

2.1.5. Types of hearing loss

There are two main types of hearing loss, sensorineural and conductive hearing loss. A combination of sensorineural and conductive hearing loss is generally referred to as mixed hearing loss (Harrell, 2002). The first and primary impact of hearing loss is on the
perception of usable information by the individual. Any disruption to this cascade of sounds, as they move from the environment through the various parts of the ear to the auditory nerve and on to the brain, poses a threat to the individual being able to hear these sounds and recognise them as speech (Rappaport & Provencal, 2002).

i) Sensorineural hearing loss (SNHL) results from damage within or malformation of the cochlea itself, where the cochlear hair cells are either damaged or destroyed. Injury to the hair cells can result from excessive noise exposures, chemical damage such as smoking or ototoxic drugs (Nomura et al, 2004), environmental agents (Rybak, 1992) or medications (Buszman, 2003), and long term wear and tear from the ageing process, which is referred to as presbyacusis. Hearing loss can also result from damage to the auditory or eighth nerve that runs from the cochlea to the brain. Sensorineural hearing loss is permanent by nature, as human cochlear hair cells do not regenerate.

ii) Conductive hearing loss occurs when problems in the middle ear prevent sound from being conducted effectively to the inner ear. A conductive loss can be transient or permanent. Common causes are middle ear infections, wax impaction, or an ossicular chain disruption such as otosclerosis. Another common cause of mild conductive
hearing loss is Eustachian tube dysfunction (Rappaport & Provencal, 2002). Otitis media is common in children and chronic otitis media can result in temporary or permanent hearing losses and scarring of the tympanic membrane which reduces its ability to conduct sound through the ossicular chain to the inner ear (Gunasekera et al. 2009).

2.1.6. Treatment and intervention for hearing loss

Aural rehabilitation is the term generally used to describe the treatment of hearing loss. Devices such as hearing aids and cochlear implants generally form part of the broader aural rehabilitation strategy which can also include communication training, informational counselling and other techniques (Davis et al. 2016). Whilst interventions such as hearing aids and cochlear implants enhance a person’s ability to communicate, the majority of people with hearing loss (85%) do not have such devices (Access Economics, 2006). In children, it has been well established that early treatment intervention has a significant, positive impact on a child’s future well-being in terms of language development, psychosocial well-being, educational achievement and employment prospects (Yoshinaga-Itano, et al., 1998). In adults, the benefits of early intervention are less studied. However, intervention during the pre-retirement period for persons has been associated with better uptake of interventions
such as hearing aids, and better outcomes from the use of these devices (Stephens et al. 1990).

2.2 Access to audiological services

Over the past three decades there has been a marked drop in mortality rates and a corresponding rise in life expectancy that has increased attention on reducing disability and handicap caused by conditions such as hearing loss (AIHW, 2014). Audiologists are generally recognised as the primary contact for the non-medical management of patients with hearing disorders (ISA, 2004). Studies evaluating the number of audiologists per capita in developing countries have indicated between one audiologist per 0.5 million people to one per 6.25 million (WHO, 1998). In contrast, the WHO reported the number of audiologists in developed countries as being closer to one per 20,000 people (indicating a ratio density of audiologists in developing to developed countries as 300 to 1).

In Australia, one of the primary barriers to the provision of hearing services is access to qualified audiologists (AIHW, 2011). Table 2.2 summarises the distribution of audiologists in Australia, showing people living in regional and remote areas have significantly poorer access to ear health services than those living in major cities. This is largely attributable to the vast land areas to be serviced in regional and remote areas.
Table 2.2: Distribution of audiologists in Australia by capita and area
(adapted from the Australian Institute of Health and Welfare Medical Labour Workforce census (AIHW, 2011).

<table>
<thead>
<tr>
<th>Audiologists</th>
<th>Major Cities</th>
<th>Inner Regional</th>
<th>Outer Regional</th>
<th>Remote</th>
<th>Very Remote</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Per 100,000 people</td>
<td>57.37</td>
<td>33.62</td>
<td>15.34</td>
<td>30.61</td>
<td>0.00</td>
<td>48.05</td>
</tr>
<tr>
<td>Per 1000 sq kms</td>
<td>60.33</td>
<td>0.67</td>
<td>0.04</td>
<td>0.01</td>
<td>0.00</td>
<td>0.14</td>
</tr>
</tbody>
</table>

As highlighted by Goulios & Patuzzi (2008), currently there are not enough ear and hearing healthcare professionals to meet the demand for audiological services and workforce shortages are a major barrier to accessing ear and hearing healthcare. Over 86% of countries stated they did not have the required audiological workforce to meet service demand. Many low and middle-income countries have limited or no provision of otolaryngological and audiological services (Fagan & Jacobs, 2009). In sub-Saharan Africa there are approximately 0.101 audiologists per 100,000 people compared to 4.1 audiologists per 100,000 in the United Kingdom, the lowest ratio in Europe (Fagan & Jacobs, 2009).

2.2.1 Population growth and increasing need for audiology services
The provision of hearing services in developed countries is facing the challenge of increased population growth. Windmill & Freeman (2013), analysing Unites States demand, assert that the significant population increase forecast over the next 30 years will result in
increases in demand for audiological services that cannot be met by current service provision levels. They highlight a 50% increase in the number of audiologists graduating from professional programs and 1% annual compound efficiency savings in audiologist productivity over the next 30 years as key strategies to meet demand for services. They suggest the utilisation of improved technological test systems and utilisation of assistants as possible methods for achieving these efficiency savings.

Similarly, Margolis & Morgan (2008) have estimated that approximately 17.6 million people required audiometric testing in the United States in the year 2000 and 31.4 million people will require audiometric testing in 2050. They predict, assuming a 1% annual increase in audiological service provision that by 2050 there will be approximately a 50% shortfall between the number of audiograms required and the capacity for audiometry testing.

Swanepoel et al. (2010) have highlighted the potential for tele-audiology to supply the unmet global need for hearing services in developing countries. The application of automated audiometry to telehealth models of care has potential to offer significant efficiency savings to meet future demand and increase access to hearing
services by reducing the amount of time an audiologist spends performing routine diagnostic audiometry assessments.

If automated audiometry can provide accurate and reliable determination of hearing thresholds and be applied to telehealth models of care, this has the potential to achieve these efficiency savings and greatly improve the provision of ear and hearing healthcare services in these under-resourced regions.

2.3 Tele-audiology

2.3.1 What is Telehealth?

Telehealth offers unique opportunities for providing access to hearing health care services to underserved populations worldwide. The term telehealth refers to the utilization of information and communication technology in health care. Alternate terminology that has been used to describe the field includes telemedicine, online health, m-health and e-health. Telehealth literally means “health care at a distance” (Wootton, 2009). The World Medical Association (WMA) defines telehealth as “…the use of information and communications technology to deliver health and healthcare services and information over large and small distances” (WMA, 2016) and the World Health Organisation (WHO) defines telehealth as “…the delivery of health
care services, where distance is a critical factor, by all health care professionals using information and communication technologies for the exchange of valid information for diagnosis, treatment and prevention of disease and injuries, research and evaluation, and for the continuing education of health care providers, all in the interests of advancing the health of individuals and their communities.” (WHO, 2010).

More recently the specific field or specialty to which telehealth is applied has been preceded by the prefix ‘tele’, e.g. tele-audiology, which refers to the application of telehealth to the practice of audiology. Telehealth can be employed in a synchronous, real-time manner (e.g. an assessment via interactive videoconferencing) or in an asynchronous, store-and-forward manner (e.g. digital picture emailed to health care provider), or a hybrid model encompassing synchronous and asynchronous aspects can be used (Krumm, 2007; Eikelboom & Swanepoel, 2016).

The aim of telehealth is to improve health care access, quality of service delivery, effectiveness and efficiency of health care, and ameliorate the inequitable distribution of health professionals globally (Wootton, 2009). Internet connectivity and technology provides a bridge between patients and health care providers who may
otherwise be separated by distance, location, geographical and weather barriers as well as economic barriers.

2.3.2 What is Tele-audiology?

The first reports of telehealth applications applied to audiology appeared in the 1990s (Cherry & Rubinstein, 1995; Schmiedge, 1997; Birkmire-Peters et al., 1999). However, it was Swanepoel et al. (2010) who together with collaborators from Africa, Europe, North and South America and Asia provided the catalyst for increased efforts to implement, standardise and evaluate telehealth applications in audiology on a global scale. The scope of tele-audiology is broad, facilitating specialist audiological services for neonates through to the geriatric population and covers simple procedures such as otoscopy, objective tests such as tympanometry, and otoacoustic emissions, through to advanced electrophysiological testing and remote fitting of hearing aids and cochlear implants (Swanepoel & Hall, 2010). Tele-audiology services can be used to provide access to a range of screening measures, diagnostic tests and interventions, as well as education and training (Swanepoel et al., 2010). This series of studies will focus on the identification of patients with a hearing loss using automated audiometry and how this technology can improve tele-audiology services.
2.3.4 Limitations of telehealth and tele-audiology

Tele-audiology has the potential to deliver huge benefits, but it is not without its challenges and disadvantages either. There should be a genuine need for a tele-audiology within the target population. There are a broad range of audiology services that can be provided via telehealth to wide-ranging population groups, but the service must match the needs of the target population to be effective. In some cases, the capacity to perform the required testing cannot yet be reliably performed via telehealth and to attempt to do so or to limit the clinical options available to the patient without offering a face-to-face appointment may disadvantage the patient. Whilst the general attitudes towards telehealth are positive (Eikelboom & Swanepoel, 2016), not all patients may wish for or benefit from this approach. Other disadvantages may include difficulty building rapport between patients and clinicians, concerns about confidentiality, the potential for telehealth equipment to fail or malfunction and poorly established remuneration pathways for clinicians that may disincentive them from expanding their telehealth services (Kaplan & Litewka, 2008; Utidjian & Abramson, 2016).

The full process from identification, counselling, assessment for candidacy for intervention such as hearing aids and cochlear
implants, the fitting of these devices and the ongoing management and rehabilitation for people using these devices needs to be improved. In this thesis on of the first steps, identification or diagnosis of hearing loss, has been specifically targeted but this is just one step towards improving the quality of life for people with hearing loss.

2.4 Early history of audiometry and the development of automated audiometry

2.4.1 Early beginnings: the “Buzzer Audiometer”

The devices and the methods used for assessing hearing thresholds in the early 20th century were rudimentary and inconsistent. Macfarlan (1928) provides one of the first reports of attempts to standardise the use of audiometry in his paper “Circuit plans for audiometry” in which he describes the use of a “Buzzer Audiometer”. Macfarlan (1939) describes that the first types of audiometers were clock-like, giving off air-borne sound to the tubes of a stethoscope. Another model used a tripped hammer to strike a metal rod and produce the testing sound. Later, with the advent of simple electric circuits, came the use of his Buzzer Audiometer, in which the bell announcer circuit with electromagnet and interrupting armature was used and the electric interruptions were sent out to a telephone receiver, and the attenuation (reduction in sound) was produced
either by a resistance in the telephone line or by simply cutting down the battery power.

2.4.2 Hughson & Westlake Method

The Hughson & Westlake method is the most common method of performing manual audiometry. It was first described by Hughson & Westlake (1944) and is still the fundamental audiometry protocol used in modern audiology practice. The protocol involves the tester presenting frequency-specific sound stimuli and adjusting the intensity of the stimuli, either decreasing or increasing (the minimum increment is typically 5 dB), according to the patient’s response or lack or response, respectively (Hughson & Westlake, 1944). This method is also referred to as a “method of limits” approach and is performed in modern day audiology to determine hearing thresholds to the ISO 8253-1:2010 standards (or ISO 8253-1:1989 if conducted prior to the 2010 revision of standards). These standard procedures are adaptations of the original Hughson & Westlake method and are considered the gold standard test for diagnosing hearing impairment in adults and typically developing children over five years of age (ASHA, 2004).
2.4.3 Bekesy method – the first automated audiometer

In 1947 Georg von Békésy reported on a self-recording threshold audiometer which automatically increased and decreased the sound intensity while sweeping through a test-frequency range (von Bekesy, 1947). The patient was required to press a response button when a test signal was heard and release it when they lost perception of the signal. This method of determining the hearing threshold was termed “Békésy audiometry” and is also referred to as a “method of adjustment” audiometry.

Subsequent systems used derivations of this technique with fixed-frequency threshold-seeking algorithms, referred to as fixed or discreet frequency Békésy audiometry, where a sweep in intensity occurs within a fixed frequency based on the patient’s behavioural response relayed through a response switch (Meyer-Bisch, 1996). Békésy audiometry is rarely used clinically and only occasionally reported in research studies (Mahomed et al., 2013), although it has the potential to be more precise than the “method of limits” and requires less testing time.

2.4.4 Automated Hughson & Westlake audiometry

Rather than Bekesy audiometry, most of the work on automated audiometry has focused on automating the Hughson & Westlake
procedures or variations of it, first reported by Sparks (1972). Here the automated audiometry systems were programmed according to conventional manual audiometry procedural steps (Sparks, 1972), typically using versions of the Hughson and Westlake threshold-seeking method (Hughson & Westlake, 1944). The audiometer automatically makes adjustments to the intensity of the presented signal, up or downward depending on the response or lack of response. This method has also been modified in some cases to include forced-choice responses from the patient (Margolis et al., 2010). Here the listener is required to listen and make a response that either indicates that a sound was heard or not. This can be done, for example, by pressing the appropriate “Yes-button” or “No-button” on a touch-screen monitor after a signal is presented.

2.4.5 Current trends in automated audiometry

Pure-tone threshold audiometry measures are especially suited to automation because they are based on predetermined sequenced steps (Margolis & Morgan, 2008). Automated Hughson & Westlake or “method of limits” audiometry has practical benefits compared to Békésy as the results and test procedure are familiar to clinicians and non-audiological health and medical specialties. In addition, when using a computer, results can be recorded automatically enabling all the advantages of electronic record keeping, such as
reduced paperwork, transfer to other clinicians, and tracking change over time. Automated testing can incorporate quality monitoring mechanisms to ensure consistent and reliable results as has recently been demonstrated (Margolis et al., 2010), and decreases the risk of bias (Margolis et al. 2016). There is also potential to improve standardisation of test protocols and procedures across clinics and within clinics.

2.5.6 A note on accuracy studies of audiometry

Typically, accuracy refers to sensitivity and specificity; that is the ability of a diagnostic test to correctly identify those with a condition (sensitivity) and those without it (specificity) (Bossuyt et al. 2003). However, definitive diagnostic criteria are not typically applied to audiometric thresholds. In this thesis, a number of diagnostic criteria have been devised and applied and accuracy is referred to in the traditional sense in terms of sensitivity and specificity (Chapters Three, Five, Six and Seven). However, the term accuracy is also used to refer to the variance in mean hearing thresholds, with an accurate hearing threshold considered to be those which fall within the current recognised ISO standards (e.g. ISO8253-1:2010). This application is descriptive, but is useful when considering hearing loss where diagnostic criteria cannot be routinely applied but “accuracy” is still important.
Agreement or concordance is also another outcome that is measured in Chapters Five, Six and Seven. This is not accuracy *per se*, rather it is a measure of reliability (Kraemer et al., 2002).

2.5 The need for further research

Automated audiometry has made a rapid progression scientifically, technically and clinically over the past decade, however there are a number of areas requiring further examination, or yet to be examined. Presented in this thesis are a series of studies that address a number of evidence gaps in the need, validity and application of automated audiometry. The background rationale and aim of each study is presented below.

2.5.1 Self-reported hearing loss in rural and urban areas

*(Chapter Three)*

The use of self-report measures has increased in both epidemiological research and public health programmes and is a potentially useful tool to assess hearing in areas where audiometry is not available (Sindhusake et al., 2001; Swanepoel et al., 2013). The impact of poor audiological service provision in Australia has not been quantified with empirical research. Whilst we know that there are fewer audiologists in rural areas it is not clear how this translates to specific health outcomes (AIHW, 2011). Intervention (e.g. hearing
aids) for hearing loss in the pre-retirement period is associated with better patient outcomes (Stephens et al., 1990).

The aim of this study was therefore to examine whether self-reported hearing difficulty is an accurate measure of hearing loss compared to standard hearing screening with pure tone audiometry in rural and urban communities. A secondary aim was to quantify if there was any difference in the uptake of hearing intervention for participants living in rural compared to urban areas.

2.5.2 Clinical validation of automated audiometry (Chapter Four)

The device used in this study (KUDUwave 5000) has previously been validated in an environment that is not sound-treated using its manual-mode (Maclennan-Smith, Swanepoel & Hall, 2013), and in a controlled noise environment in automated-mode (Storey et al., 2014). However, these studies have typically used a pre-selected population with known hearing levels (either normal hearing or a known hearing loss).

The pre-selection of participants without a hearing loss introduces significant bias into accuracy studies (Rutjes et al., 2006). The potential for bias is clear; normal hearing patients are known to have hearing within a certain range, thereby limiting the potential range of
variation between two methods of assessment. To limit bias it is therefore essential that the accuracy of automated audiometry be examined in a population that is likely to include participants with a range of hearing threshold levels, but who are not pre-selected according to hearing status or level of impairment.

Previous research has not examined automated audiometry in an unselected, clinically heterogeneous population. Indeed, some previous studies have excluded participants with conductive hearing losses (Storey et al., 2014).

The aim of the present study was to address a gap in the evidence-base by combining automated testing in an uncontrolled environment that is not sound-treated, using an unselected clinical population of patients attending otolaryngology and audiology appointments at a tertiary public hospital. The potential influence of age and presence of hearing loss was also examined to investigate the influence of patient-related variables on accuracy of automated audiometry.

2.5.3 Diagnostic accuracy of automated audiometry and asynchronous telehealth (Chapter Five)

Asynchronous telehealth models have been under-examined in audiology and offer a potential means to increase service provision.
Disabling hearing loss, conductive hearing loss and unilateral or asymmetrical hearing loss are common reasons for referral to specialist medical and audiological professionals. If these diagnoses can be correctly identified through automated audiometry and automated diagnosis this could facilitate better triaging of patients in areas without appropriate audiological or otolaryngological staff.

Whilst standard criteria exist for diagnosing different types of hearing loss there have been no previous reports of a diagnostic accuracy analysis of automated audiometry. There are a number of classification systems for audiogram interpretation, including automated classification systems such as the AMCLASS® (Margolis & Saly, 2007). However, there have been no identified reports of the application of automated diagnosis applied to automated (or manual) audiometry in the scientific literature. Despite the lack of scientific reports, a number of patents have been filed for automated diagnosis of hearing loss (e.g. Cromwell, 2007), indicating a clear need for evidence regarding diagnostic accuracy and of automated audiometry and automated diagnosis of hearing loss.

This pilot study therefore examined the feasibility of using standardised diagnostic criteria of automated audiometry. The aim of this study was to examine the diagnostic accuracy of automated
audiometry in adults with hearing loss and the accuracy of using pre-defined diagnostic protocols asynchronous telehealth model.

2.5.4 Individual participant data meta-analysis of automated audiometry accuracy (Chapter Six)

In the meta-analysis conducted by Mahomed et al. (2013), automated audiometry showed comparable accuracy to manual audiometry. However, Mahomed et al. (2013) did not identify enough studies examining automated bone-conduction audiometry to draw any meaningful conclusions regarding its accuracy, highlighting the need for further studies and meta-analyses to examine the accuracy of this method. There were also limited reports examining participants with hearing loss.

Their analysis also identified a lack of studies examining the accuracy bone-conduction audiometry was limited to, combining summary statistics of mean differences for hearing thresholds. Since the publication of their review, a number of reports of automated audiometry (including automated bone-conduction audiometry) have been published, thereby warranting a re-examination of the accuracy of automated audiometry.
An individual participant data (IPD) meta-analysis is considered the gold-standard of meta-analyses and has a number of advantages over a conventional meta-analysis of summary statistics. An IPD meta-analysis has the ability to examine the data in detail, produce consistent analyses across studies and avoid biases associated with the use of aggregate data in meta-regression (Simmonds et al., 2005; Riley et al., 2010).

The aim of this IPD meta-analysis was to provide the most precise estimates of accuracy for automated audiometry to date and to allow the diagnostic accuracy of automated audiometry to be examined, a statistical approach that cannot be conducted using only summary statistics.

2.5.5 Asynchronous interpretation of audiometry (Chapter Seven)

Remote interpretation of automated audiometry offers the potential to enable asynchronous tele-audiology assessment and diagnosis in areas where synchronous tele-audiometry may not be possible or practical. The aim of this study was to compare remote interpretation of manual and automated audiometry.
The remote interpretation of test results is common and has been validated in a number of areas of medicine to facilitate telehealth services; including interpretations of retinal images (Chiang et al., 2006), radiography (Rosen et al., 1999), echocardiograms (Choi et al., 2011) and otoscopy (Eikelboom et al., 2002; Biagio et al., 2014). However, comparisons between remote interpretations of manual and automated audiometry have not been examined.

The aim of the present study was to examine the agreement and reliability of remote audiogram interpretation by audiologists and whether the potential variation in hearing thresholds introduced by automated audiometry would affect the clinical decisions made by audiologists.
2.6 REFERENCES


CHAPTER THREE
Self-reported hearing loss and manual audiometry: A rural versus urban comparison

(Australian Journal of Rural Health, 2016; 24(2):130-5)
Foreword to Chapter Three:

This first experimental chapter examines two screening methods that are currently used to identify hearing loss. This chapter also compares the characteristics of the rural and urban populations undergoing hearing screening in a mobile unit that could potentially provide automated audiometry instead of, or in addition to, the hearing questionnaires and screening audiometry currently provided.

This chapter serves as a summary of some of the common hearing screening practices and characterises some of the differences between rural and urban areas of Australia in relation to the uptake of hearing intervention. This chapter supports the rationale for improving access to diagnostic audiometry in rural areas and also highlights the potential benefit of incorporating automated audiometry into similar community screening programs to enable quicker access to full diagnostic hearing assessments that can better characterise the need for further audiological or medical intervention.
ABSTRACT:

Objective: To examine whether self-reported hearing difficulty is an accurate measure of hearing loss compared to standard hearing screening with pure tone audiometry in rural and urban communities.

Design: Convenience sampling.

Setting: Urban and rural areas of Western Australia.

Participants: 2090 participants (923 male; 1165 female; 2 unknown) aged 20 to 100 years presenting for community-based hearing screening in urban (982) and rural (1090) areas.

Main outcome measures: Self-reported hearing difficulty assessed with the Hearing Handicap in the Elderly-Screening (HHIE-S) questionnaire. Hearing loss defined as average hearing thresholds >25 dB in the better ear using screening audiometry conducted at 500, 1000, 2000 and 4000 Hz.

Results: The HHIE-S was sensitive (≥60yrs = 76.69%; <60yrs = 71.67%) but not specific (≥60yrs = 45.15%; <60yrs = 49.63%) for identifying hearing loss. The <60 age group had a hearing loss prevalence of 25.6%, and a false-positive rate of 67.12% compared to a prevalence of 69.12% and false-positive rate of 29.77% for the ≥60 age group. For all ages, rural participants were more likely to have a disabling hearing loss (OR 2.04 [95%CI 1.55, 2.67];
$X^2(1)=27.28; \ p<0.001$) but there were no significant differences in hearing aid uptake.

**Conclusions:** Patients in rural areas presenting for hearing screenings are more likely to suffer a hearing loss than adults in urban areas. We suggest rural health practitioners incorporate a self-reported hearing loss questionnaire into health check-ups for adults, particularly patients aged $\geq60$ years due to the high prevalence of hearing loss in this group.
INTRODUCTION

Approximately 15% of the world's adult population have some degree of hearing loss, half of which is considered to be disabling (WHO, 2013). This makes hearing loss one of the few low mortality conditions listed by the World Health Organisation (WHO) in its top twenty causes of disease burden globally. Hearing loss affects many aspects of a person’s life, including social participation, emotional and behavioural well-being and employment status, which can lead to social isolation, feelings of uncertainty, anger and anxiety (Olusanya et al., 2006). However, the lack of access and availability of appropriate interventions for patients, particularly those in remote and rural areas, is of significant concern (Dalton et al., 2003). Many patients may also feel that there is a social stigma associated with hearing loss and as a result are reluctant to raise their concerns with a health practitioner (Brooks & Hallam, 1998; Wallhagen, 2010).

Screening for hearing loss often involves audiological measures such as screening pure-tone audiometry or otoacoustic emissions testing. However, these measures are costly to implement and are often not available in primary or rural healthcare settings (Lower et al., 2010). Other screening methods, such as telephone-based hearing screening (Smits et al., 2004) and the whispered voice test
(McShefftery et al., 2013), have been shown to be a valid method of screening for hearing loss. However, telephone-based screening requires initiation from the patient and the whispered voice test is often dependent on clinician experience (McShefftery et al., 2013). Self-report questionnaires for identifying hearing loss have therefore been proposed as a cost-effective and time-efficient alternative to audiological screening that can be initiated and utilised by any health professional group (Torre et al., 2006; Ramkissoon & Cole, 2011; Swanepoel et al., 2013).

Currently, the uptake of hearing aids by patients with hearing loss is low (Laplante-Lévesque et al., 2010). Identification and intervention for hearing loss in the pre-retirement period (approximately 50 to 65 years of age) is associated with a greater uptake of hearing aids and improved rehabilitation outcomes (Stephens et al., 1990; Dalton et al., 2003). Self-report questionnaires may therefore be useful in early identification of hearing-impairment, prompting early intervention and improved outcomes in those who are motivated to receive treatment and rehabilitation for their hearing loss.

The self-assessment of hearing difficulty is designed to quantify the social and emotional aspects of hearing loss (Weinstein & Ventry,
This is an important aspect in hearing loss as the personal variability in the perception of hearing loss is wide, and levels of audiometric hearing loss are reported to account for only 50% of the variation in hearing handicap status. A number of self-report measures were therefore developed and tested against pure-tone audiometry, the standard method of diagnosing hearing loss. The HHIE-S has become one of the most popular self-report measures of hearing difficulty, is widely used and has been adapted for different age and cultural groups (Newman et al., 1990; Lichtenstein & Hazuda, 1998). The HHIE-S has been extensively psychometrically validated, and compared against similar validated questionnaires and the use of a single question for diagnosing hearing impairment (Weinstein & Ventry, 1983; Oberg et al., 2007; Tomioka et al., 2013). The HHIE-S was comparable in accuracy to similar questionnaires (Oberg et al., 2007) and had higher levels of reliability than a single question, although lower sensitivity for identifying hearing losses of 25-40 dB than a single question (Tomioka et al., 2013).

This study was designed to evaluate the accuracy of self-report measures for determining hearing loss amongst self-referring patients compared to standard, manual audiometric screening and to examine the association between the uptake of hearing aids and location (rural or urban).
METHODS

Study Cohort

Data from all hearing screenings conducted on self-referred volunteers performed by the Lions Hearing Foundation in Western Australia between March 2010 and July 2013 were collated in this study. Excluded were 182 participants who declined permission for their data to be used for research purposes, 5 participants who failed to commence or complete the hearing screening, 11 people who failed to provide a date of birth, and 49 people under 20 years of age; leaving a cohort of 2090 people. For the purposes of this study, participants were divided into two groups, broadly representing a pre-retirement group. All participants were asked for their consent for their data to be used for research purposes. Only those participants who consented are presented in this study, participants received a hearing screening free of charge and no financial incentives were offered.

Materials and procedures

All participants completed the Hearing Handicap Inventory for the Elderly–Screening (HHIE-S) (Weinstein & Ventry, 1983) and a conventional audiometric hearing screening. HHIE-S scores of >8 indicated a hearing impairment in this study (Ventry & Weinstein,
An additional questionnaire covering hearing aid use and other ear and hearing related outcomes was designed by the investigators and also included in the data collection.

Ears were examined by otoscopy to ensure there was no obstruction of the ear canal prior to screening audiometry, which was conducted with a GSI 17 audiometer (Grason Stadler, Eden Prairie, MN, USA) and Peltor 211 headphones (3M, St. Paul, MN, USA), calibrated annually. Screening was conducted by trained, non-specialist volunteers in a sound-treated booth on a bus modified for mobile hearing screenings. The Lions Hearing Bus was acoustically treated to conform to WorkSafe test standards in Australia (WorkSafe, 2002).

Hearing screening commenced at low intensity (25 dBHL). Participants “passed” if they responded to low intensity sounds at 500, 1000, 2000 and 4000 Hz in both ears. This modification in the protocol was used to make the process simpler for the screeners. If they did not respond, intensity was increased in 5 dBHL steps until the participant responded. Order of testing left or right ears first was not prescribed, and the order of testing frequencies was 500, 1000, 2000 and 4000 Hz.
**Data analysis**

Hearing screening thresholds and HHIE-S data, together with date of birth, gender and location and whether they used HAs or not were collated using Microsoft® Excel 2010 (Seattle, Washington) and analysed in SPSS v21 (New York: IBM Corp). A hearing loss was defined as mild (best ear four-frequency average >25dB) or disabling (best ear four-frequency average >40dB) across the frequencies of 0.5, 1, 2 and 4 kHz, according to the WHO criteria. An urban or rural address was determined by the postcode of the participant’s residential address. Accessibility/Remoteness Index of Australia (ARIA) classification codes were used to identify an urban (ARIA code: 0) or rural/regional address (ARIA codes: 1, 2, 3, 4) (AIHW, 2004). A number of participants had missing data for one of the four variables of hearing loss, HHIE-S score, age and postcode for rurality and this resulted in slight variations in the participant numbers across analyses.

**RESULTS**

**Participant profile**

Of the 2090 participants in the study, 982 (46.99%) resided in the metropolitan area of Perth and 1090 (52.15%) in the regional, rural or remote areas of Western Australia according to ARIA classifications;
18 (0.86%) people did not provide a valid postcode for ARIA classification. More females (n=1165) than males (n=923) were included; the sex of two people was not reported. The mean age of participants was 60.0 years (SD 13.9, range 20.3 to 100.3 years), with no statistically significant difference between the gender balances or age (independent sample t-tests; p>0.05) of urban versus rural participants. Figure 3.1 shows the distribution of participants according to age, gender and hearing loss.

**Figure 3.1:** Prevalence of hearing loss (>25dBHL) according to age and sex in participants attending the bus.
Accuracy of HHIE-S self-report questionnaire

The sensitivity of the HHIE-S questionnaire was 76.69% (95%CI 73.33, 79.75), with specificity of 45.15% (95%CI 40.29, 50.01) for identifying mild hearing loss in adults aged ≥60 years.

For adults aged <60 years, the sensitivity of the HHIE-S was 71.67% (95%CI 65.35, 77.27) and specificity was 49.63% (95%CI 45.80, 53.46). The <60 group had a false-positive rate of 67.12% [95%CI 62.82, 71.16] compared to 29.77% [95%CI % 26.57, 33.19] for the ≥60 group; the false-negative rate for the <60 year old group was 16.42% [95%CI 13.01, 20.48] and for the ≥60 years group was 54.85% [95%CI 49.91, 59.71], (see Table 3.1 for frequency distributions).

Table 3.1: Two-by-two tables showing accuracy of the HHIE-S for identifying hearing loss in adults.

<table>
<thead>
<tr>
<th>Hearing loss criteria</th>
<th>&lt;60 years of age</th>
<th>≥60 years of age</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HHIE &lt;8 (%)</td>
<td>HHIE-S ≥8 (%)</td>
</tr>
<tr>
<td>Normal hearing</td>
<td>336 (36.9)</td>
<td>341 (37.5)</td>
</tr>
<tr>
<td>Hearing loss (&gt;25 dB)</td>
<td>66 (7.2)</td>
<td>167 (18.3)</td>
</tr>
<tr>
<td>Totals</td>
<td>402 (44.2)</td>
<td>508 (55.8)</td>
</tr>
</tbody>
</table>
The prevalence of hearing loss confirmed by audiometry for adults in the participant group aged ≥60 years was 69.12% (95%CI 66.32, 71.79). Adults aged <60 years had a lower prevalence (25.60% [95%CI 22.82, 28.59]) (Figure 3.1). There was an increased risk of having a HHIE-S score >8 associated with participants from rural areas compared to urban areas (OR 1.32 [95%CI 1.11, 1.59]; $X^2(1)=9.51; p=0.002$). Participants in rural populations also had an increased risk of mild hearing loss (OR 1.73 [95%CI 1.45, 2.06]; $X^2(1)=38.13; p<0.001$) and disabling hearing loss (OR 2.04 [95%CI 1.55, 2.67]; $X^2(1)=27.28; p<0.001$). There were no differences in hearing aid uptake (mild hearing loss: OR 0.93 [95%CI 0.66, 1.46]; $X^2(1)=0.008; p=0.927$) (disabling hearing loss: OR 0.54 [95%CI 0.84, 2.20]; $X^2(1)=1.614; p=0.204$) (Tables 3.2 & 3.3).

**Table 3.2:** A comparison of prevalence and degree of hearing loss (HL) and use of hearing aid(s) (HA) in urban and rural areas (all ages).

<table>
<thead>
<tr>
<th>Hearing loss criteria</th>
<th>Urban (n = 982)</th>
<th>Rural (n = 1090)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HL (%)</td>
<td>HL+HA (%)</td>
</tr>
<tr>
<td>Mild HL</td>
<td>409 (41.6)</td>
<td>49 (5.0)</td>
</tr>
<tr>
<td>Disabling HL</td>
<td>88 (9.0)</td>
<td>29 (2.9)</td>
</tr>
<tr>
<td>Totals</td>
<td>675 (50.6)</td>
<td>105 (15.6)</td>
</tr>
</tbody>
</table>
Table 3.3: Analysis of participants with hearing loss (HL) and participants with hearing loss using hearing aids (HL + HA) in urban and rural areas (all ages).

<table>
<thead>
<tr>
<th>Hearing loss criteria</th>
<th>$\chi^2(1)$</th>
<th>$p$</th>
<th>OR</th>
<th>95%CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mild HL</td>
<td>38.13</td>
<td>&lt;0.001</td>
<td>1.73</td>
<td>1.45, 2.06</td>
</tr>
<tr>
<td>Disabling HL</td>
<td>27.28</td>
<td>&lt;0.001</td>
<td>2.04</td>
<td>1.55, 2.67</td>
</tr>
<tr>
<td>Mild HL + HA</td>
<td>0.008</td>
<td>0.927</td>
<td>0.93</td>
<td>0.66, 1.46</td>
</tr>
<tr>
<td>Disabling HL + HA</td>
<td>1.614</td>
<td>0.204</td>
<td>0.54</td>
<td>0.84, 2.20</td>
</tr>
</tbody>
</table>

DISCUSSION

This study has examined the prevalence of hearing loss in a cohort of self-referring adults from rural and urban areas of Western Australia. Participants in rural areas showed a significantly higher prevalence of hearing loss but their rates of hearing aid uptake were not significantly different compared to the urban participants. There was a high prevalence of hearing loss in those attending community-based hearing screenings and the HHIE-S questionnaire was a sensitive, but not specific, measure for detecting a hearing loss of >25dB for adults aged ≥60 years in this study. Approximately 70% of participants aged over 60 years of age had a hearing loss confirmed through audiometry. However, whilst this would appear to support calls for standard hearing screening assessments in this age group,
these prevalence figures cannot be generalised to the wider population due to the sampling method.

The findings from this study highlight the specific issue of increased prevalence of hearing loss in regional and rural areas, and support previous calls (Swanepoel et al., 2013) for the better identification of hearing loss in primary care. Identifying and treating hearing loss in the pre-retirement period is associated with an increased uptake of hearing aids and improved rehabilitation outcomes which can positively affect quality of life (Stephens et al., 1990; Dalton et al. 2003). Considering the accuracy of self-report measures in the present study, the routine use of such measures for adults aged ≥60 years in a community or primary healthcare setting may be beneficial for the early identification of hearing loss, due to the high prevalence of the condition in this age group, and the quantification of the social impact of a patient’s hearing loss. However, self-report hearing measures do not have comparable accuracy to audiometry, and better access to these services in rural and remote areas is therefore still needed.

For both age groups the self-report questionnaires were sensitive (≥60yrs = 76.69%; <60yrs = 71.67%), but not specific (≥60yrs =
Therefore, for both age groups the HHIE-S could correctly identify those with the hearing loss approximately 70% of the time. However, the questionnaire could not correctly exclude participants without a hearing loss at a rate better than 50%, potentially resulting in excessive referrals for patients self-reporting a hearing difficulty in the absence of an audiometric hearing loss. This study produced similar sensitivity estimates to previous studies of the HHIE-S (Sindhusake et al. 2001; 80%), but lower specificity (Sindhusake et al. 2001; 76%).

The false-positive rate for patients aged <60 years was substantially higher (67.12%) on self-reported hearing questionnaires compared to those aged ≥60 years (29.77%). False positives refer to when a test (in this case, self-report questionnaires) incorrectly diagnose a patient as having a condition (i.e. hearing loss), in the absence of the condition according to the gold standard (in this case, a hearing loss confirmed by audiometry). The <60 year-old group had a lower false-negative rate than the ≥60 year old group, at 16.42% and 54.85%, respectively. This is expected, considering the high prevalence of hearing loss in the ≥60 years age group. False negatives, in this case, refer to patients with a hearing loss according to audiometric testing, but who do not self-report significant hearing impairment on the HHIE-S questionnaire. The majority of these false-negatives may
represent patients who do not feel they are experiencing sufficient limitations on their quality of life to warrant intervention. Counselling on the impact of hearing loss and the benefits of early intervention may be beneficial to this population (Salonen et al., 2011). The impact of false-positives in the <60 years group is likely to put greater strain on specialist services – resulting in an increase in referrals for hearing assessment in the absence of a hearing loss. Therefore, if routine hearing screening were to be used in a primary care setting, the impact of false-negatives and false-positives in relation to patient age should be considered by the referring clinician.

A recent population-based study found that 46.2% of a regional sample population presented with a self-reported hearing loss at 60-65 years of age (Swanepoel et al., 2013). This is lower than the prevalence of self-reported hearing loss in our sample which was 57.6% for the age range 60-69 years (Figure 3.1). This was expected, considering our sample were self-referred for hearing screening, however it illustrates the high prevalence of self-reported hearing loss in the general population, with almost one-in-two people self-reporting at 60-65 years of age, and presents a significant primary health concern.
The increased risk of mild and disabling hearing losses for rural compared to urban participants was significant. Whilst this study did not examine the causes of hearing loss, previous research had indicated that both workers and families living in rural areas are at a higher risk of occupational noise exposure through agricultural and production industries (Lower et al., 2010).

The cost-effectiveness of audiometric screening in community settings and the effectiveness of community screenings for adults as a whole has been debated (Chou et al., 2011). There is some evidence (Yueh et al., 2010) to suggest that hearing screening increases the uptake of treatment (i.e. hearing aids). However, both screened and not-screened groups in the Yueh et al. (2010) study showed a low-uptake of hearing aids when followed-up one year after screening. This may suggest it is patient perceptions towards treatment-seeking for hearing loss that needs to be changed. A potential way to achieve this is the inclusion of a screening questionnaire as part of regular health checks for patients aged over 60 years old, thereby increasing exposure to hearing intervention options and decreasing social stigma.
Pronk et al. (2011) have argued for better methods of hearing screening in adults and a wider range of alternative rehabilitative options, in addition to hearing aids, such as informational counselling, communication strategies and engagement of communication partners. In our study the HHIE-S self-report questionnaire showed poor specificity for identifying a clinically significant, audiometric hearing loss. This suggests that self-report measures may be a useful tool for identifying patients with listening difficulties, in the absence of audiometric hearing loss, who could benefit from these alternative rehabilitation options. This finding also highlights the need for better access to specialist hearing services in rural areas to distinguish between patients who would benefit from amplification though hearing aids, those who have a treatable condition that could benefit from medical or surgical intervention and those who require alternative interventions for improving communication tactics.

Beyond identification of hearing loss, access to intervention for hearing loss can be an even greater challenge. Remote fitting of hearing aids and cochlear implants is a relatively new concept, but is now the subject of significant clinical and research interest (Ferrari & Bernardez-Braga, 2009; Eikelboom et al., 2014). Aside from increased coverage and service provision, remote fittings offer
significant time and monetary savings over outreach clinics (Swanepoel et al. 2010).

Whilst screening questionnaires can be effective in detecting hearing loss early, these methods are not sufficient for a definitive diagnosis and exclusion of underlying pathology, which requires someone specifically trained to provide a diagnostic audiometry assessment. These services are rarely available in rural and remote areas (AIHW, 2011) and result in increased time and cost to travel to the nearest specialist, or have a specialist visit the region. The advent of automated audiometers (Swanepoel, Mngemane et al., 2010; Eikelboom et al., 2013) that can be operated either via telehealth or by local primary care staff and interpreted remotely by audiologists may serve as a means of improving access to specialist, diagnostic audiometry services for patients in rural areas. This may enable patients with a self-reported hearing loss identified in primary care to have the type and severity of their hearing loss established locally, identifying patients suitable for hearing aids, and those who require medical or surgical intervention. This technology, in conjunction with self-report measures, has the potential to streamline the clinical pathway of patients from identification to appropriate diagnosis and intervention.
LIMITATIONS

As participants voluntarily presented for hearing screening it can be
assumed that many already had concerns about their hearing and
were therefore not representative of the general population. Poor
specificity of the HHIE-S in this study is in contrast to other studies
where specificity has ranged from 70-95% (Nondahl et al., 1998;
Sindhusake et al., 2001). However, numerous population-based
studies support the recommendation of regular, primary-care
assessments for self-reported hearing loss, whether via a single-
question or validated questionnaire (e.g. Ramkissoon et al., 2011;
Swanepoel et al., 2013).

CONCLUSION

The results highlight the increased risk of hearing loss in rural areas
compared to urban areas. This study demonstrates the potential
value of self-report measures for detecting hearing impairment for
adults, particularly those aged ≥60 years in setting where audiometric
screening equipment is not available. However, provision of
audiometry services would allow more accurate identification of
hearing loss. We suggest that primary healthcare practitioners,
particularly those in rural and remote areas, include a standardised
assessment of hearing difficulty (e.g. the HHIE-S) in routine check-
ups to screen for hearing loss in patients aged ≥60 years. This suggestion would go some way in addressing the challenges faced from a lack of specialist staff and equipment suitable for a detailed assessment of hearing loss.
REFERENCES


CHAPTER FOUR
Clinical validation of automated audiometry with continuous noise-monitoring in a heterogeneous population outside a sound-treated environment.

(International Journal of Audiology, 2016; 55(9): 507-513)
Foreword to Chapter Four

Chapter Four describes the validation of automated audiometry in a clinical setting with a highly heterogeneous patient group in terms of type and level of hearing loss, building on the current evidence base in this area. In addition, this study also introduced the calculation of a participant-level mean differences score (described as an “individual absolute mean difference” in the Chapter) which allowed the effect of the potentially confounding variables of age and presence of hearing loss on automated audiometry accuracy to be examined.

The data collected from participants in this Chapter is also used in Chapter Five for compared diagnostic accuracy of automated audiometry and in Chapter Seven where the automated and manual audiograms obtained in Chapter Four are interpreted remotely. In this sense, Chapter Four provides some background methodology for the further work carried out in Chapters Five and Seven.
ABSTRACT

Objective: Examine the accuracy of automated audiometry in a clinically heterogeneous population of adults.

Design: Prospective accuracy study. Manual audiometry was performed in a sound-treated room and automated audiometry was not conducted in a sound-treated environment.

Study Sample: 42 consecutively recruited participants from a tertiary otolaryngology department in Western Australia.

Results: Absolute mean differences ranged between 4.76–9.17 dB (air-conduction) and 8.21–14.40 dB (bone-conduction). 86.5% of manual and automated 4FAs were within 10 dB (i.e. ±5 dB); 94.8% were within 15 dB. However, there were significant ($p < 0.05$) differences between automated and manual audiometry at 0.25, 0.5, 1 and 2 kHz (air-conduction) and 0.5 and 1 kHz (bone-conduction).

The effect of age (≥55 years) on accuracy ($p = 0.014$) was not significant on linear regression ($p > 0.05$; $R^2 = 0.11$). The presence of a hearing loss (better ear ≥26 dB) did not significantly affect accuracy ($p = 0.604$; air-conduction), ($p = 0.218$; bone-conduction).

Conclusions: This study provides clinical validation of automated audiometry using the KUDUwave in a clinically heterogeneous population, without the use of a sound-treated environment. Whilst threshold variations were statistically significant, future research is needed to ascertain the clinical significance of such variation.
INTRODUCTION

Assessment of hearing sensitivity thresholds is one of the key tests conducted by audiologists. The methods of assessment are well defined by the modified Hughson & Westlake protocols ISO 8253-1:2010 (ISO, 2010). In a standard manual audiometry procedure, frequency-specific sound stimuli are presented to a patient and the hearing level of the stimuli is adjusted, either decreasing or increasing, according to the patient’s response or lack of response, respectively. This method is also termed a ‘method of limits’ approach and is performed according to ISO 8253-1:2010 standards on equipment calibrated to ISO389-1:1998 (ISO, 1998) standards. In the past decade there has been an increasing interest in systems that automate these procedures (Eikelboom et al., 2013; Ho et al., 2009; Margolis et al., 2010; Swanepoel et al., 2010).

Automated audiometers are not new. Georg von Bekesy (von Bekesy, 1947) was the first to describe a self-recording threshold audiometer which automatically increased or decreased sound level whilst sweeping a specified frequency test range. Whilst this technique is still in use by some, the Hughson & Westlake method is now the most common technique for performing audiometry. A number of automated audiology systems have implemented computerised versions of the Hughson & Westlake procedure, with
the first reports of this method of automation appearing more than four decades ago (Sparks, 1972).

Following the successful clinical validation of a number of automated audiometers (Eikelboom et al., 2013; Margolis, Frisina, & Walton, 2011; Swanepoel et al., 2010), and a systematic review of their accuracy (Mahomed et al., 2013), the potential scope of these devices has expanded to include full diagnostic hearing assessments for adults, encompassing masked and not-masked air and bone-conduction thresholds.

In the meta-analysis conducted by Mahomed et al. (2013), automated audiometry showed comparable accuracy to manual audiometry, with overall average differences of 0.4 dB (6.1 SD). However, the authors noted that there was limited data on automated bone conduction audiometry and patients with different types and degrees of hearing loss. A number of studies included in the systematic review reported the accuracy of automated audiometry on participants with normal hearing only.

The inclusion of participants without hearing loss introduces significant bias into accuracy studies (Rutjes et al., 2006). The potential for bias is clear; normal hearing patients are known to have
hearing within a certain range, thereby limiting the potential range of variation between two methods of assessment. To limit bias it is therefore essential that the accuracy of automated audiometry be examined in a population that is likely to include participants with a range of hearing threshold levels, but who are not pre-selected according to hearing status or level of impairment. The exclusion of patients with known conductive hearing impairments (e.g. Storey, et al., 2014) is also a source of potential bias. These patients represent a significant part of the clinical population and it is just as important to have accuracy estimates in such cases as for patients with sensorineural hearing losses.

The inclusion of participants with normal hearing threshold levels has been a necessary and valuable step in establishing the accuracy of automated audiometry. However, the development of studies that reduce bias by examining participants from a true clinical population will provide the most valid estimates of the accuracy of automated audiometry in practice.

One of the appeals of automated audiometry over conventional manual audiometry is its potential application in teleaudiology and its use in situations where sound treated rooms are unavailable or inaccessible. Recent reports have emphasized the global shortage of
audiological services, and highlighted that these shortages are not exclusive to low and middle-income countries (Windmill & Freeman, 2013). It has also been reported that patients living in rural and remote areas of developed countries are more likely to present to primary care with a self-reported hearing loss (Brennan-Jones et al., 2015). The ability to provide automated audiometric testing in the absence of a sound-treated environment has a great potential to increase service provision to low and middle-income countries, and rural and remote areas of high-income countries that do not have these facilities. At least two of the contemporary clinically available automation-capable audiometers (Swanepoel & Biagio, 2011; Swanepoel et al., 2010; Margolis et al., 2010) use audiocups to provide attenuation from environmental sounds, and studies have demonstrated their potential feasibility in environments that are not sound-treated (Eikelboom et al., 2013; Maclennan-Smith et al., 2013).

The device used in this study (KUDUwave 5000) has previously been validated in an environment that is not sound-treated using its manual-mode (Maclennan-Smith et al., 2013), and in a controlled noise environment in automated-mode (Storey et al., 2014). The present study therefore aims to address a gap in the evidence-base by combining automated testing in an uncontrolled environment that
is not sound-treated, using an unselected clinical population of patients attending otolaryngology and audiology appointments at a tertiary public hospital. The potential influence of age and presence of hearing loss will also be examined to investigate the influence of patient-related variables on accuracy of automated audiometry.

METHODS

Participants

42 participants (20 male, 22 female) were recruited from a publicly funded combined otolaryngology and audiology clinic at Sir Charles Gairdner Hospital, Perth, Western Australia. Attendance at the clinic was free at the point of service for patients. Inclusion criteria were: 18 years or over, no known cognitive disorder, English spoken as a first language, both ears suitable for hearing assessment. Ethics approval was granted by the University of Western Australia Human Research Ethics Committee (Reference: RA/4/1/4877).

Participant sampling and recruitment

Patient recruitment was by consecutive series, with all patients attending the clinic offered enrolment in the study, subject to inclusion criteria. Recruitment was not based on presenting symptoms (except where they contra-indicated audiological
assessment) or results from previous audiometry. No incentives were given to participants involved in the study.

Data collection

Data collection was prospectively designed. The order of test administration was not randomised. Five patients had the index test administered prior to the reference test, and all other participants (n=37) received the reference test first.

Test methods

Reference test: Manual audiometry

Manual audiometry is considered the gold standard assessment of hearing thresholds in adults and children over five years of age (ASHA, 2004) and therefore served as the reference test for this study. The Hughson & Westlake method (i.e. ascending method according to ISO8253-1:2010), or adaptations of this method according to local protocols, is typically used when determining hearing thresholds with manual audiometry. Manual audiometry was conducted within a sound-treated room (mean ambient noise level 37 dBA) using Acoustic Analyser AA30 audiometer (Starkey Hearing Technologies; Minnesota), calibrated to ISO389-1:1998 and TDH-39P (Telephonics; North Carolina) supra-aural headphones and Radioear B-71 bone-conductor (Radioear Corp.; Pennsylvania),
calibrated to ISO389-3:1994. The bone-conductor was placed on the patient’s mastoid for manual testing. Patient history, otoscopy and tympanometry using a GSI 38 Auto Tymp (Grason-Stradler; Minnesota) preceded audiometry testing.

**Index test: Automated audiometry**

Automated audiometry was conducted using the KUDUwave (eMoyoDotNet; Pretoria, South Africa) a mobile Type 2B screening, diagnostic and clinical audiometer (IEC 60645-1/2) using the ascending method according to ISO8253-1:2010. A key advantage of the KUDUwave audiometer is its double attenuation via use of insert earphones and circumaural earcups and its use of continual noise monitoring, which pauses audiometric testing if ambient noise levels exceed prescribed limits, enabling accurate testing down to 0 dB with an ambient noise level of up to 59 dB SPL. The mean ambient noise level when there was no outpatient clinic in progress was measured at 46 dBA. Placing insert earphones down to the bony part of the ear canal also reduces the occlusion effect allowing for bone-conduction evaluation with occluded ears using insert earphones (Slevin et al., 2000; Swanepoel & Biagio, 2011). However, not removing the insert earphone is a limitation to the technique as insertion down to the bony portion of the ear canal cannot be confirmed or guaranteed. If the contralateral insert earphone is removed, this can adjust for the
occlusion effect, however it also means losing some attenuation that is added by the insert. The insert earphone frequency response approximated that of the ER3A within 1 dB across test frequencies. This allowed for the use of the international insert earphone standard (ISO 389-2: 1994) for calibration. These features make the KUDUwave especially suited for use without a sound-treated environment, making it appropriate for use in rural, remote or community settings, where the availability of a sound-treated environment for testing is unlikely. The audiometry procedures were automated and recorded on a laptop using the eMOYO (v3.6.7) interface developed by eMoyoDotNet. Whenever the difference between the air conduction thresholds in the test and non-test ear was 75 dB or more at frequencies ≤1000 Hz and 50 dB or more at frequencies >1000 Hz, air conduction thresholds were masked according to current guidance (Edwards, 2010; Munro & Agnew, 1999). A masking level of 30 dB above the air conduction threshold of the non-test ear was used. Bone conduction thresholds (using a B-71 bone oscillator (Kimmetrics, Smithsburg, USA)) were determined with continuous masking in the contralateral ear. A continuous masking level of 20 dB above the air conduction threshold of the non-test ear was used. Testing took place in a quiet room that was not sound treated (mean ambient noise level when there was no outpatient clinic in progress was 46 dBA). The researcher gave
standard instructions, placed the insert earphones, bone-conductor and headset on the participant and monitored the progress of the test in case of malfunction or patient discomfort.

Definitions

Hearing thresholds were presented in dB hearing level (dBHL).

Participants were tested at air conduction frequencies of 250, 500, 1000, 2000, 4000 and 8000 Hz and bone-conduction frequencies of 500, 1000, 2000, 4000 Hz for both the reference test and index test. The audiologists administering the reference test obtained hearing thresholds at additional frequencies for participants as clinically indicated; however, these additional thresholds were not examined in this analysis as in most cases no corresponding threshold from the index test was available.

The index test had lower maximum sound level limits compared to the reference test (KUDUwave limits for air conduction were 95 dB for 0.25 kHz, 100 dB for 0.5, 1, 2 and 4 kHz and 90 dB at 8 kHz; for bone-conduction 55 dB at 0.5 kHz and 70 dB at 1-4 kHz). In cases where no response was recorded because the index test reached its maximum testable limits at a lower level than the reference test, the hearing threshold level of the reference test was corrected to the maximum output level of the index test.
Test procedure

The reference test (manual audiometry) was administered by tertiary-qualified clinical audiologists (five clinical audiologists were involved in administering the reference test throughout the study). The audiologists were all registered with the Audiological Society of Australia. Interpretation of the reference test was conducted by the clinical audiologist responsible for the patient’s care. Automated audiometry was administered by researchers involved in the project. The time interval between the reference test and the index test being conducted was less than 60 minutes for all participants, as patients proceeded directly to the next test, or after a short break if requested or required.

Blinding

The audiologist administering the reference test was blinded to the results of the index test. The researcher administering the index test was not blinded to the results of the reference test as the index test was automated and therefore could not influence the results. Other information available to the audiologist and researcher was a clinical history, and a combination of tympanometry, acoustic reflexes and speech recognition threshold testing scores, as conducted by the clinical audiologist.
**Statistical methods**

The validation analysis used air-conduction thresholds for 250, 500, 1000, 2000, 4000 and 8000 Hz and bone-conduction thresholds of 500, 1000, 2000 and 4000 Hz for both manual and automated audiometry. Mean and standard deviations were calculated for each frequency as well as real and absolute mean differences between the reference and index test hearing thresholds. Absolute mean differences are a preferable measure compared to real mean differences as absolute differences can account for positive and negative variation, whereas positive and negative variance can cancel each other out when using real mean differences (Eikelboom et al., 2013). Reference test (manual audiometry) thresholds were subtracted from index test (automated audiometry) thresholds to calculate the difference, in keeping with methodologies from similar studies (Eikelboom et al., 2013; Swanepoel et al., 2010). A paired-samples t-test and ANOVA with Bonferroni’s correction applied were used to calculate significant differences in hearing thresholds, an independent samples t-test was used for age (using 55 years of age as an arbitrary cut-point) and presence of hearing loss analysis (using better ear hearing of 4FA ≥26 dB as a cut-point). Simple linear regression was also used for the analysis of age on accuracy of automated audiometry. Excel 2010 (Microsoft®, Washington) and
RESULTS

Participants

The mean age of participants was 49.9 years (SD = 17.3, range of 19.3 to 92.5 years). Patients presented with a diverse range of clinical conditions, symptoms and co-morbidities, including but not limited to: sensorineural hearing loss, tinnitus, conductive hearing loss, otosclerosis, otitis media, acoustic neuroma, Meniere’s disease, benign paroxysmal positional vertigo, perforated tympanic membrane, Eustachian tube dysfunction, ototoxic hearing loss, skull base fracture and unilateral hearing loss, as well as pre-surgical assessment and post-surgical assessments. Hearing loss was not always the primary complaint for participants and many had more than one ear or hearing related symptom at the time of testing. The patients had a wide range of hearing losses (Table 4.1). Patients who had incomplete assessment data on either manual or automated audiometry (n = 8) or had reliability questioned by the clinical audiologist (n = 4) were not included in the analysis (Figure 4.1).
Figure 4.1: Flow diagram of patient recruitment and drop-outs

Table 4.1: Ear specific level of hearing and cumulative percentage differences.

<table>
<thead>
<tr>
<th>Hearing level (dB)</th>
<th>Total</th>
<th>%</th>
<th>c.%</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 to 10</td>
<td>8</td>
<td>9.5</td>
<td>9.5</td>
</tr>
<tr>
<td>10-20</td>
<td>22</td>
<td>26.2</td>
<td>35.7</td>
</tr>
<tr>
<td>20-30</td>
<td>17</td>
<td>20.2</td>
<td>55.9</td>
</tr>
<tr>
<td>30-40</td>
<td>15</td>
<td>17.9</td>
<td>73.8</td>
</tr>
<tr>
<td>40-50</td>
<td>7</td>
<td>8.3</td>
<td>82.1</td>
</tr>
<tr>
<td>50-60</td>
<td>4</td>
<td>4.8</td>
<td>86.9</td>
</tr>
<tr>
<td>60-70</td>
<td>8</td>
<td>9.5</td>
<td>96.4</td>
</tr>
<tr>
<td>70-80</td>
<td>1</td>
<td>1.2</td>
<td>97.6</td>
</tr>
<tr>
<td>&gt;80</td>
<td>2</td>
<td>2.4</td>
<td>100.0</td>
</tr>
<tr>
<td>Totals</td>
<td>84</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Accuracy of KUDUwave automated audiometry

Summary tables of mean absolute and real differences are presented in Tables 4.2 and 4.3 respectively. The range of absolute mean differences for all air-conduction thresholds was 4.76 to 9.17 dB (SDs 5.17 to 9.84 dB) (Table 4.2). The range of absolute mean differences for 4FA (500, 1000, 2000 and 4000 Hz) air-conduction thresholds was 4.76 to 7.02 dB (SDs 5.17 to 6.54 dB) and for all bone-conduction frequencies was 8.21 to 13.45 dB (SDs 7.42 to 10.20) (Table 4.2). The percentage of 4FA automated air-conduction thresholds falling within an absolute mean difference of 5dB of the reference test was 67.8%, within 10 dB was 86.5% and within 15 dB were 94.8% of hearing thresholds (Table 4.4).

Table 4.2: Absolute mean differences, standard deviation for 42 participants for automated audiometry compared to manual audiometry.

<table>
<thead>
<tr>
<th>Hearing Thresholds</th>
<th>Frequency (HZ)</th>
<th>250</th>
<th>500</th>
<th>1000</th>
<th>2000</th>
<th>4000</th>
<th>8000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air Right AMD</td>
<td>7.86</td>
<td>6.43</td>
<td>6.19</td>
<td>6.19</td>
<td>4.76</td>
<td>8.41</td>
<td></td>
</tr>
<tr>
<td>Left AMD</td>
<td>9.17</td>
<td>7.02*</td>
<td>6.79*</td>
<td>6.67*</td>
<td>7.02*</td>
<td>6.79*</td>
<td></td>
</tr>
<tr>
<td>Bone Right AMD</td>
<td>8.69</td>
<td>6.54</td>
<td>5.50</td>
<td>6.31</td>
<td>6.54</td>
<td>7.23</td>
<td></td>
</tr>
<tr>
<td>Bone Left AMD</td>
<td>10.83*</td>
<td>13.45*</td>
<td>9.76*</td>
<td>9.63*</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

AMD: Absolute mean difference (in dB); SD: Standard deviation; - not measured
*indicates a significant (p<0.05) difference in threshold accuracy according to a one-way ANOVA with Bonferroni’s correction applied.
Analysis of variance (ANOVA) with Bonferroni’s correction applied was used to compare the mean difference between manual and automated audiometry and these results are provided in Table 4.2.

For air-conduction audiometry, the mean differences in hearing thresholds determined by manual and automated audiometry were not significantly different \((p>0.05)\) in the right ear, but were significantly different across all frequencies in the left ear. For bone-conduction audiometry, the mean differences in hearing thresholds determined by manual and automated audiometry were significantly different at all frequencies.

**Table 4.3:** Real mean differences and standard deviation and \(p\) value for pair-wise t-test for 42 participants for index test (automated audiometry) compared to reference test (manual audiometry).

<table>
<thead>
<tr>
<th>Hearing Thresholds</th>
<th>Frequency (Hz)</th>
<th>250</th>
<th>500</th>
<th>1000</th>
<th>2000</th>
<th>4000</th>
<th>8000</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Air</strong> Right</td>
<td>RMD</td>
<td>3.33</td>
<td>4.76</td>
<td>5.00*</td>
<td>4.29*</td>
<td>-1.90</td>
<td>-2.74</td>
</tr>
<tr>
<td>SD</td>
<td></td>
<td>9.54</td>
<td>7.40</td>
<td>6.98</td>
<td>7.03</td>
<td>6.80</td>
<td>12.55</td>
</tr>
<tr>
<td>Left</td>
<td>RMD</td>
<td>5.60</td>
<td>6.07</td>
<td>5.36*</td>
<td>3.57*</td>
<td>-1.31</td>
<td>-1.55</td>
</tr>
<tr>
<td>SD</td>
<td></td>
<td>11.3</td>
<td>7.45</td>
<td>6.93</td>
<td>8.50</td>
<td>9.57</td>
<td>9.85</td>
</tr>
<tr>
<td><strong>Bone</strong> Right</td>
<td>RMD</td>
<td>-</td>
<td>6.79</td>
<td>12.74</td>
<td>3.33</td>
<td>2.02</td>
<td>-</td>
</tr>
<tr>
<td>SD</td>
<td></td>
<td>-</td>
<td>12.5</td>
<td>9.64</td>
<td>12.03</td>
<td>12.55</td>
<td>-</td>
</tr>
<tr>
<td>Left</td>
<td>RMD</td>
<td>-</td>
<td>8.93</td>
<td>13.69</td>
<td>3.21</td>
<td>2.26</td>
<td>-</td>
</tr>
<tr>
<td>SD</td>
<td></td>
<td>9.41</td>
<td>10.77</td>
<td>14.26</td>
<td>12.31</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

RMD: Real mean difference (in dB); SD: Standard deviation; - not measured. *indicates a significant \((p<0.05)\) difference according to pair-wise t-test.
Pair-wise comparisons for the real mean difference between manual and automated audiometry are provided in Table 4.3. For air-conduction audiometry, the mean differences in hearing thresholds determined by manual and automated audiometry for the frequencies 4000 and 8000 Hz were not significantly different ($p>0.05$). For bone-conduction audiometry, the mean differences in hearing thresholds determined by manual and automated audiometry for the frequencies 2000 and 4000 Hz were also significantly associated bilaterally ($p>0.05$). All other pair-wise comparisons in Table 4.3 were significantly different ($p<0.05$).

**Table 4.4:** Difference distribution of air-conduction hearing thresholds for mid-frequencies (500, 1000, 2000 and 4000 Hz).

<table>
<thead>
<tr>
<th>dB Diff</th>
<th>500 Hz</th>
<th>1000 Hz</th>
<th>2000 Hz</th>
<th>4000 Hz</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>right</td>
<td>left</td>
<td>right</td>
<td>left</td>
</tr>
<tr>
<td>0</td>
<td>11</td>
<td>11</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>18</td>
<td>13</td>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td>10</td>
<td>6</td>
<td>13</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>15</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>25</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>30</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>35</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Totals</td>
<td>42</td>
<td>42</td>
<td>42</td>
<td>42</td>
</tr>
</tbody>
</table>

% within 5 dB

<table>
<thead>
<tr>
<th></th>
<th>500 Hz</th>
<th>1000 Hz</th>
<th>2000 Hz</th>
<th>4000 Hz</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total % ≤5 dB</td>
<td>67.8</td>
<td>69.0</td>
<td>73.8</td>
<td>66.7</td>
</tr>
<tr>
<td>% within 10 dB</td>
<td>83.3</td>
<td>88.1</td>
<td>88.1</td>
<td>87.5</td>
</tr>
<tr>
<td>Total % ≤10 dB</td>
<td>86.5</td>
<td>83.3</td>
<td>83.3</td>
<td>92.9</td>
</tr>
<tr>
<td>% within 15 dB</td>
<td>98.4</td>
<td>95.2</td>
<td>95.2</td>
<td>89.5</td>
</tr>
<tr>
<td>Total % ≤15 dB</td>
<td>94.8</td>
<td>97.6</td>
<td>95.2</td>
<td>94.9</td>
</tr>
</tbody>
</table>

\[\text{Total} \% \leq 15 \text{ dB} = 94.8\]
Age differences and presence of hearing loss

To calculate the effect of age differences and the presence of a hearing loss on the accuracy of automated audiometry, each participant had an individual average absolute mean difference (individual AMD) calculated using frequencies tested for air-conduction (250–8000 Hz) and bone-conduction (500–4000 Hz) on both manual and automated audiometry. This created a summary score of audiometric variance between manual and automated audiometry for each participant to enable analysis. 16 participants had a hearing loss of ≥26 dB in both ears. Bilateral hearing loss was not significantly associated with increased variation in individual AMDs in thresholds between automated and manual audiometry for air-conduction $t(40) = -0.523; p = 0.604$ or bone-conduction $t(40) = 1.251; p = 0.218$. The mean age of participants with a bilateral hearing loss was 56.1 (SD 18.9) compared to 46.1 (SD 14.4) for those without a hearing loss, this difference was marginally not significant ($\beta = 10.10$ [95%CI -0.706, 20.913]; $p = 0.066$).

A statistically significant difference was found between age (≥55 years, $n = 14$) and hearing threshold accuracy for air-conduction $t(40) = 1.599; p = 0.014$, but not for bone-conduction $t(40) = 1.334; p = 0.190$. However, whilst linear regression showed a slight upward trend of increased individual AMDs (i.e. decreased accuracy) with
age ($R^2 = 0.11$), the relationship was not statistically significant when
analysed independently ($\beta = 0.019 \ [95\% \text{CI} -0.037, 0.074]; \ p = 0.504$)
(See Figure 4.2) or once adjusted for presence of hearing loss ($\beta =
0.025 \ [95\% \text{CI} -0.034, 0.083]; \ p = 0.398$).

**Figure 4.2:** Scatterplot showing individual absolute mean differences
(AMD) for air-conduction between automated and manual audiometry
against participant age at testing ($R^2 = 0.11$).
Excluded participants

Details of excluded participants are provided in Table 4.5. The mean age is lower than included population and there were equal number of normal hearing and hearing impaired participants excluded.

Table 4.5: Data from excluded participants (XP)

<table>
<thead>
<tr>
<th>Participant</th>
<th>Gender</th>
<th>Age (yrs)</th>
<th>4FA (dB)</th>
<th>Reason for exclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>XP1</td>
<td>M</td>
<td>73.6</td>
<td>70.00</td>
<td>Spoke English as second language</td>
</tr>
<tr>
<td>XP2</td>
<td>F</td>
<td>49.6</td>
<td>13.75</td>
<td>Spoke English as second language</td>
</tr>
<tr>
<td>XP3</td>
<td>F</td>
<td>25.4</td>
<td>7.50</td>
<td>Audiologist reported poor reliability</td>
</tr>
<tr>
<td>XP4</td>
<td>F</td>
<td>69.8</td>
<td>41.25</td>
<td>Audiologist reported poor reliability</td>
</tr>
<tr>
<td>XP5</td>
<td>F</td>
<td>37.9</td>
<td>18.75</td>
<td>Audiologist reported poor reliability</td>
</tr>
<tr>
<td>XP6</td>
<td>F</td>
<td>n/a</td>
<td>n/a</td>
<td>Device failure</td>
</tr>
<tr>
<td>XP7</td>
<td>F</td>
<td>61.2</td>
<td>50.00</td>
<td>Audiologist reported poor reliability</td>
</tr>
<tr>
<td>XP8</td>
<td>M</td>
<td>61.8</td>
<td>n/a</td>
<td>Incomplete test</td>
</tr>
<tr>
<td>XP9</td>
<td>F</td>
<td>50.8</td>
<td>n/a</td>
<td>Device failure</td>
</tr>
<tr>
<td>XP10</td>
<td>F</td>
<td>21.1</td>
<td>n/a</td>
<td>Device failure</td>
</tr>
<tr>
<td>XP11</td>
<td>F</td>
<td>56.4</td>
<td>n/a</td>
<td>Device failure</td>
</tr>
<tr>
<td>XP12</td>
<td>M</td>
<td>75.7</td>
<td>n/a</td>
<td>Device failure</td>
</tr>
<tr>
<td>Mean age</td>
<td></td>
<td></td>
<td></td>
<td>53.0</td>
</tr>
<tr>
<td>SD age</td>
<td></td>
<td></td>
<td></td>
<td>18.5</td>
</tr>
</tbody>
</table>
DISCUSSION

This study presents data relating to the accuracy of automated audiometry in a clinical population, using consecutive series recruitment. The present study also examined a range of patient-related factors that may affect accuracy between manual and automated audiometry. The study cohort represents a wide range of type and severity of hearing losses, which has been highlighted as a key limitation of previous studies (Mahomed et al., 2013).

According to the current ISO standard 8253-1:2010, the standard variability for determining a hearing threshold level at frequencies below 4 kHz is 4.9 dB, in a sound-treated environment, without masking and assuming no other uncertainties. To account for uncertainties, the standard acceptable variability in audiometry is an absolute difference of 10 dB, representing the typical ±5 dB test-retest criteria that is practiced widely by audiologists and present in most audiological standards (ASHA, 2004).

Participants in this study had their automated audiometry thresholds tested in a room that was not sound-treated, whilst the manual audiometry testing was performed in a sound-treated environment, potentially introducing further accuracy variation from previous
studies (Maclennan-Smith et al., 2013; Storey et al., 2014). The placement of the bone-conductor also differed, with mastoid placement used for the reference test and forehead placement used for the index test and ambient noise levels for the reference test also exceeded the recommended ISO 8253-1:2010 standard. These factors may therefore have contributed additional variation between the reference and index tests.

Therefore, a number of potential influences may have affected variability for this study, with standard test variability, inter-tester differences for manual audiometry, calibration differences and the test environment all likely to influence variation in addition to that caused by automation. Considering the compounded variation of these variables, the variability due to automation appears acceptable, with 86.5% of four-frequency thresholds (500, 1000, 2000 and 4000 Hz) within the accepted ISO absolute variation of 10 dB and 94.8% of participant’s thresholds within a further 5 dB of this (i.e. absolute mean threshold difference within 15 dB, real mean difference of ±7.5 dB). This shows that for approximately 95% of participants in this study the additional variation introduced was within an additional 5 dB absolute difference (±2.5 dB relative difference) of ISO standards. It is also comparable to previous clinical validation studies of the KUDUwave which showed 91% accuracy (Swanepoel et al., 2010)
and 92% accuracy (Storey et al., 2014) of obtaining hearing thresholds within a 10 dB absolute difference in sound-treated and 40 dBA multi-talker background noise environments.

Whilst the additional 5 dB variation above the ISO standard may be considered low once the confounding factors are accounted for, the accuracy of automated audiometry in this study is lower than previous studies which have fallen within the ISO standard threshold variation limits (Eikelboom et al., 2013; Margolis et al., 2011; Storey et al., 2014; Swanepoel et al., 2010). Use of the same KUDUwave audiometer in manual and automated mode, as in Swanepoel et al. (2010), eliminates the calibration differences that were introduced in the present study, and could easily account for some of the increased differences in hearing thresholds. The use of sound-treated environments in previous studies (Eikelboom et al., 2013; Margolis et al., 2011; Swanepoel et al., 2010), or artificial background noise environments (Storey et al. 2010) can also help account for the slightly better accuracy estimates in these studies. This study therefore provides clinical validation and accuracy levels for automated audiometry in environments with variable noise levels, typical of an outpatient clinic or health professional’s office.
There were a number of outlying hearing thresholds in the study cohort, with six individual threshold differences of 30-35 dB occurring across five different participants, which may have skewed mean differences given the size of our study sample. We have found no clear reason for these outliers but future studies with increased sample sizes may be able to better account for such variation. Whilst previous studies have excluded participants as outliers (Storey et al., 2014), the outliers in the present study were specific to thresholds, not individuals themselves. Inter and intra-tester variability scores were not possible to assess in this study, but previous studies have indicated that the average level of inter-tester variation can be between 2.3–6.0 dB for air-conduction thresholds and 2.9–7.9 dB for bone-conduction thresholds (Margolis et al., 2010). This does not include clinician-specific variation around certain frequencies such as 4000 Hz (Margolis et al., 2013), which may also have increased variation in the results. Data from participants with an outlying threshold have not been removed as it could be argued that their results give a clearer picture of the potential issues facing automated audiometry, namely identifying participants who would benefit from manual audiometry due to patient-related factors that may potentially affect reliability, such as memory, attention, reaction times and physiological aspects of ageing (e.g. ear canal structure) (Landry, 1999).
To date, the analysis of results comparing automated and manual devices has been predominately descriptive; here we have included analysis of the pair-wise relationship between automated and manual audiometry results in conjunction with an analysis of variance approach. We have also included independent tests of the effect of age, presence of hearing loss on accuracy using a multivariate regression model and used simple linear regression to examine the influence of age on accuracy (Figure 2). The results from these statistical tests are mixed. For the pair-wise tests of automated versus manual audiometry accuracy, the low and mid frequencies for both air and bone conduction showed statistically significant differences with one another. However, at high frequencies for air conduction (4000 and 8000 Hz) and bone conduction (2000 and 4000 Hz) there were no significant statistical differences between the two testing methods. These results from pair-wise tests at high-frequencies were unanticipated. There is a recognized positive skew in manual thresholds referred to by (Margolis et al., 2015) as the “Good enough” bias which is believed to be the result of manual testers not acquiring accurate thresholds below 0 dB because this is deemed to be a sufficient level of hearing. Automated audiometry does not have this basis and may therefore introduce variation. There are also recognised calibration discrepancies at 4000 Hz for
bone conduction that may account for additional variation, but in this case, the influence of these differences were not significant on pair-wise analysis (Margolis et al., 2013). For the ANOVA analysis, a significant difference was seen for left ear air conduction thresholds but not right ear air conduction thresholds. These results highlight the random error (as opposed to systematic error) associated with behavioural testing and order effects. We therefore suggest that automated audiometry include a number of pre-assessment trials to familiarise the patient with the automated procedure before the hearing threshold assessment begins.

For air-conduction, 8000 Hz thresholds presented the most variability according to absolute mean difference, although these differences were not statistically significant according to pair-wise tests. It has been established that high-frequency audiometric testing at or above 8000 Hz is more susceptible to variation from differences in the coupling of headphones or earphones and individual physiological differences, with additional variation differences of up to 10 dB (Gössing, 2003). It is therefore possible that the use of insert earphones (KUDUwave) compared to manual audiometry (supra-aural headphones) may have introduced additional variation at higher frequencies.
Age was examined as a potential source of variability in this study. We found statistically significant differences in threshold accuracy between manual and automated audiometry for participants aged ≥55 years using air-conduction audiometry when using a t-test approach. However, this effect was not significant using linear regression or once adjusted for the presence of hearing loss. Therefore, any effect of age on threshold accuracy using automated audiometers (see Figure 4.2) was either very weak or non-existent (i.e. a Type I error or “false-positive”). However, future studies examining this variable would be beneficial. Despite the variation that may have been introduced by a heterogeneous clinical population, we detected no significant association between the presence of hearing loss and accuracy of automated audiometry.

It could be argued that whilst statistically significant differences have been identified at certain frequencies, the variation in hearing thresholds may be of minimal clinical significance. Whilst the absolute mean variation in thresholds exceeded the ISO standard by 5 dB in this study, this is still the minimal measurable difference in most conventional audiometers and the clinical implications of the difference may be similar to equivalent inter-tester, environmental or
calibration differences. Future research should focus on isolating key variables that increase threshold variation as optimal test conditions and patient factors deteriorate and investigate the clinical, rather than statistical, significance of audiometric variation and the effect on audiometry interpretation due to automation.

CONCLUSION

This study has described the clinical validation of automated audiometry in an unselected, clinically heterogeneous population, without the use of a sound-treated environment and with numerous manual audiometry testers, potentially introducing a high degree of inter- and intra-tester variability. Considering this, the difference in hearing thresholds is low, with 86.5% of 4FAs within 10 dB and 94.8% within 15 dB. This study did however reveal that in these least optimal conditions for automated audiometry, the majority of automated hearing thresholds were statistically different to manual thresholds, with the exception of high frequency air-conduction (4000 and 8000 Hz) and bone-conduction (2000 and 4000 Hz) frequencies, and this should be considered when interpreting audiograms produced via this method. However, whilst this variation was statistically significant, future research is needed to ascertain the clinical significance of such variation.
REFERENCES


CHAPTER FIVE
Diagnosis of hearing loss using automated audiometry in an asynchronous telehealth model:
A pilot accuracy study
(Journal of Telemedicine and Telecare, 2016; Apr 6)
Foreword to Chapter Five

Chapter Five provides the rationale, pilot data and a potential service model for an asynchronous tele-audiology program incorporating automated audiometry. The original rationale for examining the potential benefit of an asynchronous tele-audiometry model came from observations on remote clinical visits that many remote Western Australian communities often did not have a reliable internet connection that could sustain synchronous remote tele-audiometry assessments.

This Chapter used diagnostic accuracy methods to ascertain sensitivity and specificity levels for determining various types and levels of hearing loss. To do this required the application of pre-defined diagnostic protocols to hearing threshold data (obtained in Chapter Four) to create a series of dichotomous outcomes for various classifications of hearing impairment.
ABSTRACT:

Introduction: Standard criteria exist for diagnosing different types of hearing loss, yet audiologists interpret audiograms manually. This pilot study examined the feasibility of standardised interpretations of audiometry in a telehealth model of care. The aim of this study was to examine diagnostic accuracy of automated audiometry in adults with hearing loss in an asynchronous telehealth model using pre-defined diagnostic protocols.

Materials and Methods: We recruited 42 study participants from a public audiology and otolaryngology clinic in Perth, Western Australia. Manual audiometry was performed by an audiologist either before or after automated audiometry. Diagnostic protocols were applied asynchronously for normal hearing, disabling hearing loss, conductive hearing loss and unilateral hearing loss. Sensitivity and specificity analyses were conducted using a two-by-two matrix and Cohen’s kappa was used to measure agreement.

Results: The overall sensitivity for the diagnostic criteria was 0.88 (range: 0.86 – 1) and overall specificity was 0.93 (range: 0.86 – 0.97). Overall Kappa (k) agreement was “substantial” $k = 0.80$ [95%CI 0.70, 0.89] and significant at $p<0.001$. 

- 107 -
Discussion: Pre-defined diagnostic protocols applied asynchronously to automated audiometry provide accurate identification of disabling, conductive and unilateral hearing loss. This method has the potential to improve synchronous and asynchronous tele-audiology service delivery.
INTRODUCTION

There is a global shortage of audiological services that is not exclusive to low and middle-income countries (Olusanya et al., 2014). In developed economies, people living in rural areas have been shown to have higher rates of hearing loss with lower uptake of interventions (Lower et al., 2010; Brennan-Jones, Taljaard et al., 2016), and as many as 25% of children may require diagnostic audiometric testing to identify hearing loss associated with recurrent ear disease (Brennan-Jones et al., 2015). Telehealth models of care utilising automated audiometers have therefore been proposed as a potential means to increase access to hearing services in underserved regions (Margolis & Morgan, 2008; Swanepoel, Clark et al., 2010; Windmill & Freeman, 2013).

A number of automated audiometers have recently been clinically validated, including the AMTAS (Margolis et al., 2010; Eikelboom et al., 2013) and KUDUwave (Swanepoel, Mngemane et al., 2010). The studies examining clinical validation typically utilise a comparative or test-retest design, aimed at comparing the accuracy of individual hearing thresholds. These studies have helped determine the strengths and weaknesses of automated audiometry. The effects of introducing background noise, along with other variables, have been
examined and the consensus across studies is that automated audiometry is a suitable alternative to manual audiometry (Storey et al., 2014). However, these approaches all rely on the availability of an audiologist to interpret the audiometry results. Pre-defined diagnostic protocols that can correctly identify patients with a hearing diagnosis that requires further examination using automated audiometry may have the potential to increase the efficiency of audiological services. In synchronous models of tele-audiology, pre-assessment using diagnostic protocols applied to automated audiometry could reduce the number of audiologist-administered audiograms required per session and decrease the time to diagnosis and referral in asynchronous models of care.

Classification systems applied to audiometry have been the subject of previous efforts to standardise the interpretation and reporting of audiometric results (Guild, 1932; Carhart, 1945). More recently, the World Health Organisation has issued guidance on the level of severity for hearing loss (WHO, 2013). Margolis & Saly (2007) have developed a comprehensive automated system for audiogram classification (AMCLASS). However, the severity of hearing loss is a general descriptor, and not specific to a diagnosis (Clark, 1981). As such, previous studies have expressed the need for a consensus on diagnosing the site of lesion for a given hearing loss (i.e. is the
Margolis & Saly (2007) have previously focused on improving agreement between clinical audiologists by standardising the classification and configuration of audiograms by creating the AMCLASS software which incorporates 161 audiogram classifications. However, the AMCLASS is currently only available for use with the AMTAS automated audiometry software. This study will therefore utilise simple, freely available, diagnostic guidelines issued by professional bodies that can be applied to any type of manual or automated audiometer and interpreted by telehealth facilitators with minimal training. The primary focus of the diagnostic criteria used in this study was to identify patients for suitable for interventions, either medical or audiological. Disabling hearing loss, conductive hearing loss and unilateral or asymmetrical hearing loss are common reasons for referral to specialist medical and audiological professionals. However, using diagnostic criteria can streamline the referral process and limit unnecessary medical and audiological referrals. Patients with a bilateral sensorineural hearing loss will,
generally, be considered initially for audiological intervention (e.g. hearing aids) without the need for a medical referral. For patients with a significant conductive hearing loss, both medical intervention (to assess whether any hearing can be restored through surgical or non-surgical intervention) and audiological intervention will typically be required. Patients with a significant unilateral or asymmetrical hearing loss will require a medical and often a radiological referral to exclude acoustic neuroma in additional to an audiological referral (BAA, 2009). The aim of the study was to examine the accuracy of standard diagnostic criteria applied to automated audiometry to identifying a number of key audiometric characteristics which can guide further referral to specialist services.

METHODS

Participants

Study population

We recruited 42 study participants (20 male, 22 female) from a publicly-funded combined audiology and otolaryngology clinic at Sir Charles Gairdner Hospital, Perth, Western Australia. Attendance at the clinic was free at the point of service for patients. Inclusion criteria were: 18 years or over, no known cognitive disorder, spoke
English as a first language, both ears suitable for hearing assessment. Ethics approval was granted by The University of Western Australia Human Research Ethics Committee.

Participant recruitment and sampling

Patient recruitment was by consecutive series, with all patients attending the clinic offered enrolment in the study, subject to exclusion and inclusion criteria. Recruitment was not based on presenting symptoms (except where they contra-indicated audiological assessment) or results from previous audiometry. No incentives were offered to participants involved in the study.

Test methods

Reference test: Manual audiometry

Manual audiometry is considered the gold standard assessment of hearing thresholds in adults and children over five years of age (ASHA, 2004) and therefore served as the reference test for this study. The Hughson & Westlake method, or adaptations of this method according to local protocols, is typically used when determining hearing thresholds with manual audiometry. Manual audiometry was conducted within a sound-treated room (mean ambient noise level 37 dBA) using Acoustic Analyser AA30 audiometer (Starkey Hearing Technologies; Minnesota), calibrated to

Index test: Automated audiometry

Automated audiometry was conducted using the KUDUwave (eMoyoDotNet; Pretoria, South Africa) a mobile Type 2B screening, diagnostic and clinical audiometer (IEC 60645-1/2) using the ascending method according to ISO8253-1:2010. A key advantage of the KUDUwave audiometer is its double attenuation via use of insert earphones and circumaural earcups and its use of continual noise monitoring, which pauses audiometric testing if ambient noise levels exceed prescribed limits, enabling accurate testing down to 0 dB with an ambient noise level of up to 59 dB SPL. Ambient noise levels are monitored in octave bands through an external microphone on each circumaural earcup. The noise monitoring function of the KUDUwave uses low-pass (<125 Hz), even single octave band-pass (125, 250, 500, 1000, 2000, 4000, and 8000 Hz) and high-pass (>8000 Hz) filters to separate the incoming sound. The filters have a stop-band attenuation of 90 dB and pass-band ripple of 0.003 dB. The outputs
of these filters are monitored in real-time and the peak value passes to the user interface software (eMOYO) every 100-ms, which is visually represented within the software (Figure 5.1). By representing the peak and not the average ambient noise values averaged over each 100-millisecond period, the device provides an aggressive monitoring function. The noise level indicated represents the peak value for the microphone at the ear undergoing testing. Placing insert earphones down to the bony part of the ear canal also reduces the occlusion effect allowing for bone-conduction evaluation with occluded ears using insert earphones (Dean & Martin, 2000; Swanepoel & Biagio, 2011). The insert earphone frequency response approximated that of the ER3A within 1 dB across test frequencies. This allowed for the use of the international insert earphone standard (ISO 389-2, 1994) for calibration. These features make the KUDUwave especially suited for use without a sound-treated environment, making it appropriate for use in rural, remote or community settings, where the availability of a sound-treated environment for testing is unlikely.

The audiometry procedures were automated and recorded on a laptop using the eMOYO (v3.6.7) interface developed by eMoyoDotNet. The diagnostic criteria were calculated post-testing and were not programmed into the eMOYO software. Whenever the
difference between the air conduction thresholds in the test and non-test ear was 75 dB or more at frequencies ≤1000 Hz and 50 dB or more at frequencies >1000 Hz, air conduction thresholds were masked. A masking level of 30 dB above the air conduction threshold of the non-test ear was used. Bone conduction thresholds (using a centrally placed B-71 bone oscillator, Radioear B-71 (Radioear Corp.; Pennsylvania)) were determined with continuous masking in the contralateral ear. A continuous masking level of 20 dB above the air conduction threshold of the non-test ear was used. Testing took place in a quiet room that was not sound treated (mean ambient noise level when there was no outpatient clinic in progress was 46 dBA). The researcher gave standard instructions, placed the insert earphones, bone-conductor and headset on the participant and monitored the progress of the test in case of malfunction or patient discomfort.

**Figure 5.1:** KUDUwave automated audiometry screen display during testing.
Hearing thresholds were presented in dB hearing level (dB HL).

Participants were tested at air conduction frequencies of 250, 500, 1000, 2000, 4000 and 8000 Hz and bone conduction frequencies of 500, 1000, 2000, 4000 Hz for both the reference standard and index test. The audiologists administering the reference standard obtained hearing thresholds at additional frequencies for participants as clinically indicated; however, these additional thresholds were not examined in this analysis as no corresponding threshold from the index test was available. The index test had lower maximum testable sound level limits compared to the reference standard (KUDUwave limits for air conduction were 95 dB for 250 Hz, 100 dB for 500, 1000, 2000 and 4000 Hz, and 90 dB at 8000 Hz; sound level limits for bone conduction were 55 dB at 500 Hz and 70 dB at 1000, 2000 and 4000 Hz). In cases where both the index test and the reference standard reach maximum testable limits before a participant's hearing threshold was established, a predicted threshold at the index tests maximum level was imputed.

**Diagnostic criteria**

All participants were examined to identify whether they had normal hearing, a disabling hearing loss, a conductive or sensorineural hearing loss and whether their hearing thresholds were bilateral, unilateral or asymmetrical except where individuals had missing data...
that prevented the diagnostic criteria being calculated ($n = 3$). Normal hearing in this study was defined according the World Health Organisation (2013) criteria of $\geq 26$ dB for hearing loss (i.e. normal hearing is $< 26$ dB) in the better ear, with hearing thresholds averaged across 500, 1000, 2000 and 4000 Hz. Disabling hearing loss in this study was defined according the World Health Organisation (2013) criteria of $\geq 41$ dB hearing loss in the better ear, with hearing thresholds averaged across 500, 1000, 2000 and 4000 Hz.

Conductive hearing loss was defined in this study as an air-bone gap of $\geq 20$ dB at two or more adjacent frequencies out of 500, 1000, 2000 and 4000 Hz (BAA, 2009). The individual frequency bone-conduction thresholds were subtracted from the individual frequency air-conduction thresholds for both the index and reference test to obtain the air-bone gap.

Unilateral or asymmetrical hearing loss was defined in this study as a $\geq 20$ dB difference between the left and right bone-conduction thresholds at two or more adjacent frequencies out of 500, 1000, 2000 or 4000 Hz (BAA, 2009).
**Test procedure**

The reference standard (manual audiology) was administered by one of five clinical audiologists. The audiologists were all full members of Audiology Australia. Automated audiology was facilitated by researchers involved in the project. The time interval between reference test and the index test being conducted was less than 60 minutes for all participants, as patients proceeded directly to the next test, or after a short break if requested.

**Blinding**

The audiologist administering the reference standard (manual audiology) was blinded to the results of the index test (automated audiology). The researcher administering the index test was not blinded to the results of the reference test as the index test was automated and therefore could not influence the results.

**Statistical methods**

The diagnostic accuracy analysis used air-conduction thresholds for 500, 1000, 2000, 4000 Hz and bone-conduction thresholds of 500, 1000, 2000 and 4000 Hz for both manual and automated audiology. Sensitivity and specificity along with positive and negative predictive value were calculated for each category of hearing loss.
Agreement between automated and manual audiometry across the diagnostic categories was assessed using Cohen’s Kappa statistic. The Landis & Koch (1977) recommendations of agreement classification were applied, with $k<0$ indicating no agreement, $k = 0–0.20$ indicating “Slight” agreement, $k = 0.21–0.40$ indicating “Fair” agreement, $k = 0.41–0.60$ indicating “Moderate” agreement, $k = 0.61–0.80$ indicating “Substantial” agreement and $k = 0.81–1.00$ indicating “Almost perfect” agreement.

**RESULTS**

*Participants*

The mean age of participants was 49.9 years (SD = 17.3, range of 19.3 to 92.5 years). Patients presented with a diverse range of symptoms and comorbidities, including but not limited to: sensorineural hearing loss, tinnitus, conductive hearing loss, otosclerosis, otitis media, acoustic neuroma, Meniere’s disease, benign paroxysmal positional vertigo, perforated tympanic membrane, Eustachian tube dysfunction, ototoxic hearing loss, skull base fracture, unilateral hearing loss, pre-surgical and post-surgical assessment. Hearing loss was not the primary complaint of all participants and many had more than one ear or hearing related disorder at the time of testing. An additional eleven patients who did
not complete both hearing assessments (n = 7) or had reliability questioned (i.e. suspected attention or cognitive ability issues) by the clinical audiologist (n = 4) were not included in the analysis.

Diagnostic accuracy of clinical protocols for automated audiometry

Table 5.1 presents 2x2 tables of the diagnostic data for detecting disabling, conductive and unilateral hearing loss using pre-defined clinical protocols applied to automated audiometry. Three participants had incomplete data that prevented the calculation of diagnostic criteria (one patient for disabling hearing loss, two for unilateral hearing loss). Table 5.2 presents sensitivity and specificity results; overall sensitivity was 0.88 [95%CI 0.75, 0.95] and overall specificity was 0.93 [95%CI 0.88, 0.97]. The highest sensitivity and specificity was for identifying disabling hearing loss (1.00 and 0.97 respectively), and the lowest was for identifying a normal ear (0.86 and 0.86 respectively). There were generally greater 95% confidence intervals for sensitivity than for specificity. The positive predictive values (PPV) ranged between 75% and 86% across the diagnostic criteria and negative predictive values (NPV) ranged from 92% and 100% (Table 5.2).
Table 5.1: Participant classification according to diagnostic criteria presented in 2x2 format.

<table>
<thead>
<tr>
<th>Diagnostic criteria</th>
<th>HL Absent</th>
<th>HL Present</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Normal</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test Positive</td>
<td>4</td>
<td>12</td>
<td>16</td>
</tr>
<tr>
<td>Test Negative</td>
<td>24</td>
<td>2</td>
<td>26</td>
</tr>
<tr>
<td>Totals</td>
<td>28</td>
<td>14</td>
<td>42</td>
</tr>
<tr>
<td><strong>Disabling HL</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test Positive</td>
<td>1</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Test Negative</td>
<td>36</td>
<td>0</td>
<td>36</td>
</tr>
<tr>
<td>Totals</td>
<td>37</td>
<td>4</td>
<td>41</td>
</tr>
<tr>
<td><strong>Conductive HL</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test Positive</td>
<td>3</td>
<td>18</td>
<td>21</td>
</tr>
<tr>
<td>Test Negative</td>
<td>60</td>
<td>3</td>
<td>63</td>
</tr>
<tr>
<td>Totals</td>
<td>63</td>
<td>21</td>
<td>84</td>
</tr>
<tr>
<td><strong>Unilateral HL</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test Positive</td>
<td>2</td>
<td>11</td>
<td>13</td>
</tr>
<tr>
<td>Test Negative</td>
<td>26</td>
<td>1</td>
<td>27</td>
</tr>
<tr>
<td>Totals</td>
<td>28</td>
<td>12</td>
<td>40</td>
</tr>
</tbody>
</table>

HL – Hearing loss

Agreement

Cohen's Kappa agreement ranged from “substantial” to “almost perfect” across the diagnostic criteria, $k = 0.80$ [95%CI 0.70, 0.89]; $p<0.001$. 
Table 5.2: Diagnostic accuracy of clinical protocols applied to automated audiometry.

<table>
<thead>
<tr>
<th>Diagnostic criteria</th>
<th>L95%CI</th>
<th>U95%CI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Normal</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.86</td>
<td>0.56</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.86</td>
<td>0.66</td>
</tr>
<tr>
<td>PPV</td>
<td>0.75</td>
<td>0.47</td>
</tr>
<tr>
<td>NPV</td>
<td>0.92</td>
<td>0.73</td>
</tr>
<tr>
<td><strong>Disabling HL</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>1.00</td>
<td>0.40</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.97</td>
<td>0.84</td>
</tr>
<tr>
<td>PPV</td>
<td>0.80</td>
<td>0.30</td>
</tr>
<tr>
<td>NPV</td>
<td>1.00</td>
<td>0.88</td>
</tr>
<tr>
<td><strong>Conductive HL</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.86</td>
<td>0.63</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.95</td>
<td>0.86</td>
</tr>
<tr>
<td>PPV</td>
<td>0.86</td>
<td>0.63</td>
</tr>
<tr>
<td>NPV</td>
<td>0.95</td>
<td>0.86</td>
</tr>
<tr>
<td><strong>Unilateral HL</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.92</td>
<td>0.60</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.93</td>
<td>0.75</td>
</tr>
<tr>
<td>PPV</td>
<td>0.85</td>
<td>0.54</td>
</tr>
<tr>
<td>NPV</td>
<td>0.96</td>
<td>0.79</td>
</tr>
<tr>
<td><strong>Overall</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.88</td>
<td>0.75</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.93</td>
<td>0.88</td>
</tr>
<tr>
<td>PPV</td>
<td>0.82</td>
<td>0.69</td>
</tr>
<tr>
<td>NPV</td>
<td>0.96</td>
<td>0.91</td>
</tr>
</tbody>
</table>

HL – hearing loss  
PPV – positive predictive value  
NPV – negative predictive value  
L95%CI – lower 95% confidence interval  
U95%CI – upper 95% confidence interval
DISCUSSION

This is the first study to examine the accuracy of pre-defined diagnostic criteria applied to automated audiometry in a telehealth context. The aim of the study was to examine whether applying standard diagnostic criteria applied to automated audiometry is an accurate method of identifying a number of key audiometric characteristics which can guide further referral to specialist services. The primary use of this method would be to increase access to audiology services for patients in remote areas as part of a teleaudiology program or increase efficiency of audiology practices in urban areas by facilitating patient triaging. The overall agreement ranged from substantial to almost perfect and was highly significant ($p<0.001$), indicating that pre-defined diagnostic criteria may be an accurate and effective method of identifying hearing disorders using automated audiometry. However, there was a wide range of variance in agreement (0.70 - 0.89) and further research beyond a pilot study with additional participants will enable a more precise range of variance to be calculated. Sensitivity, specificity, positive and negative predictive values appeared to be high across conditions, indicating that automated audiometry may be an accurate diagnostic measure when using pre-defined diagnostic protocols. However, there is no direct comparison available for the sensitivity and specificity values obtained in this study so we are unable to ascertain
whether the values obtained truly represent high sensitivity and specificity. Screening audiometry in unselected populations perhaps offers the closest comparison and typically has sensitivity values between 50% to 75%, specificity of 78% to 98.5%, PPV values ranging from 7.6% to 52.9% and NPV values of approximately 89-99% (Halloran et al., 2009; Botasso et al., 2015; Mahomed-Asmail et al., 2016). Whilst the results from the present study compare favourably with screening audiometry in terms of accuracy, it should be noted that the objectives, target population and disease prevalence in screening audiometry are not directly comparable to diagnostic audiometry in this study.

Limitations of this study include the low number of participants with the target hearing conditions, particularly for disabling hearing loss, which contributed to the wide confidence intervals. As a pilot study, the current results lack the statistical power to provide conclusive evidence to support adopting pre-defined diagnostic protocols. However, it demonstrates proof of concept and shows that this method warrants further investigation as it has significant potential for improving audiology service delivery. Sensitivity and specificity for identifying normal hearing (the largest clinical group) was 86%. We would therefore expect future studies to have a similar level of accuracy, albeit with more precise confidence intervals.
Role of diagnostic protocols in a telehealth framework

With over 360 million people around the world with disabling hearing loss (WHO, 2013), and a shortage of hearing health specialists (Goulios & Patuzzi, 2008), new models of diagnosis and care are required to provide equity of access to health care and to reduce the burden of disease, particularly in countries such as Australia, where the availability of specialist healthcare services is limited outside of urban centres. Automated audiometry is becoming increasingly sophisticated and has recently been validated for use without a sound treated environment (Swanepoel et al., 2015), and in a clinically heterogeneous population (Brennan-Jones, Eikelboom et al., 2016), meaning it has the potential to overcome some of the obstacles in access to hearing services. However, the results of a hearing test still need to be interpreted to determine the next step in the clinical pathway. Diagnostic protocols such as those applied in this pilot study have the potential to address this, for synchronous, asynchronous and hybrid tele-audiology services (Swanepoel & Hall, 2010). An example of a possible clinical pathway using diagnostic protocols for automated audiometry is shown in Figure 5.2. Applying these protocols to automated audiometry could enable local health workers to triage patients with ear and hearing disorders according to their type and level of hearing loss. In particular, the identification of participants with normal hearing levels will directly reduce the
workload of audiologists and enable more patients to be seen or allow more time spent on complex cases, whether synchronous or asynchronous methods are used.

**Figure 5.2:** Flowchart of the potential patient journey through a service incorporating pre-defined diagnostic criteria applied to automated audiometry.

---

**Clinical relevance & further research**

This study presents a method of identifying patients with hearing disorders that can be applied to automated audiometry within telehealth practices. Assistants currently working to facilitate both synchronous and asynchronous tele-audiology services would be able to apply, conduct and interpret these results from pre-defined diagnostic protocols using automated audiometry, reducing the time and administrative burden placed on clinical audiologists.
Classification systems do have their limitations (Clark, 1981). Classifying hearing loss by severity does not provide any medical information (conductive or sensorineural hearing loss) or an understanding of the patient’s perceived experience of their hearing loss (i.e. low-frequency versus high-frequency hearing loss) or quality of life.

The diagnostic accuracy of audiometry will often be strengthened by interpretation by an experienced audiologist. However, standard diagnostic criteria are available for identifying disabling hearing losses, conductive (versus sensorineural hearing losses) and unilateral versus bilateral losses (BAA, 2009). Disabling, conductive and unilateral hearing losses all require further investigation and management, usually from a team comprised of otolaryngologists and audiologists. Clinical protocols that adhere to these diagnostic criteria could allow basic treatment and referral decisions to be made based on results of automated audiometry conducted in a primary care or tele-audiology setting where experienced audiologists may not be available. In addition to complementing tele-audiology models of care, the diagnosis of these conditions could be an effective method of triaging appropriate referral to audiologists and otolaryngologists in secondary and tertiary care.
It is expected that there will be disparities in clinical guidelines locally, nationally and internationally which would affect the definitions used for the diagnostic protocols. Indeed, this study has incorporated guidance from a number of sources to derive the diagnostic definitions included in this study as no single audiological professional body lists universally agreed upon definitions for diagnosis of hearing loss. The selection of clinical guidelines for this study was therefore limited to those that are known to the authors and publicly available; all diagnostic criteria represented national or international guidelines. However, this study serves to provide a proof of concept and numerous iterations of pre-defined diagnostic protocols are possible for identifying target conditions (e.g. BAA, 2009; WHO, 2013). Current automated audiometers can easily accommodate different diagnostic criteria into their programming, but the use of alternative guidelines will vary the sensitivity and specificity of the diagnostic test as a result. It is therefore suggested that a formal analysis of alternative diagnostic definitions should be conducted before implementation.

Further research should also consider patient perceptions of automated audiometry testing as part of a telehealth service. Patient non-acceptance of telehealth applications in audiology is decreasing, but has ranged from 9-30% in previous studies (Swanepoel & Hall,
Quality assessment of automated audiograms is an area for future research, as the development of a validated tool for use with the KUDUwave, similar to those used for the AMTAS automated audiometer would help to identify patients for whom automated testing may not be suitable (Margolis et al., 2007).

CONCLUSION

The current study demonstrates significant potential for the use of diagnostic protocols applied to automated audiometry to complement current telehealth models of care in audiology. Pilot data demonstrates that pre-defined diagnostic protocols applied to automated audiometry asynchronously are sensitive, specific and have substantial to almost perfect agreement for identifying disabling, conductive and unilateral hearing. We have demonstrated proof of concept for this method to be used in a telehealth model of care to improve synchronous and asynchronous tele-audiology service delivery. However, further studies utilising a greater number of participants with the target conditions are now required.
REFERENCES


CHAPTER SIX
The accuracy of automated audiometry:
An individual participant data meta-analysis
Foreword to Chapter Six

Whilst the number of validation studies of automated audiometry has considerably increased in the past decade, many have been based on small participant samples. Consequently, there has been variation in the range of absolute mean differences between studies and therefore it has been difficult to get a clear indication of the expected variance when using automated audiometry.

The use of diagnostic accuracy techniques, such as those described in Chapter Five, is a recent addition to the methodology used to quantify the comparability of automated and manual audiometry. This approach can be applied to previous reports through meta-analysis, however, as individual-level data is required to calculate diagnostic accuracy, only an individual participant data meta-analysis will enable the calculation of diagnostic accuracy using sensitivity and specificity scores.

Through replicating the general approaches used to describe the validation in Chapter Five and diagnostic accuracy in Chapter Five to a meta-analysis of over 300 participants, this chapter provides the most precise estimates of automated audiometry accuracy to date.
ABSTRACT:

Objectives: To conduct an individual participant data meta-analysis to provide precise accuracy estimates for automated audiometry.

Study sample: 329 participants from six included studies.

Design: Prospective accuracy study using an individual participant data meta-analysis.

Results: Absolute mean differences between automated and manual audiometry ranged from 4.07 dB to 5.19 dB (SD 4.79-6.53) for air-conduction and 8.47 dB to 13.99 dB (SD 10.16-16.05) for bone-conduction. The sensitivity of automated audiometry for identifying participants with a bilateral hearing loss was 99.07 [95%CI 94.15, 99.95] and specificity was 92.39 [95%CI 87.52, 95.53].

Conclusions: Automated air-conduction audiometry is a valid method of obtaining hearing thresholds. However, the validity of automated bone-conduction audiometry requires further examination as the variance is currently beyond typically expected variability.
INTRODUCTION

A number of automated audiometers have recently been clinically validated, including the AMTAS (Margolis, Glasberg et al. 2010, Eikelboom, Swanepoel et al., 2013) and KUDUwave (Swanepoel, Mngemane et al. 2010; Brennan-Jones et al. 2016a). The methods used for clinical validation typically utilise a comparative or test-retest design, aimed at comparing the variation of individual hearing threshold data between manual and automated audiometers. These studies have helped determine the effects of introducing background noise, along with other variables on automated audiometry (Storey et al., 2014). However, many of these studies are limited to small sample sizes and there is limited research into clinical validity of automated bone-conduction audiometry. Only one study to date has adopted a diagnostic test accuracy approach to examine the sensitivity and specificity of automated audiometry for identifying hearing loss (Brennan-Jones et al. 2016b).

A meta-analysis of automated audiometry has previously been conducted by Mohamed et al. (2013). Their analysis identified a lack of studies examining the accuracy of bone-conduction audiometry and was limited to combining summary statistics of mean differences for hearing thresholds. Since the publication of their review, a
A number of reports of automated audiometry (including automated bone-conduction audiometry) have been published thereby warranting a re-examination of the accuracy of automated audiometry (e.g., Eikelboom et al., 2013; Storey et al., 2014; Brennan-Jones et al., 2016a).

An individual participant data (IPD) meta-analysis is considered the gold-standard of meta-analyses and have a number of advantages over a conventional meta-analysis of summary statistics. An IPD meta-analysis has the ability to examine the data in detail, produce consistent analyses across studies and avoid biases associated with the use of aggregate data in meta-regression (Simmonds et al. 2005, Riley et al. 2010).

The objectives of this IPD meta-analysis are to provide the most precise estimates of accuracy for automated audiometry to date. Examination of individual participant data also allows the diagnostic accuracy of automated audiometry to be examined. A meta-analysis of diagnostic accuracy aims to investigate whether a test is sufficiently specific or sensitive to fit its role in practice, to compare the accuracy of two or more diagnostic tests, or to investigate where existing variation in results comes from. In addition to the comparison
of thresholds for clinical validation, the utilisation of a diagnostic test accuracy approach, where defined diagnostic criteria are used to identify different types of hearing loss (i.e. sensorineural or conductive; mild or severe) is a method that has shown potential for improving asynchronous tele-audiology service delivery (Brennan-Jones et al. 2016b). Therefore, a diagnostic test accuracy meta-analysis will help inform the validity of this approach.

This meta-analysis will examine the automated Hughson and Westlake method (method of limits approach) compared with the current gold standard assessment for determining hearing thresholds, manual audiometry (ISO 8253-1:2010).

**METHODS**

*Reference test (manual audiometry)*

Manual audiometry is considered the gold standard assessment of hearing thresholds in adults and children over five years of age (ASHA, 2004) This meta-analysis examined studies using manual audiometry performed according to ISO 8253-1:2010 standards (or ISO 8253-1:1989 if conducted prior to the 2010 revision of standards) on equipment calibrated to the same standards. No restriction was placed on the manufacturer of audiometers, headphones, bone
conductors and environmental sound proofing used (provided they adhered to ISO 8253-1:2010 or ISO 8253-1:1989 standards).

Index test (automated audiometry)

No restriction was placed on the manufacturer of automated audiometers, headphones, bone conductors and environmental sound proofing used. Only index tests that adhered to an automated method of limits approach, comparable to ISO 8253-1:2010 or ISO 8253-1:1989 standards, were considered for evaluation (i.e. not screening audiometry).

Identification of study samples: search strategy

Firstly, we reviewed the search results of Mohamed et al. (2013) to identify studies that met the inclusion criteria for the present study. Secondly, we performed an updated systematic search using the strategy employed by Mohamed et al. (2013) (Table 6.1).

Table 6.1: Search Strategy, conducted in July, 2015.

<table>
<thead>
<tr>
<th>Database</th>
<th>Search strategy</th>
<th>Identifiers</th>
<th>Results</th>
<th>Limiters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pubmed/Medline</td>
<td>Reports indicating findings of automated audiological testing. Terms occurring in the title, abstract, or keywords of articles.</td>
<td>“Automatic” OR “computerized” OR “computer-based” OR “pc-based” OR “automation” OR “automated” OR “audiocan” AND “audiometry” OR “hearing measurement” OR “hearing thresholds” OR “auditory thresholds” OR “hearing assessment” OR “hearing evaluation”</td>
<td>716</td>
<td>Reports published prior to 1946 not included</td>
</tr>
<tr>
<td>Google Scholar</td>
<td>As above</td>
<td>“Automated audiometry”</td>
<td>155</td>
<td>No limits</td>
</tr>
</tbody>
</table>
Abstracts of all reports resulting from the searches were reviewed to determine whether the report complied with the inclusion criteria. Where an abstract was unavailable, the full article was reviewed. After all duplicates and unrelated reports had been excluded, the remaining reports were reviewed in full to determine whether they met the inclusion criteria.

A secondary search was used to supplement the findings of the primary search. This involved reviewing the reference lists of reports already identified for inclusion during for additional, relevant reports. Prominent authors in the field were also contacted for information on potential further sources of published or unpublished (i.e. grey literature) data that could be eligible for inclusion in this study.

The reports selected for review were scrutinized according to the audiological threshold-seeking method used (method of adjustment or method of limits), type of evaluation (diagnostic or screening), AC or BC thresholds, type of transducers and audiometer used, age and hearing status of participants and type of statistical analysis for accuracy and test-retest reliability.
Inclusion Criteria

For inclusion, studies must have examined the accuracy of diagnostic automated audiometry using a method of limits approach (i.e. Hughson & Westlake audiometric methods), compared to manual method of adjustment audiometry. Studies could be prospective, retrospective or cross-sectional in nature and not be conducted on children <5 years of age, as test procedures are often varied for children in this age group (ASHA, 2004).

Data collection

The primary investigators of all selected studies were contacted via email and asked to provide the raw data for their study: air and bone conduction thresholds for each participant for the index and reference test. The obtained data were checked for consistency and any queries were resolved by correspondence with the responsible trial investigator. As is customary in IPD meta-analyses (Veroniki, Straus et al. 2016), authorship was offered to the primary investigators supplying data in accordance with the International Committee of Medical Journal Editors (ICMJE) criteria (ICMJE, 2015).
Statistical analysis

We estimated the accuracy of automated audiometry compared to manual audiometry for determining frequency-specific hearing thresholds using a meta-analysis of mean differences (real and absolute) between hearing thresholds reported in included studies, using individual patient data. This analysis examined the frequencies of 500, 1000, 2000 and 4000 Hz extracted from the included studies as these were the frequencies most consistently measured across studies. We used analysis of variance (ANOVA) with Bonferroni’s correction applied to examine statistical significance for absolute mean differences and pair-wise comparisons for the real mean difference, in accordance with previous studies (Eikelboom et al. 2013, Brennan-Jones et al. 2016a). However, the relevance of such statistical tests has been questioned as they do not account for clinical significance which is typically deemed as a real difference between hearing thresholds of ±5 dB or absolute difference of 10 dB (ISO 8253-1:2010).

A diagnostic test accuracy meta-analysis to compare the sensitivity and specificity of the index tests of automated audiometry compared to manual audiometry was also performed. To obtain a series of two-by-two matrices for diagnostic test accuracy analysis we used individual hearing threshold data obtained from study authors and categorised hearing impairment as positive or negative in subgroups
for the type of hearing impairment (sensorineural or conductive) and level of hearing loss according to predefined criteria. We entered the resulting two-by-two matrices for the calculation of sensitivity and specificity.

*Diagnostic criteria for 2x2 matrices*

Normal hearing in this study was defined according the World Health Organisation criteria using a four-frequency average (4FA) of hearing thresholds (500, 1000, 2000 and 4000 Hz) with a better ear ≥26 dB 4FA defined as a hearing loss (i.e. normal hearing is <26 dB 4FA) (WHO, 2013). Disabling hearing loss in this study was defined according the World Health Organisation criteria of ≥41dB 4FA in the better ear (WHO, 2013).

Conductive hearing loss was defined in this study as an air-bone gap of ≥20 dB at two or more adjacent frequencies of 500, 1000, 2000 and 4000 Hz (BAA, 2009). The individual frequency bone-conduction thresholds were subtracted from the individual frequency air-conduction thresholds for both the index and reference test to obtain the air-bone gap.

Unilateral or asymmetrical hearing loss was defined in this study as a ≥20 dB difference between the left and right bone-conduction
thresholds at two or more adjacent frequencies of 500, 1000, 2000 or 4000 Hz (BAA, 2009).

**Assessment of heterogeneity**

An adapted QUADAS-2 (Quality Assessment of Diagnostic Accuracy Studies) checklist was developed to capture specific study design characteristics and patient characteristics that may account for heterogeneity in this study (Whiting et al. 2011).

**Protocol and ethics**

A prospective protocol for this study was registered with PROSPERO (CRD42015013880). All included studies were granted ethical approval from their relevant authority and all raw data obtained from study investigators was de-identified and presented with their unique ID number from the original study.

**RESULTS**

**Study Identification**

Ten of the reports identified in Mahomed et al. (2013) were eligible for inclusion in this study. We contacted the authors of seven of these studies and were unable to obtain contact details for the remaining three studies. (Sparks, 1972; Wood et al. 1973; Picard et
The updated search identified a further 16 reports that appeared relevant to the research question based on study title (see Figure 6.1, PRISMA flowchart (Moher, et al. 2009)). On abstract examination, 7 studies were excluded (as they did not examine diagnostic audiometry) and the authors of the remaining 9 studies were contacted (comprising 5 research groups). In total, authors for six published studies (comprising four research groups) responded and three of the research groups (comprising data from six separate studies (Mahomed et al. 2016; Brennan-Jones et al. 2016a; Swanepoel et al. 2010; Eikelboom et al. 2013; Margolis et al. 2007; Margolis et al. 2010) were willing and able to provide individual participant data.

**Participants**

A total of 329 participants were included. Only three included studies provided individual participant data for age (Swanepoel et al. 2010; Mahomed et al. 2015; Brennan-Jones et al. 2016a), representing less than half of the total cohort (43.5%). A separate meta-analysis for mean age was therefore not conducted. However, it is noted that at least 64 children aged between six and ten years of age reported in Mahomed et al. (2015) were included in the analysis. The prevalence of hearing loss (better ear 4FA ≥26 dB) in the study cohort was 35.2%.
Validity: Mean differences

Absolute mean differences between automated and manual audiometry thresholds ranged from 4.07 dB to 5.19 dB (SD 4.79-6.53) for air-conduction and 8.47 dB to 13.99 dB (SD 10.16-16.05) for bone-conduction. Analysis of variance (ANOVA) with Bonferroni’s correction applied was used to compare the mean difference between manual and automated audiometry (Table 6.2). For air-conduction audiometry, the mean differences in hearing thresholds determined by manual and automated audiometry were significantly different at all air and bone-conduction thresholds ($p<0.05$).

Table 6.2: Absolute mean differences (pooled analysis)
<table>
<thead>
<tr>
<th>Hearing Thresholds</th>
<th>Frequency (Hz)</th>
<th>500</th>
<th>1000</th>
<th>2000</th>
<th>4000</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Air</strong> Right</td>
<td>AMD</td>
<td>4.07*</td>
<td>5.09*</td>
<td>5.06*</td>
<td>4.84*</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>4.79</td>
<td>6.24</td>
<td>6.31</td>
<td>6.00</td>
</tr>
<tr>
<td>Left</td>
<td>AMD</td>
<td>4.27*</td>
<td>5.03*</td>
<td>4.52*</td>
<td>5.19*</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>5.62</td>
<td>6.04</td>
<td>6.53</td>
<td>6.37</td>
</tr>
<tr>
<td><strong>Bone</strong> Right</td>
<td>AMD</td>
<td>8.47*</td>
<td>10.09*</td>
<td>13.27*</td>
<td>13.99*</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>10.16</td>
<td>13.07</td>
<td>13.05</td>
<td>13.71</td>
</tr>
<tr>
<td>Left</td>
<td>AMD</td>
<td>10.64*</td>
<td>10.09*</td>
<td>13.25*</td>
<td>13.99*</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>10.17</td>
<td>12.89</td>
<td>15.69</td>
<td>16.05</td>
</tr>
</tbody>
</table>

AMD: Absolute mean difference (in dB); SD: Standard deviation; - not measured
*indicates a significant (p<0.05) difference in threshold accuracy according to a
one-way ANOVA with Bonferroni’s correction applied.

Real mean differences between automated and manual audiometry
thresholds ranged from -2.48 dB to -0.88 dB (SD 6.16-8.21) for air-
conduction and -1.80 dB to 2.32 dB (SD 13.20-19.29) for bone-
conduction. Pair-wise comparisons for the real mean difference
between manual and automated audiometry are provided in Table
6.3. For air-conduction audiometry, the mean differences in hearing
thresholds determined by manual and automated audiometry were
significantly different for all frequencies (p<0.05). For bone-
conduction audiometry, the mean differences in hearing thresholds
determined by manual and automated audiometry were not
significantly different (p>0.05), except for 4000 Hz right ear bone-
conduction thresholds (p<0.05).
Table 6.3: Real mean differences (pooled analysis)

<table>
<thead>
<tr>
<th>Hearing Thresholds</th>
<th>500</th>
<th>1000</th>
<th>2000</th>
<th>4000</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Air</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right RMD</td>
<td>-1.27*</td>
<td>-1.50*</td>
<td>-2.48*</td>
<td>-1.15*</td>
</tr>
<tr>
<td>SD</td>
<td>6.16</td>
<td>7.92</td>
<td>7.70</td>
<td>7.63</td>
</tr>
<tr>
<td>Left RMD</td>
<td>-1.77*</td>
<td>-2.27*</td>
<td>-1.34*</td>
<td>-0.88*</td>
</tr>
<tr>
<td>SD</td>
<td>6.84</td>
<td>7.53</td>
<td>7.84</td>
<td>8.21</td>
</tr>
<tr>
<td><strong>Bone</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right RMD</td>
<td>-0.92</td>
<td>-1.80</td>
<td>2.32</td>
<td>1.99*</td>
</tr>
<tr>
<td>SD</td>
<td>13.20</td>
<td>16.25</td>
<td>15.50</td>
<td>16.58</td>
</tr>
<tr>
<td>Left RMD</td>
<td>-0.47</td>
<td>-1.80</td>
<td>2.32</td>
<td>1.99</td>
</tr>
<tr>
<td>SD</td>
<td>13.59</td>
<td>15.03</td>
<td>18.69</td>
<td>19.29</td>
</tr>
</tbody>
</table>

RMD: Absolute mean difference (in dB); SD: Standard deviation; - not measured
*indicates a significant (p<0.05) difference in threshold accuracy according to pairwise t-test.

**Diagnostic accuracy – sensitivity and specificity**

Table 6.4 presents diagnostic accuracy data in a 2x2 contingency table. Not all participants had complete data to enable classification into each diagnostic criteria (i.e. missing AC thresholds for calculation of better ear 4FA or missing BC thresholds for calculation of conductive hearing loss). The sensitivity of automated audiometry varied from 99.07 [95%CI 94.15, 99.95] for identifying normal hearing to 54.05 [95%CI 44.36, 65.46] for detecting a conductive hearing loss (Table 6.5). Specificity ranged from 95.16 [95%CI 91.49, 97.36] for disabling hearing loss to 91.73 [95%CI 87.58, 94.63] for unilateral hearing loss (Table 6.5).
Table 6.4: 2x2 contingency tables of diagnostic test accuracy criteria

<table>
<thead>
<tr>
<th>Type of HL</th>
<th>HL Absent</th>
<th>HL Present</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test Positive</td>
<td>15</td>
<td>106</td>
<td>121</td>
</tr>
<tr>
<td>Test Negative</td>
<td>182</td>
<td>1</td>
<td>183</td>
</tr>
<tr>
<td>Totals</td>
<td>197</td>
<td>107</td>
<td>304</td>
</tr>
<tr>
<td>Disabling HL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test Positive</td>
<td>12</td>
<td>51</td>
<td>63</td>
</tr>
<tr>
<td>Test Negative</td>
<td>236</td>
<td>5</td>
<td>241</td>
</tr>
<tr>
<td>Totals</td>
<td>248</td>
<td>56</td>
<td>304</td>
</tr>
<tr>
<td>Conductive HL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test Positive</td>
<td>27</td>
<td>60</td>
<td>87</td>
</tr>
<tr>
<td>Test Negative</td>
<td>375</td>
<td>51</td>
<td>426</td>
</tr>
<tr>
<td>Totals</td>
<td>402</td>
<td>111</td>
<td>513</td>
</tr>
<tr>
<td>Unilateral HL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test Positive</td>
<td>21</td>
<td>46</td>
<td>67</td>
</tr>
<tr>
<td>Test Negative</td>
<td>243</td>
<td>20</td>
<td>263</td>
</tr>
<tr>
<td>Totals</td>
<td>264</td>
<td>66</td>
<td>329</td>
</tr>
</tbody>
</table>

Table 6.5: Sensitivity and specificity of diagnostic test accuracy criteria.

<table>
<thead>
<tr>
<th>Type of HL</th>
<th>L95%CI</th>
<th>U95%CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>99.07</td>
<td>94.15</td>
</tr>
<tr>
<td>Specificity</td>
<td>92.39</td>
<td>87.52</td>
</tr>
<tr>
<td>PPV</td>
<td>87.60</td>
<td>80.01</td>
</tr>
<tr>
<td>NPV</td>
<td>99.45</td>
<td>96.53</td>
</tr>
<tr>
<td>Disabling HL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>91.07</td>
<td>79.63</td>
</tr>
<tr>
<td>Specificity</td>
<td>95.16</td>
<td>91.49</td>
</tr>
<tr>
<td>PPV</td>
<td>80.95</td>
<td>68.71</td>
</tr>
<tr>
<td>NPV</td>
<td>97.92</td>
<td>94.95</td>
</tr>
<tr>
<td>Conductive HL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>54.05</td>
<td>44.36</td>
</tr>
<tr>
<td>Specificity</td>
<td>93.28</td>
<td>90.26</td>
</tr>
<tr>
<td>PPV</td>
<td>68.97</td>
<td>58.02</td>
</tr>
<tr>
<td>NPV</td>
<td>88.03</td>
<td>84.47</td>
</tr>
<tr>
<td>Unilateral HL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>70.00</td>
<td>56.99</td>
</tr>
<tr>
<td>Specificity</td>
<td>91.73</td>
<td>87.58</td>
</tr>
<tr>
<td>PPV</td>
<td>67.65</td>
<td>55.09</td>
</tr>
<tr>
<td>NPV</td>
<td>92.42</td>
<td>88.37</td>
</tr>
</tbody>
</table>
Assessment of heterogeneity & risk of bias

There was high homogeneity in the presentation of data obtained from each study as the hearing thresholds were measured and reported in accordance with relevant ISO and professional standards. The potential sources and risk of bias for studies identified using the adapted QUADAS-2 checklist are presented in Table 6.6. The main potential risk of bias were from patient selection, the inherent bias in the reference standard (manual audiometry) and conducting the reference standard with knowledge of the index test.

Table 6.6: QUADAS-2 Assessment of risk of bias

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient selection</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Unclear</td>
<td>Unclear</td>
<td>Unclear</td>
</tr>
<tr>
<td>Index test procedure</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Reference test procedure</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Study flow and timing</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

DISCUSSION

This study provides the first estimates of validity and accuracy for automated audiometry using a gold-standard individual participant data (IPD) meta-analysis approach (Simmonds et al. 2005, Riley et al. 2010). This work builds on a previous systematic review and meta-analysis of automated audiometry (Mahomed et al., 2013).
Although unable to include a number of historical reports due to difficulties in contacting authors, the present analysis benefits from the inclusion of a number of recent reports that have examined some key limitations in the evidence base according to a previous review by Mahomed et al. (2013), namely inclusion of both adult and child participants (Mahomed-Asmail et al., 2016), bone-conduction audiometry (Eikelboom et al., 2013) and unselected participants with hearing loss (Brennan-Jones et al., 2016a). Whilst participant selection methods differed across studies, the reported prevalence rate of hearing loss in the study cohort was 35.2%, similar to the 38.1% prevalence rate reported by Brennan-Jones et al. (2016a) in an unselected population recruited from a tertiary otolaryngology and audiology department. Sensitivity and specificity values are influenced by prevalence, particularly when the diagnostic test being examined is not inherently dichotomous, as is the case with hearing loss (Brenner & Gefeller, 1997). Therefore, considering the prevalence of hearing loss in the cohort, the results presented in this are more generalizable to a specialist audiological or otolaryngological setting and may differ when compared to prevalence rates in primary care or community care screening programs.
Validity

Validity, in the context of audiometry, refers to the ability of a new method to measure what it is supposed to, which means the method must be compared with a well-established one in terms of accuracy and test-retest reliability, in this case, manual audiometry. Dobie (1983) and Bland & Altman (1986) recommend the use of absolute mean differences and standard deviation as a more appropriate measure of correspondence because it provides an indicator of the expected spread in variability. Therefore, whilst we have conducted both real and absolute differences for completeness, absolute mean differences and standard deviations should be the primary outcome measure from which conclusions can be drawn regarding the validity of automated audiometry compared to manual audiometry. However, the real differences are also to be assessed as these measure the extent of a threshold offset between the two methods, and can indicate an instrument variability. If calibration procedures are carried out correctly and consistently, this difference should be low, as is this case in this study.

The findings from this IPD meta-analysis of mean differences showed that automated air-conduction audiometry is a valid method for determining hearing thresholds as the variation observed was within international guidelines and previous reports of test-retest
reliability of manual audiometry (ISO 8253-1:2010; Swnaepoel et al. 2010; Eikelboom et al., 2013). The measurement of bone-conduction thresholds requires further examination as absolute mean differences ranged from 8.47 to 13.99 dB (i.e. ±4.2 to 7.0 dB); whilst many of these differences were not statistically significant, they exceed the recommended variation according to ISO standards of 10 dB absolute difference (i.e. ±5 dB). These results were higher than some previous studies (Eikelboom et al., 2013) and lower than others (Brennan-Jones et al., 2016a). It is unclear whether the compounded variability associated with known effects related to tester bias, calibration differences, masking technique and bone-conductor placement (cf. Margolis, 2008; Margolis et al., 2013, 2016), or whether automation is also a contributor. However, the increased variation is less than one test increment outside the ISO standards (assuming test increments of 5 dB) and therefore is likely to be of limited clinical significance but further examination is needed to determine the extent of these effects. Statistically significant differences were almost universal between manual and automated air-conduction audiometry, despite the mean differences being within the clinically recommended ±5 dB ISO test-retest variation. This is likely due to comparatively larger sample size of the present study which makes it more likely to detect significant differences at smaller effect sizes compared to previous reports with smaller samples. The
presence of numerous statistically significant differences also
highlights the difference between statistical and clinical significance
when considering threshold differences as has previously been noted

Accuracy

Diagnostic accuracy methods were utilised in this study to examine
the sensitivity and specificity of automated audiometry compared to
manual audiometry. The sensitivity and specificity of automated
audiometry for categorising participants into normal hearing and
disabling hearing loss was high, indicating this is a useful utility in
triaging patients. The sensitivity for identifying conductive and
unilateral sensorineural hearing losses was below expected based
on pilot data (Brennan-Jones et al., 2016b), although specificity was
high for both criteria. The false-negative rates were high for both
conditions and the variation in bone-conduction thresholds may
explain this poor sensitivity. However, these findings were in contrast
to a recent pilot study, which gave higher accuracy estimates for
conductive and unilateral hearing loss in a population with a
comparable prevalence of hearing loss, examining the use of pre-
defined similar diagnostic criteria applied to audiometry to facilitate
automated diagnosis of hearing impairment in an asynchronous
telehealth model (Brennan-Jones et al. 2016b). The current findings
suggest such pre-defined classification criteria may be best utilised as a “red-flag” mechanism in asynchronous tele-audiology programs to prompt a detailed review of a patient’s audiogram and should not be used to provide a definitive diagnosis remotely without interpretation by a suitably qualified healthcare professional.

When used in a diagnostic accuracy capacity, the primary aim of automated audiometry is to determine whether a) a patient has a hearing loss and b) what type of hearing loss it is so that a patient can be appropriately referred for treatment. Due to the nature of hearing loss as a spectrum of impairment the primary impact of false-positives and false-negatives will be the mis-categorisation (mild or moderate; conductive or sensorineural), rather than the misdiagnosis of a hearing loss itself. The impact of false-negatives and false-positives is of particular importance when determining the type of hearing loss (conductive or SNHL). For this analysis, a low false-negative rate is of importance, as patients may have a conductive hearing loss with middle-ear pathology that goes undetected. A low false-negative rate was not achieved with the current criteria. Whilst a low false-positive rate is also clearly beneficial, these patients are less likely to undergo inappropriate treatment due to the nature of audiological practice which requires confirmation of pathology by an objective test, known as the cross-check principle (Jerger and Hayes...
1976). Therefore, adherence to the cross-check principle by including additional tests such as tympanometry alongside automated audiometry, will be important to reduce the influence of false-positives and false-negatives on patients.

Limitations

Whilst this study does present a comparatively large sample, the estimates of automated audiometry accuracy are based on only a subset of suitable, identified studies from which individual participant data could be obtained. Data from additional studies would add further power to this analysis.

CONCLUSION

Automated air-conduction audiometry is a valid method of obtaining air-conduction hearing thresholds with mean threshold variation comparable to manual audiometry. The use of automated bone-conduction audiometry requires caution and some further examination as the threshold variation exceeds ISO recommendations by ±2 dB in the present study. Automated audiometry can accurately be used to determine the severity of hearing loss to guide treatment and management decisions in asynchronous tele-audiology programs. For the diagnosis of more complex pathology, such as conductive hearing loss or significant
hearing asymmetries, pre-defined criteria can be utilised to provide a “red-flag” for further review but should not be used for a definitive diagnosis without interpretation by a healthcare professional. Future research should examine the potential sources of variance in automated bone-conduction audiometry.
REFERENCES


CHAPTER SEVEN
Asynchronous interpretation of manual and automated audiometry: Agreement and reliability

(Journal of Telemedicine and Telecare, 2016; Sept 20)
Foreword to Chapter Seven

Chapter Seven moves from statistical significance to clinical significance and examines more practical aspects regarding the use of automated audiometry in an asynchronous tele-audiology model. Chapter Seven expands from the validation and accuracy studies carried out in Chapters Four, Five and Six, which all identified differing levels of variation between manual and automated audiometry, some within accepted ISO standards, and some outside these standards. This Chapter examines whether this variation is of clinical significance in the form of an audiologist’s interpretation of the audiogram and subsequent recommendations for treatment or referral.
ABSTRACT:

Introduction: Remote interpretation of automated audiometry offers the potential to enable asynchronous tele-audiology assessment and diagnosis in areas where synchronous tele-audiometry may not be possible or practical. The aim of this study was to compare remote interpretation of manual and automated audiometry.

Materials and methods: Five audiologists each interpreted manual and automated audiograms obtained from 42 patients. The main outcome variable was the audiologist’s recommendation for patient management (which included treatment recommendations, referral or discharge) between the manual and automated audiometry test. Cohen’s Kappa and Krippendorff’s Alpha were used to calculate and quantify the intra and inter-observer agreement, respectively and McNemar’s test to assess the reliability. Audiograms were randomised and audiologists were blinded as to whether they were interpreting a manual or automated audiogram.

Results: Intra-observer agreement was substantial for management outcomes when comparing interpretations for manual and automated audiograms. Inter-observer agreement was moderate between clinicians for determining management decisions when interpreting both manual and automated audiograms. Audiologists were 2.8 times
more likely to question the reliability of an automated audiogram compared to a manual audiogram.

**Discussion:** There is a lack of agreement between audiologists when interpreting audiograms, whether recorded with automated or manual audiometry. The main variability in remote audiogram interpretation is likely to be individual clinician variation, rather than automation.
INTRODUCTION

The current, conventional method for assessing hearing in adults involves a clinician manually performing pure-tone audiometry in a suitably sound-treated environment and interpreting the results on-site (Dobie, 1983). However, higher rates of hearing loss and lower rates of intervention uptake in rural and remote populations, coupled with the shortage of audiological services in these areas has been of significant concern in both developed and developing countries (Swanepoel, Clark et al. 2010; Brennan-Jones et al. 2016a).

Telehealth solutions and the automation of audiometry have been proposed as a potential means to increase access to hearing services in underserved populations (Margolis & Morgan, 2008; Swanepoel, Clark et al. 2010; Windmill & Freeman, 2013).

A number of automated audiometers have recently been clinically validated, including the AMTAS (Margolis et al., 2010; Eikelboom et al., 2013) and KUDUwave (Swanepoel, Mngemane et al., 2010). The consensus across studies is that automated audiometry is a suitable alternative to manual audiometry (Mahomed et al. 2013), although some studies have identified an absolute mean difference of up to 10dB in air conduction thresholds compared to manual hearing thresholds, with further increased variance of approximately 15 dB for bone-conduction thresholds (Brennan-Jones 2016b). Automated audiometry has been validated for use without a sound treated
environment, and in a clinically heterogeneous population, meaning it has the potential to overcome some of the obstacles associated with testing in rural and remote areas (Storey et al. 2014, Brennan-Jones et al. 2016b).

Synchronous or “live” tele-audiology assessment, where the clinician administers and interprets the hearing assessment simultaneously, may not be possible due to connectivity issues in many rural and remote areas or due to limited clinician time (NACCHO, 2012). An alternative may be assessment with automated audiometry and remote interpretation of the results in an asynchronous telehealth model. This offers benefits such as greater coverage for difficult to access and transient populations, such as some Indigenous Australian communities, and allows for opportunistic assessments performed by local health workers which may be more efficient than scheduled appointments in some populations (Brennan-Jones et al., 2016c; Eikelboom & Swanepoel, 2016). It may also offer benefits to clinicians, reducing travel and enabling flexible working environments (e.g. working from home). The remote interpretation of test results is common and has been validated in a number of areas of medicine to facilitate telehealth services, including interpretations of retinal images (Chiang et al. 2006), radiography (Rosen et al. 1999), echocardiograms (Choi et al. 2011) and otoscopy (Eikelboom et al. 2016).
However, comparisons between remote interpretations of manual and automated audiometry have not been reported previously.

The aims of the present study were to examine the agreement and reliability of remote audiogram interpretation by audiologists and whether the potential variation in hearing thresholds introduced by automated audiometry would affect the clinical decisions made by audiologists.

METHODS

This study compares the intra and inter-observer agreement between remote interpretations of manual and automated audiometry. Results of agreement studies are intended to provide information about the amount of error inherent in any diagnosis, score, or measurement (Kottner et al., 2011).

Participants for this study were five audiologists recruited from the Ear Science Institute Australia. Audiologists who were more than three years post-qualification and maintained more than one day per week of clinical audiology practice were invited to take part in the
study. The audiologists analysed existing data collected from patients in a validation study of automated audiometry which recruited 42 adults (≥18 years) presenting with suspected hearing loss at public audiology and otolaryngology clinics at Sir Charles Gairdner Hospital, Perth, Western Australia; see Brennan-Jones et al. (2016b) for further details of the study population and test procedures.

*Procedure:* Manual audiometry, automated audiometry, tympanometry and participant demographic (age and gender) were extracted and standardised using an audiogram generator so that the manual and automated audiograms could not be distinguished (manual audiometry results were originally recorded by hand, whereas automated results are recorded electronically). The audiograms (and accompanying clinical information) were anonymised, randomised by allocation to a unique, randomly generated 4-digit number and then sorted in ascending order for interpretation.

The five audiologists participating in this experiment, blinded to whether manual or automated audiometry was used, independently interpreted the audiograms, together with the other available information (age, gender and tympanometry; a full patient history was
not available to participating audiologists). Audiologists were aware that some of the audiograms for interpretation were obtained by automated testing and that these would be compared to manual audiograms. However, they were not aware that they would be interpreting matched pairs of manual and automated audiograms. They were asked to provide a determination of: 1) the level and type of hearing loss; 2) a management plan for the patient given their audiometric results; and 3) their judgement of the reliability of the audiogram. There are no global standards for determining the type and level of hearing loss, management plan or reliability of the audiogram; these are normally the result of professional training and local clinic protocols.

For the purposes of this study, 1) the options for level of hearing loss were normal hearing, slight hearing loss or significant hearing loss requiring intervention and the options for type of hearing loss were normal hearing, sensorineural, conductive or mixed hearing loss; 2) the management options were no intervention, referral for hearing aids, referral for medical treatment and other, and 3) audiologists were given the option to comment on the reliability of the audiogram (see Figure 7.1 for an example of the audiometric interpretation criteria).
Data analysis: Firstly, Cohen’s Kappa was used to calculate the intra-observer agreement of the audiologist’s asynchronous interpretations for determining the type and severity of hearing loss, each ear separately, and the recommendation for patient management. This was to examine whether clinicians agreed with themselves for manual versus automated interpretations. Cohen’s kappa (Cohen, 1960) provides a chance-corrected index of agreement in studies employing the same two observers to score subjects on a nominal scale (Gisev et al., 2013). In this case, we have one observer scoring a subject that has been assessed using two different methods. Thus we used Cohen’s kappa to evaluate agreement between an individuals' score between the two methods of scoring. In this way
Cohen’s kappa was used to evaluate intra-observer agreement, as used in previous studies of similar subjects by Scheltens et al. (1997); McCluggage et al. (1998).

Secondly, Krippendorff’s Alpha and q-statistic was used to calculate the inter-observer agreement of remote interpretations for determining the type and severity of hearing loss, each ear separately, and the recommendation for patient management (Hayes & Krippendorff, 2007). Krippendorff’s Alpha allows comparison of agreement between multiple coders, in this case audiologists, and was therefore used to compare whether clinicians agreed with each other when interpreting both manual and automated audiograms. The q statistic represents the probability of reaching $\alpha >0.6$ (substantial agreement).

Thirdly, the main outcome variable, agreement between audiologist’s recommendation for patient management between the manual and automated audiometry test, was examined using Cohen’s Kappa and p-value for intra-observer agreement and Krippendorff’s Alpha and q-statistic for inter-observer agreement. Finally, the audiologist-related reliability of audiograms was examined using McNemar’s test; that is,
odds ratios for paired nominal data, in this case, dichotomous audiogram reliability scores.

The Landis & Koch (1977) recommendations of agreement classification were applied to Cohen’s Kappa and Krippendorff’s Alpha analyses, with $a<0$ indicating no agreement, $a=0–0.20$ indicating “Slight” agreement, $a=0.21–0.40$ indicating “Fair” agreement, $a=0.41–0.60$ indicating “Moderate” agreement, $a=0.61–0.80$ indicating “Substantial” agreement and $a=0.81–1.00$ indicating “Almost perfect” agreement.

RESULTS

Intra-observer pooled agreement for clinician’s interpretations for: (i) the level of hearing loss for manual versus automated audiograms ranged from moderate to almost perfect agreement ($a = 0.637 [95\% CI 0.452 to 0.822]; p<0.001$); (ii) the type of hearing loss ranged from fair to moderate agreement ($a = 0.407 [95\% CI 0.207, 0.613]); p <0.001; (iii) management outcomes ranged from moderate to almost perfect ($a = 0.693 [95\% CI 0.521 to 0.865]; p <0.001$), see Tables 7.1 and 7.2.
Table 7.1: Individual clinician (intra-observer) reliability for manual versus automated audiogram interpretations measured with Kappa agreement.

<table>
<thead>
<tr>
<th>Level of hearing loss</th>
<th>Type of hearing loss</th>
</tr>
</thead>
<tbody>
<tr>
<td>α</td>
<td>[95%CI]</td>
</tr>
<tr>
<td>Clinician 1</td>
<td>0.756</td>
</tr>
<tr>
<td>Clinician 2</td>
<td>0.628</td>
</tr>
<tr>
<td>Clinician 3</td>
<td>0.629</td>
</tr>
<tr>
<td>Clinician 4</td>
<td>0.543</td>
</tr>
<tr>
<td>Clinician 5</td>
<td>0.631</td>
</tr>
<tr>
<td>Clinician agreement (pooled)</td>
<td>0.637</td>
</tr>
</tbody>
</table>

Table 7.2: Individual clinician (intra-observer) reliability for manual versus automated audiogram interpretations management measured with Kappa agreement.

<table>
<thead>
<tr>
<th>Management outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>α</td>
</tr>
<tr>
<td>Clinician 1</td>
</tr>
<tr>
<td>Clinician 2</td>
</tr>
<tr>
<td>Clinician 3</td>
</tr>
<tr>
<td>Clinician 4</td>
</tr>
<tr>
<td>Clinician 5</td>
</tr>
<tr>
<td>Clinician agreement (pooled)</td>
</tr>
</tbody>
</table>

Inter-observer agreement varied from moderate to substantial in regards to the level of hearing loss, type of hearing loss and management for both automated and manual audiometry (Table 7.3). Inter-observer agreement was not significant for right or left type of hearing loss for either manual or automated audiometry. Inter-observer agreement was significant for both right and left ear interpretations of level of hearing loss using manual audiometry. For
automated audiometry however, only the right ear level of hearing loss interpretations showed significant inter-observer agreement ($\alpha = 0.68$ [95%CI 0.60, 0.76]; $q = 0.01$) (Table 7.3, Figures 7.2 and 7.3).

There was no significant inter-observer agreement for management outcomes for either manual or automated audiometry (Table 7.3 and 7.4, Figure 7.2 and 7.3).

Table 7.3: Overall clinician (inter-observer) reliability across the five remote interpretation outcomes using Kirpendorff’s alpha and probability ($q$) of reaching $\alpha >0.6$.

<table>
<thead>
<tr>
<th>Interpretation Outcomes</th>
<th>Manual audiometry</th>
<th>Automated audiometry</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\alpha$</td>
<td>[95%CI]</td>
</tr>
<tr>
<td>Right Type HL</td>
<td>0.582</td>
<td>0.519, 0.642</td>
</tr>
<tr>
<td>Left Type HL</td>
<td>0.551</td>
<td>0.490, 0.615</td>
</tr>
<tr>
<td>Right Level HL</td>
<td>0.692</td>
<td>0.593, 0.778</td>
</tr>
<tr>
<td>Left Level HL</td>
<td>0.696</td>
<td>0.569, 0.790</td>
</tr>
<tr>
<td>Management</td>
<td>0.569</td>
<td>0.505, 0.630</td>
</tr>
</tbody>
</table>
The paired samples t-test was also used to determine the correlation coefficient between manual and automated audiometry for management decisions between audiologists. Correlation was high and significant (0.823, \( p < 0.001 \)). Concurrence for discharge/no treatment was highest (93.8%), followed by audiological referral and
medical referral (respectively 81.8% and 81.4%), whereas combined medical referral was lower at 56.8% (Table 7.4).

Audiologists were more likely to question the reliability of automated audiograms than manual audiograms (OR = 2.848; p<0.001) (Table 7.5).

Table 7.4: Distribution of management decisions for remote interpretations of manual and automated audiometry (paired sample t-test corr. coeff. = 0.823, p<0.001).

<table>
<thead>
<tr>
<th>Automated audiometry management</th>
<th>Discharge</th>
<th>Manual audiometry management, n(%)</th>
<th>Aud ref</th>
<th>Med ref</th>
<th>Aud + Med</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discharge</td>
<td>45 (93.8)</td>
<td>2 (4.2) 1 (2.1) 0 (0) 0 (0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aud ref</td>
<td>2 (9.1)</td>
<td>18 (81.8) 0 (0) 2 (9.1) 0 (0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Med ref</td>
<td>6 (16.2)</td>
<td>3 (8.1) 21 (56.8) 7 (18.8) 0 (0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aud+Med</td>
<td>2 (2.0)</td>
<td>7 (6.9) 10 (9.8) 83 (81.4) 0 (0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>1 (100.0)</td>
<td>0 (0) 0 (0) 0 (0) 0 (0)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Aud = audiology; Med = medical; ref = referral)

Table 7.5: Subjective clinician rated reliability of automated versus manual audiograms using McNemar’s test.

<table>
<thead>
<tr>
<th>Audiometry reliability</th>
<th>OR</th>
<th>[95%CI]</th>
<th>X²</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinician 1</td>
<td>2.091</td>
<td>0.439, 9.961</td>
<td>0.898</td>
<td>.118</td>
</tr>
<tr>
<td>Clinician 2</td>
<td>3.222</td>
<td>0.395, 26.255</td>
<td>1.296</td>
<td>.064</td>
</tr>
<tr>
<td>Clinician 3</td>
<td>1.111</td>
<td>0.260, 4.754</td>
<td>0.020</td>
<td>.238</td>
</tr>
<tr>
<td>Clinician 4</td>
<td>7.361</td>
<td>0.689, 78.714</td>
<td>3.454</td>
<td>.006</td>
</tr>
<tr>
<td>Clinician 5</td>
<td>0.951</td>
<td>0.888, 1.019</td>
<td>0.051</td>
<td>.999</td>
</tr>
<tr>
<td>Pooled reliability</td>
<td>2.848</td>
<td>1.246, 6.508</td>
<td>6.532</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>
DISCUSSION

The present study shows that agreement for management decisions for participants remains relatively high when automated audiograms are interpreted remotely, with intra-observer agreement ranging from moderate to almost perfect. There was no significant inter-observer agreement for patient management decisions for manual or automated audiometry, indicating that management decisions differed between clinicians regardless of whether manual or automated audiometry was used.

There was significant, moderate to substantial intra-observer agreement for audiologists when determining the level and type of hearing loss using automated compared to manual audiometry. However, for inter-observer agreement, only decisions relating to the level of hearing loss were significant, but this was true for both manual and automated audiometry. This shows a lack of agreement between audiologists when determining the type of hearing loss and management decisions when interpreting manual audiometry and highlights that the main source variable in the agreement between decisions made based on remote audiogram interpretation is likely to be individual clinician variation, rather than automation.
Determination of the type and level of hearing loss was deliberately subjective in this experiment with no set quantitative criteria given to the audiologists. Whilst standard criteria exist for classifying the level of hearing loss (WHO, 2013), clinical recommendations vary and the applicability of these arbitrary cut-offs to clinical practice has been questioned (WHO, 2001; Loy & Irwig, 2004; Bucks et al., 2016). Therefore, the audiologists were presented with options of clinical significance (i.e. whether referral for intervention was necessary, and if so, what type). With a lack of definition between sensorineural, conductive and mixed hearing loss the classification of patients into these groups was diverse and agreement was poor between test method and clinicians.

Despite good intra-observer agreement and being blinded as to which audiograms were automated and which were manual, audiologists were 2.8 times more likely to question the reliability of an automated audiogram (Table 7.5). However, it has been argued that the use of automated audiometry may actually limit bias; that is, increase reliability in audiometric assessment (Margolis et al. 2013, 2015) The tester bias associated with audiometry is well-documented and many audiologists may consciously or sub-consciously alter hearing thresholds to adhere to certain rules or an expected pattern (Margolis, 2008). This perception of poor reliability may be in part
influenced by the fact that automated audiometry does not perform these adjustments. When audiometric results are shown that do not fit the expected pattern or conventions, such as where bone conduction threshold appear worse than air conduction thresholds, this may be interpreted as an unreliable assessment. However, it is also recognised that for some patients, audiometry can be a difficult task to perform and the reliability or accuracy of the audiogram can be compromised with automation (Margolis et al. 2007). Therefore, the capacity to provide synchronous tele-audiometry assessment for patient’s suspected to have poor reliability would be beneficial to complement a predominately asynchronous model (Swanepoel, Mngemane et al. 2010).

Audiometry is a key part of any test battery for the assessment of hearing and its impact on daily function and quality of life for patients. The remote interpretation of audiometry potentially offers significant efficiency and financial savings for telehealth programs and potential to improve access to services using automated technology, without the need for a clinician to travel to remote areas. However, the diverse presentation of patients with hearing loss means that audiometric assessment alone is often not sufficient and that the effect of hearing loss on daily function and quality of life should be
ascertained before clinical decisions on patient treatment and
management are made based on remote audiogram interpretation.

The findings from this study suggest that an asynchronous tele-
audiology model, where automated audiometry is performed
remotely with results forwarded for interpretation by audiologists or
medical personnel is practicable. This could facilitate much wider
coverage of ear and hearing services, streamlining metropolitan
specialist services and reducing the need for specialists to travel to
rural and remote regions to administer services. The addition of
further clinical, contextual and quality of life information may improve
both the inter- and intra-observer reliability for remote interpretations.

CONCLUSION

Remote interpretation of automated audiometry appears to be a
reliable approach for diagnosing hearing loss and identifying
appropriate interventions. Clinician interpretations vary significantly,
both for manual and automated audiograms. It is thought that this
variation is not exclusive to remote interpretation of audiometry in a
telehealth context, rather it is reflective of the diverse needs of
patients with hearing loss and a clinician’s personal experience.
The findings from this study highlight that the use of remote interpretations of automated audiometry as a method for assessing hearing ability has equivalent agreement to audiologists interpreting manual audiometry and is therefore feasible in the context of a comprehensive tele-audiological program.
REFERENCES


CHAPTER EIGHT
General Discussion
Preamble

In this concluding chapter, the main findings from each study will be reviewed and critiqued. The translational impact of the research conducted in this thesis will be discussed along with future research directions.

A review and critique of the main aims and findings of this thesis

The aim of this series of experiments was to examine the potential need, accuracy and application of automated audiometry for use in tele-audiology service models through a series of distinct but closely related studies. A range of methodological approaches were employed to address these questions including epidemiology, validation and diagnostic test accuracy analyses, meta-analyses and inter and intra-observer agreement analyses. The studies that were conducted to examine these aims are summarised, discussed and critiqued below.
8.1 The detection of hearing loss with self-report and screening audiometry, and characterising the need for hearing services in rural Australia (Brennan-Jones et al. 2016a, Chapter 3).

This study examined the role of self-report questionnaire and screening audiometry in the identification of hearing loss and provided a rationale for improving access to diagnostic audiometry in rural areas. The study built on the existing evidence base examining the accuracy of self-report measures of hearing loss.

Determining the accuracy of self-perceived hearing disability (as measured by self-report) versus hearing impairment as measured by audiometry is a nuanced task. The challenge is to weigh the importance of being able to identify a hearing impairment significant enough to be suitable for standardised interventions (audiometric hearing loss) versus identifying potential quality of life issues related to hearing that may not meet the minimum requirements for typical interventions (e.g. self-reported hearing difficulty in the absence of an audiometric hearing loss). Hearing difficulty is also not extrinsically severity of audiometric hearing loss, as outlined in the WHO ICF, hearing difficulties may be limited to certain environments and contexts, slight hearing loss may have severe impacts for some and
severe hearing losses may only have slight impacts for others (WHO, 2001). Whilst this study (and thesis) has focused on maximising the efficiency of identification of hearing loss due to the extreme demand for services globally it should be noted that where possible, hearing assessment and referral processes lend themselves to an individualistic approach.

As this study examined data collected from people who volunteered to attend a hearing screening, the results could not be used to reflect on the prevalence of hearing loss. However, the nature of the cohort allowed comparisons between rural and urban participants to be made. Data on use of hearing aids were available. The limited access to hearing services in rural and remote areas is a concern that has been raised previously (Swanepoel, Clark et al. 2010), and the risk of hearing loss associated with occupations likely to occur in rural rather than urban areas had also been established (Lower, Fragar et al. 2010). However, the present study highlighted the significant difference in the presence of hearing loss in rural populations compared to urban populations in a cohort self-referring participants, rather than targeting specific occupational groups. This study also showed that whilst self-reported hearing loss could be a useful tool in primary care, the false-positive rate was high, particularly in adults aged under 60 years of age, meaning it could
not be effectively used as a general screening tool in this population.

This was in contrast to a number of studies using population-samples which had shown higher specificity scores and lower sensitivity scores (Nondahl, Cruickshanks et al. 1998, Sindhusake, Mitchell et al. 2001, Swanepoel 2013). This study contributed to a more complete understanding of the benefits that automated audiometry and tele-audiology services could provide in rural areas in Australia and beyond.

The strengths of this study were its population sample which was fairly equally balanced between urban and rural participants and the use of the Accessibility/Remoteness Index of Australia (ARIA) classification codes to define rurality of participants (AIHW 2004). The availability of additional data relating to hearing aid use also added strength to the study and enabled some novel comparisons to be made. Limitations of the study are, ironically, also the population sample, which was not a consecutive or random sample but rather a convenience sample from participants presenting for hearing screenings. This does of course limit some of the interpretations from an epidemiological perspective, particularly specificity and it can be assumed that many participants presented for hearing screening because of concerns they already had, thereby skewing their response to the self-report questionnaire. However, from a clinical or
service-based audiology perspective, this sample is reflective of a population seeking hearing services for diagnosis and treatment where appropriate. From this perspective, the study highlights the value of and the need for better provision of diagnostic audiometry services in rural areas of Australia.
8.2 Validation of automated audiometry in a clinical population and assessment of hearing loss and age as potential confounders (Brennan-Jones et al. 2016b, Chapter 4).

This study provided the primary data collection, technical and methodological foundation for the thesis. Previous validation research for automated audiometry had been subject to problems with selection bias (Swanepoel, Mngemane et al. 2010, Eikelboom et al. 2013) and the exclusion of participants with conductive hearing losses (Storey et al. 2014). The use of a consecutively recruited, heterogeneous clinical population helped address these previous methodological issues. The use of multiple manual audiometry testers and the lack of a sound-proof environment for automated testing provided the most challenging validation environment for this device (the KUDUwave) to date. The result was increased variation beyond the ISO standard which was in contrast to previous studies. However, the variation (for air-conduction thresholds) was of limited clinical significance, being only ±2.5 dB beyond the standard ±5 dB variation. At this point, potential concerns over the accuracy of the bone-conduction thresholds were raised. However, it was unclear whether automation or other confounding variables (e.g. calibration, complex masking, ambient noise levels or transducer placement, multiple audiologists undertaking the gold standard audiometry) were
responsible for this difference; this was one of the limitations of introducing so many variables into one validation study.

The influence of order effects was also a limitation of this study, as alluded to by statistically significant hearing threshold differences between manual and automated audiometry in the left ear. The KUDUwave was programmed to test the left ear first and this was not altered throughout data collection. Counter-balancing the order of presentation would likely have removed this limitation (McCall, 1973). However, this finding did highlight the potential benefit of pre-testing or control trials before formal testing commences. This is now incorporated into recent versions of the software.

Aside from these limitations, the statistical approach taken was novel and a strength of this study. The calculation of participant-level averaged absolute mean differences (called an “individual absolute mean difference” in the study) enabled the quantification of potential confounders for automated threshold accuracy such as the presence of hearing loss and age at testing. Whilst a weak linear relationship was demonstrated with age, neither age nor presence of hearing loss were statistically significant predictors of accuracy. This study added significantly to the evidence base supporting the validation of automated audiometry by expanding validation of air-conduction
testing to a clinical population with reduced selection bias,

highlighting potential variation in bone-conduction thresholds and

applying a new statistical approach to quantify potential confounders.
8.3 Diagnostic accuracy of automated audiometry using an asynchronous telehealth model (Brennan-Jones et al. 2016c, Chapter 5).

The aim of this study was to examine the potential utility of using pre-defined diagnostic protocols to identify the type (or site of lesion) of hearing loss, severity of hearing loss and whether the hearing loss was asymmetrical, as these three features are commonly used to determine onward referral to an ear or other medical specialist and/or intervention by an audiologist. Considering the slightly increased variation in hearing thresholds found in Chapter 4 (Brennan-Jones et al. 2016b), we were also interested to see if this altered the classification of participants into standard criteria for hearing loss.

Classification systems applied to audiometry have been the subject of previous efforts to standardise the interpretation and reporting of audiometric results (Margolis & Saly 2007, WHO, 2013). However, whilst Margolis & Saly (2007) have developed an automated system for audiogram classification (AMCLASS), this system is only available for AMTAS-enabledaudiometers, narrowing its utility for global tele-audiology programs (Margolis and Saly 2008). Furthermore, the simpler classification systems address only severity of hearing loss of an ear, at times with no regard to the individual,
which do not assist the practitioner in reaching a diagnosis or
deciding the next step on the referral pathway.

The results from this study showed that the use of pre-defined
diagnostic protocols applied asynchronously to automated
audiometry could provide identification of disabling, conductive and
unilateral hearing loss, with a high overall sensitivity and specificity
compared to screening audiometry and significant (“good” to “almost
perfect”) Kappa agreement (Landis & Koch, 1977).

This was the first study to examine the accuracy of these diagnostic
criteria for automated threshold audiometry. Whilst the results
appeared promising, the interpretation and impact of the results was
hampered by the small sample size and a lack of comparable studies
of sensitivity and specificity for diagnostic audiometry. However, this
study provided a novel contribution by describing the potential
application of automated audiometry for asynchronous telehealth
using diagnostic criteria that had potential to triage patients and
increase efficiency in tele-audiology programs.
8.4 Validation and accuracy of automated audiometry and establishing a global individual participant database for meta-analysis (Chapter 6).

The aims of this chapter was to provide the most precise estimates of accuracy for automated audiometry in the world to date. Following the results from Chapter 4 (Brennan-Jones et al. 2016b) that showed increased variation of automated audiometry and Chapter 5 (Brennan-Jones et al. 2016c) that provided an encouraging yet underpowered diagnostic accuracy analysis of automated audiometry, it became clear that the most efficient way to answer the questions of validity and in particular diagnostic accuracy would be through a meta-analysis of individual participant data from previously completed validity and accuracy studies. Mahomed et al. (2013) had provided a previous meta-analysis of automated audiometry, however their analysis was limited to combining summary statistics of mean differences from previous studies and they identified a lack of studies examining the accuracy of bone-conduction audiometry. In addition, a diagnostic accuracy meta-analysis cannot be conducted with summary statistics alone unless sensitivity and specificity are reported and to date only Chapter 5 of this thesis had provided a report of sensitivity and specificity for automated audiometry. With the addition of a number of validation reports of automated
audiometry that included bone-conduction results becoming available in the time since the Mahomed et al. (2013) meta-analysis, it became clear that a primary study or an individual participant data meta-analysis that could examine air and bone-conduction validity and diagnostic accuracy would be of significant benefit to the field.

The main findings from this study examining the validation of automated audiometry were mixed. This study confirmed that automated air-conduction audiometry is a valid method for determining hearing thresholds, with pooled absolute mean differences within the recommended ISO 8253-1:2010 limits. However, for bone-conduction audiometry, pooled absolute mean differences for hearing thresholds obtained through automate audiometry were outside the recommended ISO limits. Similarly, for diagnostic accuracy, this study showed that automated audiometry was highly accurate for identifying normal hearing or diagnosing disabling hearing loss. However, this high level of accuracy was not maintained for diagnosing unilateral or conductive hearing losses.

The finding that automated bone-conduction audiometry variation exceeds ISO standards was not wholly unforeseen. It was originally hypothesised that the individual participant data meta-analysis would show less variation in automated bone-conduction thresholds than
was found in Chapter 4, but the analysis showed this not to be the case. This study has clearly identified a need for further research to refine automated bone-conduction testing and identify what confounding variables are contributing to this increased variation and what portion of the variability can be independently attributed to automation. There are a number of well-reported or potential confounding variables, including tester bias, calibration differences, ambient noise levels, masking technique and bone-conductor placement. As the automated threshold seeking procedure for air and bone-conduction testing is identical, the finding that automated air-conduction audiometry can achieve acceptable levels of variation supports the argument that the increased influence of confounding variables associated with bone-conduction testing, rather than automation itself, is the source of this observed increase in absolute mean differences. It is up to future investigators to assess whether this matter warrants further examination; this present study has been crucial in providing current estimated of threshold accuracy and identifying the need for this further research.

The examination of the sensitivity and specificity of automated audiometry using diagnostic criteria for hearing loss is a unique addition to the literature. These findings allow the quantification of the accuracy of audiometry using an alternative method to threshold
mean differences and build on the concept reported in Chapter 5 of using pre-defined diagnostic criteria applied to the results of automated audiometry to diagnose or triage patient care in an asynchronous tele-audiology model. However, this study benefits from having a sufficient sample size to allow a more meaningful interpretation of the results than was possible in Chapter 5. The results showed that identifying normal hearing loss and diagnosing disabling hearing loss could be achieved with high sensitivity and specificity using automated audiometry. For more complex diagnoses such as conductive, unilateral or asymmetrical hearing losses the accuracy of pre-defined diagnostic criteria decreased. This could be due in part to variability in automated testing or patient variability.

The inclusion of diagnostic criteria in addition to analysis of mean differences of individual hearing thresholds is useful as it provides a more clinically applicable interpretation of a participant results for manual and automated audiometry testing. For instance, a number of government and health authorities will base the eligibility of a patient for services based on their 4FA (e.g. Better Start, 2012). However, whilst pre-defined diagnostic criteria are also useful in comparing accuracy for specific types and severities of hearing loss, one of the limitations of this method is that small differences between tests can result in the misclassification of participants. For instance, if a
participant has a manual audiometry 4FA of 25 dB that will be
classed as “normal hearing” and if on automated audiometry testing
that same participant has a near identical 4FA of 26 dB they will
instead be considered to have a hearing loss and be listed as a false-
positive even though the functional difference between the two
averages is negligible. The use of mean differences for individual
thresholds for quantifying accuracy is the most common method and
has obvious appeal as this is the basic unit of an audiogram.
However, threshold accuracy can vary across frequencies and does
not give an indication of the overall characteristics of a patients
hearing loss. Therefore, examination of both mean differences for
individual hearing thresholds and using diagnostic accuracy
measures to examine broader aspects such as hearing loss severity
criteria can be useful when quantifying the accuracy between manual
and automated audiometry.

One of the most challenging aspects of this study was to obtain and
standardise the individual participant data from numerous sources
around the world. Whilst individual participant data meta-analyses
are considered the gold standard of systematic reviews and
evidence-based healthcare, they are still subject to the potential
biases of the original studies. However, the findings from this study
offer the most comprehensive assessment of automated audiometry
accuracy to date and has provided the platform for an ongoing international database of accuracy studies of automated audiometry. This will be of significant benefit now and into the future. The findings from this study confirmed that automated air-conduction audiometry, whether used for the identification of hearing thresholds or the diagnosis of varying degrees of hearing loss using 4FA, was valid and accurate and highlighted the need for a further examination of automated bone-conduction audiometry accuracy.
8.5 Agreement and reliability of asynchronous audiogram interpretations: examining the effect of automation (Chapter 7).

The aim of this chapter was to examine the agreement and reliability of remote audiogram interpretation by audiologists and whether the potential variation in hearing thresholds introduced by automated audiometry would affect the clinical decisions made by audiologists, particularly when applied in an asynchronous telehealth context. This research question arose as a result of the findings from Chapters 4, 5 and 6, particularly the variation in air and bone-conduction thresholds above ISO standards observed in Chapter 4. In Chapter 4, air-conduction thresholds showed an approximate 15 dB (or ±7.5 dB) absolute difference compared to manual audiometry, outside the 10 dB (or ±5 dB) absolute difference recommended by ISO standards. It was argued that as this value was smaller than the minimum testable difference (i.e. 5 dB) allowable under the ISO standards that the clinical significance would be minimal. Another motivation for this study was that in both the studies reported in Chapter 4 and Chapter 6 a number of comparisons of hearing thresholds obtained by automated audiometry were statistically significantly different to manual audiometry, despite the absolute thresholds differences being below the ISO standard clinical variation (i.e. absolute difference). This study, therefore, aimed to examine if this statistical
significance translated to clinical significance in audiogram interpretations between automated and manual audiometry. Finally, the findings from Chapters 5 and 6 highlighted that whilst pre-defined diagnostic criteria applied to automated audiometry was an accurate method for automatically determining the level of hearing loss, this method was not consistently accurate for identifying the type of hearing loss (i.e. site of lesion) or significant asymmetries. Remote interpretation of the automated audiograms would be required to identify these more complex presentations and this study therefore sought to examine whether this would be a feasible approach.

The main findings from this study were that the method of audiometry did not significantly affect the management decisions of individual audiologists (intra-observer agreement) for remotely interpreting audiograms. On the other hand, there was no significant agreement between audiologists (inter-observer agreement) for patient management decisions for either manual or automated audiometry, indicating that management decisions differed significantly between clinicians regardless of the method of audiometry.

It is acknowledged that the management decisions made by audiologists were based solely on the clinical information provided.
and there was a lack of contextual and personal information about the patient that is often elicited in a clinical history and can be hugely important in determining what the most appropriate treatment or management for a patient may be. Considered in the context of the WHO (ICF) introduced in Chapter 2, the relevance of severity of hearing loss in lessened, as it is the degree that their hearing loss affects their activities of daily life that will be the most important factor in guiding treatment and this will vary across individuals (WHO, 2001).

The intra-observer findings highlighted that the use of remote interpretations of automated audiometry would be a valid method for assessing and managing treatment referrals in the context of an asynchronous tele-audiology program. The inter-observer findings also implied that clinician variation amongst audiologists is more likely to account for differences in remote audiogram interpretation than automation. This was a novel and useful finding that goes some way to showing that the statistical differences between hearing thresholds observed in Chapters 4 and 6 is of limited clinical significance in this context.
The limitations of this study were the absence of a comparator for intra-observer reliability. That is, for intra-observer agreement an individual audiologist's diagnosis and management decisions were compared for manual and automated audiometry. If an individual audiologist's diagnosis and management decisions were obtained twice for manual audiometry (using a randomised, anonymised method such that the audiologist was not aware that they were interpreting the same audiogram twice) then a baseline measure of intra-observer agreement would have been obtained and this could inform the comparability of intra-observer agreement between manual and automated audiometry. However, as we found substantial and significant intra-observer agreement between interpretations of manual and automated audiometry, the lack of an intra-observer baseline agreement level for manual audiometry had little impact on our interpretation of the results, but it would have been beneficial nonetheless.

Additionally, it is recognised that even small amounts of increased variation in hearing thresholds may have some clinical impact (e.g. resulting in changes to prescription settings for hearing aid amplification). However, the impact of such variation in determining the onward referral to further audiological or medical services in the
context of the present study appears negligible compared to standard clinician variation.

One result of interest was the clinician-rated reliability of audiograms and our finding that audiologists were 2.8 times more likely to question the reliability of an automated audiogram compared to a manual audiogram. The reasons for this were not determined. However, the tester bias associated with audiometry is well-documented and many audiologists may consciously or subconsciously alter hearing thresholds to adhere to certain rules or an expected pattern (Margolis 2008). As automated audiometry does not perform these adjustments it is possible that there may be a perception of poor reliability may if audiometric results do not fit the expected pattern or conventions, even if they are genuine thresholds. It is also recognised that for some patients, audiometry can be a difficult task to perform and the reliability or accuracy of the audiogram can be compromised, with automated testing being less accommodating for patients with slower reaction times (Margolis, Saly et al. 2007). This, coupled with the increased querying of automated audiometry in the present study, highlights the benefit of the availability of face-to-face or synchronous tele-audiometry services to complement asynchronous tele-audiology assessment models for patients suspected to have poor reliability. The findings
also highlighted the need for the development of standardised methods to quantify the reliability of an audiogram, whether based on reaction times, prevalence of false-positives or the use of a forced choice method (Margolis, Glasberg et al. 2010).

This was the first study to examine and report the agreement and reliability of audiogram interpretation in a hypothetical asynchronous telehealth model using automated audiometry. This study built on the research in Chapters 4, 5 and 6 by showing that statistical differences in threshold accuracy between manual and automated audiometry did not translate to clinically significant differences in diagnosis and management decisions and that the remote interpretation of automated audiograms in an asynchronous telehealth model was effective.

8.6 Translational implications

The research conducted as part of this thesis has already began to have a translational impact on clinical practice and health policy. Automated audiometry is now used routinely in metropolitan Lions Hearing Clinics (audiology clinics allied to the candidate’s host institution) in Western Australia. The use of the KUDUwave is also being trailed for diagnostic audiometry assessments in remote
Aboriginal communities in Western Australia where hearing services are extremely limited. The candidate has also been approached by the Director of the Office of Hearing Services, Commonwealth Department of Health to contribute to the development of a position paper on government policy to expand and improve the uptake and utilisation of telehealth applications in audiology.

Whilst the research contained in this thesis alone is not the sole reason for these developments, it has been a significant contributing factor. Given the specific context of this research being undertaken in Australia, where there is a large rural and remote population for whom the uptake of telehealth services offers significant health benefits and potential savings for public and private health care providers this has lead to a drive to rapidly implement research findings into clinical practice.

8.7 Future directions

This thesis has provided a firm evidence base for the need, accuracy and application of automated audiometry into asynchronous telehealth service delivery. Whilst the validation studies in this thesis contributed to an already increasing evidence base, the main advance in the field from this work has been demonstrating the feasibility of incorporating automated audiometry into asynchronous
telehealth and further work will continue to develop this approach further. However, there are numerous avenues for further research. Importantly, there needs to be more studies that provide accuracy estimates for automated bone-conduction audiometry as well as detailed studies that can better characterise the potential sources of variability in automated bone-conduction testing. There is a need for a standardised measure to quantify the accuracy of automated audiometry. There is also a new for further research in a paediatric and Indigenous Australian population (where conductive hearing losses requiring bone-conduction assessment will be more common), and an analysis of the time efficiencies and health economic benefits of automated audiometry once implemented into clinical service delivery. The continued growth and development of the international database for automated audiometry accuracy established in this thesis (Chapter 6) is a resource for further research to examine accuracy, potential sources of variability and further applications of automated audiometry and will continue to be refined and expanded. Finally, complementary applications to diagnostic tele-audiology services such as tele-fitting of hearing aids and cochlear implants have been the subject of previous research endeavours but require more refinement to enable the widespread implementation of comprehensive tele-audiology programs that can enable remote diagnosis, intervention and rehabilitation.
8.8 Conclusions and final comments

This thesis has demonstrated the need for better provision of hearing services in rural areas of Australia and how automated audiometry can be utilised in an asynchronous telehealth model to improve access to hearing services within Australia and globally. The use of automated air-conduction audiometry has been shown to be consistently accurate using a gold-standard individual participant data meta-analysis, pre-defined criteria for identifying hearing loss has been shown to be accurate and the remote interpretation of automated audiometry have been shown to be subject to the same clinician-related variables as interpretations of manual audiometry. Overall, this series of studies has significantly progressed the evidence-base to support the use of asynchronous telehealth applications applied to audiology. This approach, in conjunction with synchronous methods, offers a means of considerably improving access to diagnostic audiological services.
REFERENCES


Better Start. (2012). *Better Start Early Intervention Service Provider Panel Operational Guidelines*. Department of Families,
Community Services and Indigenous Affairs. Canberra, Commonwealth of Australia.


Appendix A – Full-text Published and *In press* Manuscripts from Studies Presented as Part of this Thesis.
Self-reported hearing loss and manual audiometry: A rural versus urban comparison

Christopher G. Brennan-Jones, BSc(Hons),1,2 Dunay S. Taljaard, AuD, PhD,1,2,3 Sophie E.F. Brennan-Jones, MBChB,1,2,4 Rebecca J. Bennett, MAud,1,2 De Wet Swanepoel, PhD,1,2,5 and Robert H. Eikelboom, PhD1,2,5

1Ear Science Institute Australia, 2Ear Sciences Centre, School of Surgery, 4School of Population Health, The University of Western Australia, 1Department of Audiology, Princess Margaret Hospital, Perth, Western Australia, Australia; and 5Department of Speech-Language Pathology and Audiology, University of Pretoria, Pretoria, South Africa

Abstract
Objective: To examine whether self-reported hearing difficulty is an accurate measure of hearing loss compared with standard hearing screening with pure tone audiometry in rural and urban communities.

Design: Convenience sampling.
Setting: Urban and rural areas of Western Australia.
Participants: A total of 2090 participants (923 men; 1165 women; 2 unknown) aged 20–100 years presenting for community-based hearing screening in urban (982) and rural (1090) areas.

Interventions: Self-reported hearing difficulty assessed with the Hearing Handicap Inventory for the Elderly – Screening questionnaire. Hearing loss defined as average hearing thresholds >25 dB in the better ear using screening audiometry conducted at 500, 1000, 2000 and 4000 Hz.

Main outcome measures: Nil.

Results: The Hearing Handicap Inventory for the Elderly – Screening was sensitive (≥60 years = 76.69%; <60 years = 71.67%) but not specific (≥60 years = 45.15%; <60 years = 49.63%) for identifying hearing loss. The <60 age group had a hearing loss prevalence of 25.6%, and a false-positive rate of 67.12% compared with a prevalence of 69.12% and false-positive rate of 29.77% for the ≥60 age group. For all ages, rural participants were more likely to have a disabling hearing loss (odds ratio 2.04 (95% confidence interval, 1.55–2.67); χ²(1) = 27.28; P < 0.001), but there were no significant differences in hearing aid uptake.

Conclusions: Patients in rural areas presenting for hearing screenings are more likely to suffer hearing loss than adults in urban areas. We suggest rural health practitioners incorporate a self-reported hearing loss questionnaire into health check-ups for adults, particularly patients aged ≥60 years due to the high prevalence of hearing loss in this group.

KEY WORDS: hearing loss, hearing screening, primary health, rural health, self-report.

Introduction

Approximately 15% of the world’s adult population has some degree of hearing loss, half of which is considered to be disabling.1 This makes hearing loss one of the top 20 causes of disease burden globally.1 Hearing loss affects interpersonal communication, psychosocial well-being, quality of life and economic independence.2 The lack of access and availability of appropriate interventions for patients, particularly those in remote and rural areas, is of significant concern.3

Screening for hearing loss involves audiological tests that are costly and are often not available in rural healthcare settings.4 Other screening methods, such as telephone-based hearing screening and the whispered voice test, have been shown to be valid methods of screening for hearing loss. However, telephone-based
screening requires initiation from the patient, and the whispered voice test is often dependent on clinician experience.\(^5,6\) Self-report questionnaires for identifying hearing loss have therefore been proposed as a cost-effective and time-efficient method of audiological screening that can be used by any health professional group.\(^7-9\)

The uptake of hearing aids by patients with hearing loss is low.\(^10\) Identification and intervention for hearing loss in the pre-retirement period (approximately 50–65 years of age) is associated with a greater uptake of hearing aids\(^11\) and improved rehabilitation outcomes.\(^5-9\)

Self-report questionnaires can therefore be useful in early identification of hearing loss, prompting early intervention and improved outcomes in those who are motivated to receive treatment and rehabilitation for their hearing loss.

The Hearing Handicap Inventory for the Elderly – Screening (HHIE-S)\(^12\) is a widely used measure of self-reported hearing difficulty and has been adapted for different age and cultural groups.\(^11\) It has been extensively psychometrically validated,\(^12\) has comparable accuracy to similar validated questionnaires\(^14\) and higher accuracy than a single question for diagnosing hearing impairment.\(^15\)

This study was designed to evaluate the accuracy of this self-report measure for determining hearing loss among self-referring patients compared with audiometric screening, and to examine the association between the uptake of hearing aids and location (rural or urban).

**What is already known on this subject?**
- Access to treatment for hearing loss might be limited in rural and remote areas.
- Intervention for hearing loss in the pre-retirement period is associated with better patient outcomes.
- Early identification of hearing loss in primary care using the standard audiometric screening can be costly, requiring specialised equipment and training.

**What this study adds:**
- There is a significant difference in the presence of hearing loss in rural populations compared with urban populations in self-referring participants.
- The prevalence of self-reported hearing loss in adults aged over 60 years is extremely high in a self-referring population.
- Self-report measures are cost-effective methods of screening for hearing loss in primary care and of identifying patients, but they lack accuracy compared with audiometric screening.

were 182 potential participants who declined permission for their data to be used, 5 who did not complete the audiometric screening, 11 who did not provide a date of birth and 49 who were under the age 20 years; leaving a cohort of 2090 participants. For the purposes of this study, participants were divided into two groups at 60 years of age, broadly representing a pre-retirement group.

**Materials and procedures**

All participants completed the HHIE-S, a conventional audiometric hearing screening and an additional questionnaire including items on hearing aid use and outcomes designed by the investigators. HHIE-S scores of >8 indicated a hearing impairment. Ears were examined by otoscopy to ensure there was no obstruction of the ear canal prior to screening audiometry, which was conducted with a GSI 17 audiometer (GrassonStadler, Eden Prairie, MN, USA) and Peltor 211 headphones (3M, St Paul, MN, USA), calibrated annually. Screening was conducted by trained, non-specialist volunteers in a sound-treated booth on a bus modified for mobile hearing screenings. Hearing screening commenced at 25 dBHL. Participants 'passed' if they responded to low intensity sounds at 500, 1000, 2000 and 4000 Hz in both ears. If they did not respond, intensity was increased in 5 dBHL steps until the participant responded. Order of testing left or right ears first was not prescribed, and the order of testing frequencies was 500, 1000, 2000 and 4000 Hz.

**Data analysis**

Hearing screening and HHIE-S data, together with age, gender, location and hearing aid use, were analysed in SPSS v21 (IBM Corp, New York, NY, USA). Hearing...
loss was defined as mild (best ear four-frequency average >25 dB) or disabling (best ear four-frequency average >40 dB) across the frequencies of 500, 1000, 2000 and 4000 Hz, according to the World Health Organization criteria. Accessibility/Remoteness Index of Australia (ARIA) classification codes were used to identify an urban (ARIA code: 0) or rural/regional address (ARIA codes: 1, 2, 3, 4) according to postcode.

Results

Participant profile

Of the 2090 participants in the study, 982 (46.99%) resided in the metropolitan area of Perth and 1090 (52.15%) in the regional, rural or remote areas of Western Australia according to ARIA classifications; 18 (0.86%) people did not provide a valid postcode for ARIA classification. More women (n = 1165) than men (n = 923) were included; the gender of two people was not reported. The mean age of participants was 60.0 years (standard deviation 13.93, range 20.3–100.32 years), with no statistically significant difference between the gender balances or age of urban versus rural participants (independent sample t-tests; P > 0.05). Figure 1 shows the distribution of participants according to age, gender and hearing loss.

Accuracy of HHIE-S self-report questionnaire

The sensitivity of the HHIE-S questionnaire was 76.69% (95% confidence interval (CI), 73.33–79.75), with specificity of 45.15% (95% CI, 40.29–50.01) for identifying mild hearing loss in adults aged ≥60 years. For adults aged <60 years, the sensitivity of the HHIE-S was 71.67% (95% CI, 65.35–77.27) and specificity was 49.63% (95% CI, 45.80–53.46). The <60 group had a false-positive rate of 67.12% (95% CI, 62.82–71.16) compared with 29.77% (95% CI, 26.57–33.19) for the ≥60 group; the false-negative rate for the <60-year-old group was 16.42% (95% CI, 13.01–20.48) and for the ≥60 years group was 54.85% (95% CI, 49.91–59.71) (Table 1).

The prevalence of audiometric hearing loss in participants aged ≥60 years was 69.12% (95% CI, 66.32–71.79). Participants aged <60 years had a lower prevalence (25.60% (95% CI, 22.82–28.59)) (Figure 1). There was an increased risk of having an HHIE-S score >8 associated with participants from rural areas com-

FIGURE 1: Prevalence of hearing loss (>25 dBHL) according to age and gender in participants attending the Lions Hearing Bus. (□) Men, (□) women, (□) all.

TABLE 1: Two-by-two tables showing accuracy of the HHIE-S for identifying hearing loss in adults

<table>
<thead>
<tr>
<th>Hearing loss criteria</th>
<th>&lt;60 Years of age</th>
<th>≥60 Years of age</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HHIE &lt; 8 (%)</td>
<td>HHIE-S ≥ 8 (%)</td>
</tr>
<tr>
<td>Normal hearing</td>
<td>336 (36.9)</td>
<td>341 (37.5)</td>
</tr>
<tr>
<td>Hearing loss (&gt;25 dB)</td>
<td>66 (7.2)</td>
<td>167 (18.3)</td>
</tr>
<tr>
<td>Totals</td>
<td>402 (44.2)</td>
<td>508 (55.8)</td>
</tr>
</tbody>
</table>

HHIE-S, Hearing Handicap Inventory for the Elderly – Screening.
TABLE 2: A comparison of prevalence and degree of hearing loss (HL) and use of hearing aid(s) (HA) in urban and rural areas (all ages)

<table>
<thead>
<tr>
<th>Hearing loss criteria</th>
<th>Urban (n = 982)</th>
<th>Rural (n = 1090)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HL (%)</td>
<td>HL + HA (%)</td>
</tr>
<tr>
<td>Mild HL</td>
<td>409 (41.6)</td>
<td>49 (5.0)</td>
</tr>
<tr>
<td>Disabling HL</td>
<td>88 (9.0)</td>
<td>29 (2.9)</td>
</tr>
<tr>
<td>Totals</td>
<td>675 (50.6)</td>
<td>105 (15.6)</td>
</tr>
</tbody>
</table>

TABLE 3: Analysis of participants with hearing loss (HL) and participants with hearing loss using hearing aids (HL + HA) in urban and rural areas (all ages)

<table>
<thead>
<tr>
<th>Hearing loss criteria</th>
<th>Rural compared with urban locality</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\chi^2(1)$</td>
</tr>
<tr>
<td>Mild HL</td>
<td>38.13</td>
</tr>
<tr>
<td>Disabling HL</td>
<td>27.28</td>
</tr>
<tr>
<td>Mild HL + HA</td>
<td>0.008</td>
</tr>
<tr>
<td>Disabling HL + HA</td>
<td>1.614</td>
</tr>
</tbody>
</table>

CI, confidence interval; OR, odds ratio.

pared with urban areas (odds ratio (OR) 1.32 (95% CI, 1.11–1.59); $\chi^2(1) = 9.51$; $P = 0.002$). Participants in rural populations also had an increased risk of mild hearing loss (OR 1.73 (95% CI, 1.45–2.06); $\chi^2(1) = 38.13$; $P < 0.001$) and disabling hearing loss (OR 2.04 (95% CI, 1.55–2.67); $\chi^2(1) = 27.28$; $P < 0.001$). There were no differences in hearing aid uptake (mild hearing loss: OR 0.93 (95% CI, 0.66–1.46); $\chi^2(1) = 0.008$; $P = 0.927$) (disabling hearing loss: OR 0.54 (95% CI, 0.84–2.20); $\chi^2(1) = 1.614$; $P = 0.204$) (Tables 2, 3).

Discussion

This study examined the prevalence of hearing loss in a cohort of self-referring adults from rural and urban areas of Western Australia. Participants in rural areas showed a significantly higher prevalence of hearing loss, but their rates of hearing aid uptake were not significantly different. There was a high prevalence of hearing loss in those attending community-based hearing screenings, and the HHIE-S questionnaire was a sensitive, but not specific, measure for detecting hearing loss for adults aged ≥60 years in this study. Approximately 70% of participants aged ≥60 years had an audiometric hearing loss. However, these prevalence figures cannot be generalised to the wider population because of the sampling method. Population-based estimates of self-reported hearing loss are 46.2% for a regional population at 60–65 years of age. Although lower than our sample, this illustrates the high prevalence of self-reported hearing loss in the general population. Although this study did not examine the causes of hearing loss, both workers and families living in rural areas are at a higher risk of occupational noise exposure through agricultural and production industries.5

Table 3 shows an increased prevalence of hearing loss in regional and rural areas, and support previous calls for the better identification of hearing loss in primary care. Identifying and treating hearing loss in the pre-retirement period is associated with an increased uptake of hearing aids and improved rehabilitation outcomes.3,11 Considering the results of this study, the routine use of questionnaires for adults aged ≥60 years in a primary care can enable the early identification of hearing loss because of the high prevalence of the condition in this age group. However, self-report hearing measures do not have comparable accuracy with audiometry, and better access to these services in rural and remote areas is therefore still needed.

For both age groups, the HHIE-S was sensitive (≥60 years = 76.69%; <60 years = 71.67%), but not specific (≥60 years = 45.15%; <60 years = 49.63%). Low specificity could result in excessive referrals for patients who self-report a hearing difficulty in the absence of an audiometric hearing loss. This study produced similar sensitivity estimates to previous studies of the HHIE-S (Sindhusake et al. 2001; 80%),17 but with lower specificity (Sindhusake et al. 2001; 76%).

The HHIE-S false-positive rate for patients aged <60 years was substantially higher (67.12%) compared with those aged ≥60 years (29.77%). False positives incorrectly diagnose a patient as having a condition (i.e. self-reported hearing loss), in the absence of the condition according to the gold standard (i.e. hearing loss confirmed by audiometry). The impact of false positives in the <60 years group is likely to put greater strain on specialist services – resulting in an increase in referrals for hearing assessment in the absence of a hearing loss. However, it can be argued that a client who self-reports difficulty can still benefit from consultation.

The <60-year-old group had a lower false-negative rate than the ≥60-year-old group, at 16.42% and
54.85%, respectively. This is expected considering the high prevalence of hearing loss in the 260 years age group. These false negatives represent patients with hearing loss according to audiometric testing, who did not self-report a hearing loss using the HHIE-S. Although these patients might not feel they are experiencing sufficient limitations on their quality of life to warrant intervention, counselling on the impact of hearing loss and the benefits of early intervention can be beneficial.

The cost-effectiveness of audiometric screening in community settings has been debated, and there is some evidence that hearing screening increases the uptake of hearing aids. However, both screened and not-screened groups in the Yueh et al. (2010) study showed a low-uptake of hearing aids when followed-up 1 year after screening.

Pronk et al. have argued for better methods of hearing screening in adults and a wider range of alternative rehabilitative options in addition to hearing aids, such as informational counselling and communication strategies. In our study, the HHIE-S showed poor specificity for identifying a clinically significant, audiometric hearing loss. This suggests that self-report measures can identify patients with listening difficulties, in the absence of audiometric hearing loss, who could benefit from these alternative rehabilitation options. This finding also highlights the need for better access to specialist hearing services in rural areas to distinguish between patients who would benefit from hearing aids, medical or surgical intervention, or communication tactics.

Although screening questionnaires can be effective in detecting hearing loss early, these methods are not sufficient for a definitive diagnosis, which requires a diagnostic audiometry assessment. These services are rarely available in rural and remote areas and result in increased time and cost to travel to the nearest specialist, or require a specialist visit the region. The advent of automated audiometers that can be operated via telehealth or local primary care staff and interpreted remotely by audiologists can serve as a means of improving access to these specialist services. This technology, in conjunction with self-report measures, has the potential to streamline the clinical pathway of patients from identification to appropriate diagnosis and intervention.

**Limitations**

As participants voluntarily presented for hearing screening, it can be assumed that many already had concerns about their hearing and were therefore not representative of the general population. Poor specificity of the HHIE-S in this study is in contrast to other studies where specificity has ranged from 70% to 95%. However, numerous population-based studies support the recommendation of regular, primary care assessments for self-reported hearing loss, whether via a single-question or validated questionnaire.

**Conclusion**

The results highlight the increased risk of hearing loss in rural areas. This study demonstrates the potential value of self-report measures for detecting hearing impairment for adults, particularly those aged ≥60 years in settings where audiometric screening equipment is not available. However, greater provision of audiometry services would allow more accurate identification of hearing loss.

**Acknowledgements**

CGBJ received research funding from the Lions Hearing Foundation (WA). We would like to acknowledge the Lions Hearing Foundations of Western Australia who provide free hearing screening across Western Australia and the volunteers who were involved in the collection of data used in this study. We would also like to particularly acknowledge the contribution of Ms Christine Smelt, Chairperson of the Lions Hearing Foundation of Western Australia, and Ms Sharon Safstrom, Manager, The Avant CENTER, Ear Science Institute Australia.

**References**


© 2015 National Rural Health Alliance Inc.


Clinical validation of automated audiometry with continuous noise-monitoring in a clinically heterogeneous population outside a sound-treated environment

Christopher G. Brennan-Jones1,2, Robert H. Eikelboom1,2,3, De Wet Swanepoel1,2,3, Peter L. Friedland1,2,4,5 & Marcus D. Atlas1,2,5

1Ear Science Institute Australia, Subiaco, Perth, Australia, 2Ear Sciences Centre, School of Surgery, The University of Western Australia, Crawley, Australia, 3Department of Speech-Language Pathology and Audiology, University of Pretoria, South Africa, 4School of Medicine, University of Notre Dame, Fremantle, Australia, and 5Department of Otolaryngology, Head, Neck and Skull Base Surgery, Sir Charles Gairdner Hospital, Nedlands, Australia

Abstract

Objective: Examine the accuracy of automated audiometry in a clinically heterogeneous population of adults using the KUDUwave automated audiometer. Design: Prospective accuracy study. Manual audiometry was performed in a sound-treated room and automated audiometry was not conducted in a sound-treated environment. Study sample: 42 consecutively recruited participants from a tertiary otolaryngology department in Western Australia. Results: Absolute mean differences ranged between 5.12–9.68 dB (air-conduction) and 8.26–15 dB (bone-conduction). A total of 86.5% of manual and automated 4FAs were within 10 dB (i.e. ±5 dB); 94.8% were within 15 dB. However, there were significant (p<0.05) differences between automated and manual audiometry at 250, 500, 1000, and 2000 Hz (air-conduction) and 500 and 1000 Hz (bone-conduction). The effect of age (≥55 years) on accuracy (p = 0.014) was not significant on linear regression (p > 0.05; R² = 0.11). The presence of a hearing loss (better ear ≥26 dB) did not significantly affect accuracy (p = 0.604; air-conduction), (p = 0.218; bone-conduction). Conclusions: This study provides clinical validation of automated audiometry using the KUDUwave in a clinically heterogeneous population, without the use of a sound-treated environment. Whilst threshold variations were statistically significant, future research is needed to ascertain the clinical significance of such variation.

Key Words: Automated audiometry; audiometry; hearing loss; teleaudiology; KUDUwave

Assessment of hearing sensitivity thresholds is one of the key tests conducted by audiologists. The methods of assessment are well defined by the modified Hughson-Westlake protocols ISO 8253-1:2010 (ISO, 2010). In a standard manual audiometry procedure, frequency-specific sound stimuli are presented to a patient and the hearing level of the stimuli is adjusted, either decreasing or increasing, according to the patient’s response or lack of response, respectively. This method is also termed a ‘method of limits’ approach and is performed according to ISO 8253-1:2010 standards on equipment calibrated to ISO 389-1:1998 (ISO, 1998) standards. In the past decade there has been an increasing interest in systems that automate these procedures (Eikelboom et al, 2013; Ho et al, 2009; Margolis et al, 2010; Swanepoel et al, 2010). Automated audiometers are not new. Georg von Bekesy (von Bekesy, 1947) was the first to describe a self-recording threshold audiometer which automatically increased or decreased sound level whilst sweeping a specified frequency test range. Whilst this technique is still in use by some, the Hughson-Westlake method is now the most common technique for performing audiometry. A number of automated audiometry systems have implemented computerized versions of the Hughson-Westlake procedure, with the first reports of this method of automation appearing more than four decades ago (Sparks, 1972).

Following the successful clinical validation of a number of automated audiometers (Eikelboom et al, 2013; Margolis et al, 2011; Swanepoel et al, 2010), and a systematic review of their accuracy (Mahomed et al, 2013), the potential scope of these devices has expanded to include full diagnostic hearing assessments for adults, encompassing masked and not-masked air and bone conduction thresholds.
In the meta-analysis conducted by Mahomed et al (2013), automated audiometry showed comparable accuracy to manual audiometry, with overall average differences of 0.4 dB (6.1 SD). However, the authors noted that there was limited data on automated bone conduction audiometry and patients with different types and degrees of hearing loss. A number of studies included in the systematic review reported the accuracy of automated audiometry on participants with normal hearing only.

The inclusion of participants without hearing loss introduces significant bias into accuracy studies (Rutjes et al, 2006). The potential for bias is clear; normal-hearing patients are known to have hearing within a certain range, thereby limiting the potential range of variation between two methods of assessment. To limit bias it is therefore essential that the accuracy of automated audiometry be examined in a population that is likely to include participants with a range of hearing threshold levels, but who are not pre-selected according to hearing status or level of impairment. The exclusion of patients with known conductive hearing impairments (e.g. Storey et al, 2014) is also a source of potential bias. These patients represent a significant part of the clinical population and it is just as important to have accuracy estimates in such cases as for patients with sensorineural hearing losses.

The inclusion of participants with normal hearing threshold levels has been a necessary and valuable step in establishing the accuracy of automated audiometry. However, the development of studies that reduce bias by examining participants from a true clinical population will provide the most valid estimates of the accuracy of automated audiometry in practice.

One of the appeals of automated audiometry over conventional manual audiometry is its potential application in teleaudiology and its use in situations where sound treated rooms are unavailable or inaccessible. Recent reports have emphasized the global shortage of audiological services, and highlighted that these shortages are not exclusive to low and middle-income countries (Windmill & Freeman, 2013). It has also been reported that patients living in rural and remote areas of developed countries are more likely to present to primary care with a self-reported hearing loss (Brennan-Jones et al, 2016). The ability to provide automated audiometric testing in the absence of a sound-treated environment has a great potential to increase service provision to low and middle-income countries, and rural and remote areas of high-income countries that do not have these facilities. At least two of the contemporary clinically available automation-capable audimeters use audiocups to provide attenuation from environmental sounds (Margolis et al, 2010; Swanepoel & Biagio, 2011), and studies have demonstrated their potential feasibility in environments that are not sound-treated (Eikelboom et al, 2013; Macleanan-Smith et al, 2013).

The device used in this study (KUDUwave 5000) has previously been validated in an environment that is not sound-treated, using its manual-mode (Macleanan-Smith et al, 2013), and in a controlled noise environment in automated-mode (Storey et al, 2014). The present study therefore aims to address a gap in the evidence-base by combining automated testing in an uncontrolled environment that is not sound-treated, using an unselected clinical population of patients attending otolaryngology and audiology appointments at a tertiary public hospital. The potential influence of age and presence of hearing loss will also be examined to investigate the influence of patient-related variables on accuracy of automated audiometry.

**Methods**

**Participants**

Forty-two participants (20 male, 22 female) were recruited from a publicly funded combined otolaryngology and audiology clinic at Sir Charles Gairdner Hospital, Perth, Western Australia. Attendance at the clinic was free at the point of service for patients. Inclusion criteria were: 18 years or over, no known cognitive disorder, English spoken as a first language, both ears suitable for hearing assessment. Ethics approval was granted by the University of Western Australia Human Research Ethics Committee (Reference: RA/4/1/4877).

**Participant sampling and recruitment**

Patient recruitment was by consecutive series, with all patients attending the clinic offered enrolment in the study, subject to inclusion criteria. Recruitment was not based on presenting symptoms (except where they contra-indicated audiological assessment) or results from previous audiometry. No incentives were given to participants involved in the study.

**Data collection**

Data collection was prospectively designed. The order of test administration was not randomized. Five patients had the index test administered prior to the reference test, and all other participants (n = 37) received the reference test first.

**Test methods**

**Reference test: Manual audiometry**

Manual audiometry is considered the gold standard assessment of hearing thresholds in adults and children over five years of age and therefore served as the reference test for this study (ASHA, 2004). The Hughson-Westlake method (i.e. ascending method according to ISO 8253-1:2010), or adaptations of this method according to local protocols, is typically used when determining hearing thresholds with manual audiometry. Manual audiometry was conducted within a sound-treated room (mean ambient noise level 37 dBA) using an Acoustic Analyser AA30 audiometer (Starkey Hearing Technologies; Minnesota, USA), calibrated to ISO 389-1:1998; and TDH-39P (Telephonics; North Carolina, USA) supra-aural headphones and Radioear B-71 bone-conductor (Radioear Corp.; Pennsylvania, USA), calibrated to ISO389-3:1994. The bone-conductor was placed on the patient’s mastoid for manual testing. Patient history, otoscopy, and tympanometry using a GSI 38 Auto Tymp (Grason-Stradler; Minnesota) preceded audiometry testing.

**Index test: Automated audiometry**

Automated audiometry was conducted using the KUDUwave (eMoyoDotNet; Pretoria, South Africa) a mobile Type 2B screening, diagnostic, and clinical audiometer (IEC 60645-1/2) using the ascending method according to ISO 8253-1:2010. A key advantage of the KUDUwave audiometer is its double attenuation via use of insert earphones and circumaural earcups and its use of continual noise monitoring, which pauses audiometric testing if ambient noise
levels exceed prescribed limits, enabling accurate testing down to 0 dB with an ambient noise level of up to 59 dB SPL. The mean ambient noise level when there was no outpatient clinic in progress was measured at 46 dBA. Placing insert earphones down to the bony part of the ear canal also reduces the occlusion effect allowing for bone-conduction evaluation with occluded ears using insert earphones (Slevin et al, 2000; Swanepoel & Biagio, 2011). However, not removing the insert earphone is a limitation to the technique as insertion down to the bony portion of the ear canal cannot be confirmed or guaranteed. If the contralateral insert earphone is removed, this can adjust for the occlusion effect, however it also means losing some attenuation that is added by the insert. The insert earphone frequency response approximated that of the ER3A within 1 dB across test frequencies. This allowed for the use of the international insert earphone standard (ISO 389-2:1994) for calibration. These features make the KUDUwave especially suited for use without a sound-treated environment, making it appropriate for use in rural, remote, or community settings, where the availability of a sound-treated environment for testing is unlikely. The audiometry procedures were automated and recorded on a laptop using the eMOYO (v3.6.7) interface developed by eMoyoDotNet. Whenever the difference between the air conduction thresholds in the test and non-test ear was 75 dB or more at frequencies ≤1000 Hz and 50 dB or more at frequencies >1000 Hz, air conduction thresholds were masked according to current guidance (Munro & Agnew, 1999; Edwards, 2010). A masking level of 30 dB above the air conduction threshold of the non-test ear was used. Bone conduction thresholds (using a B-71 bone oscillator; Kimmetrics, Smithsburg, USA) were determined with continuous masking in the contralateral ear. A continuous masking level of 20 dB above the air conduction threshold of the non-test ear was used. Testing took place in a quiet room that was not sound treated (mean ambient noise level when there was no outpatient clinic in progress was 46 dBA). The researcher gave continuous masking level of 20 dB above the air conduction threshold of the non-test ear was used. Testing took place in a quiet room that was not sound treated (mean ambient noise level when there was no outpatient clinic in progress was 46 dBA). For each participant, the researcher gave a detailed instruction, placed the insert earphones, bone-conductor, and headset on the participant and monitored the progress of the test in case of malfunction or patient discomfort.

**Definitions**

Hearing thresholds were presented in dB hearing level (dBHL). Participants were tested at air conduction frequencies of 250, 500, 1000, 2000, 4000, and 8000 Hz, and bone-conduction frequencies of 500, 1000, 2000, 4000 Hz for both the reference test and index test. The audiologists administering the reference test obtained hearing thresholds at additional frequencies for participants as clinically indicated; however, these additional thresholds were not examined in this analysis as in most cases no corresponding threshold from the index test was available.

The index test had lower maximum sound levels compared to the reference test (KUDUwave limits for (1) air conduction were 95 dB for 250 Hz, 100 dB for 500, 1000, 2000, and 4000 Hz, and 90 dB at 8000 Hz; for (2) bone-conduction: 55 dB at 500 Hz and 70 dB at 1000–4000 Hz). In cases where no response was recorded because the index test reached its maximum testable limits at a lower level than the reference test, the hearing threshold level of the reference test was corrected to the maximum output level of the index test.

**Test Procedure**

The reference test (manual audiometry) was administered by tertiary-qualified clinical audiologists (five clinical audiologists were involved in administering the reference test throughout the study). The audiologists were all registered with the Audiological Society of Australia. Interpretation of the reference test was conducted by the clinical audiologist responsible for the patient’s care. Automated audiometry was administered by researchers involved in the project. The time interval between the reference test and the index test being conducted was less than 60 minutes for all participants, as patients proceeded directly to the next test, or after a short break if requested.

**Blinding**

The audiologist administering the reference test was blinded to the results of the index test. The researcher administering the index test was not blinded to the results of the reference test as the index test was automated and therefore could not influence the results. Other information available to the audiologist and researcher were a clinical history, and a combination of tympanometry, acoustic reflexes, and speech recognition threshold testing scores, as conducted by the clinical audiologist.

**Statistical Methods**

The validation analysis used air-conduction thresholds for 250, 500, 1000, 2000, 4000, and 8000 Hz and bone-conduction thresholds of 500, 1000, 2000, and 4000 Hz for both manual and automated audiometry. Mean and standard deviations were calculated for each frequency as well as real and absolute mean differences between the reference and index test hearing thresholds. Absolute mean differences are a preferable measure compared to real mean differences as absolute differences can account for positive and negative variation, whereas positive and negative variance can cancel each other out when using real mean differences (Eikelboom et al, 2013). Reference test (manual audiometry) thresholds were subtracted from index test (automated audiometry) thresholds to calculate the difference, in keeping with methodologies from similar studies (Eikelboom et al, 2013; Swanepoel et al, 2010). A paired-samples t-test and ANOVA with Bonferroni’s correction applied were used to calculate significant differences in hearing thresholds; an independent samples t-test was used for age (using 55 years of age as an arbitrary cut-point) and presence of hearing loss analysis (using better ear hearing of 40 dB ≥26 dB as a cut-point). Simple linear regression was also used for the analysis of age on accuracy of automated audiometry. Excel 2010 (Microsoft©, Washington, USA) and SPSS v21 (IBM Corp, New York, USA) were used for the analysis.

**Results**

**Participants**

The mean age of participants was 49.93 years (SD = 17.35, range of 19.33 to 92.55 years). Patients presented with a diverse range of clinical conditions, symptoms, and co-morbidities, including but not limited to: sensorineural hearing loss, tinnitus, conductive hearing loss, otosclerosis, otitis media, acoustic neuroma, Ménière’s disease, benign paroxysmal positional vertigo, perforated tympanic membrane, Eustachian tube dysfunction, ototoxic hearing loss, skull base fracture, and unilateral hearing loss, as well as pre-surgical and post-surgical assessments. Hearing loss was not always the primary complaint for participants and many had more than one ear or hearing-related symptom at the time of testing. The patients had a
wide range of hearing losses (Table 1). Patients who had incomplete assessment data on either manual or automated audiometry ($n = 8$), or had reliability questioned by the clinical audiologist ($n = 4$), were not included in the analysis (Figure 1).

**Accuracy of KUDUwave automated audiometry**

Summary tables of mean absolute and real differences are presented in Tables 2 and 3 respectively. The range of absolute mean differences for all air-conduction thresholds was 5.12 to 9.68 dB (SDs 5.17 to 9.59 dB), and for bone-conduction was 8.26 to 15.00 dB (SDs 5.17 to 6.46 dB), and for all bone-conduction frequencies was 8.26 to 15.00 dB (SDs 7.44 to 10.58) (Table 2). The range of absolute mean differences for 4FA (500, 1000, 2000, and 4000 Hz) air-conduction thresholds was 5.12 to 6.98 dB (SDs 5.17 to 6.46 dB), and for all bone-conduction frequencies was 8.26 to 15.00 dB (SDs 7.44 to 10.58) (Table 2). The percentage of 4FA automated air-conduction thresholds falling within an absolute mean difference of 5 dB of the reference test was 67.8%, within 10 dB was 86.5%, and within 15 dB were 94.8% of hearing thresholds (Table 4).

Analysis of variance (ANOVA) with Bonferroni’s correction applied was used to compare the mean difference between manual and automated audiometry and these results are provided in Table 2. For air-conduction audiometry, the mean differences in hearing thresholds determined by manual and automated audiometry were not significantly different ($p > 0.05$) in the right ear, but were significantly different across all frequencies in the left ear. For bone-conduction audiometry, the mean differences in hearing thresholds determined by manual and automated audiometry were significantly different at all frequencies.

Pair-wise comparisons for the real mean difference between manual and automated audiometry are provided in Table 3. For air-conduction audiometry, the mean differences in hearing thresholds determined by manual and automated audiometry for the frequencies 4000 and 8000 Hz were not significantly different ($p > 0.05$). For bone-conduction audiometry, the mean differences in hearing thresholds determined by manual and automated audiometry for the frequencies 2000 and 4000 Hz were also significantly associated bilaterally ($p > 0.05$). All other pair-wise comparisons in Table 3 were significantly different ($p < 0.05$).

### Table 1. Ear specific 4FA level of hearing loss and cumulative percentage differences.

<table>
<thead>
<tr>
<th>Hearing level (dB)</th>
<th>Total</th>
<th>%</th>
<th>c.%#</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–10</td>
<td>8</td>
<td>9.5</td>
<td>9.5</td>
</tr>
<tr>
<td>10–20</td>
<td>22</td>
<td>26.2</td>
<td>35.7</td>
</tr>
<tr>
<td>20–30</td>
<td>17</td>
<td>20.2</td>
<td>55.9</td>
</tr>
<tr>
<td>30–40</td>
<td>15</td>
<td>17.9</td>
<td>73.8</td>
</tr>
<tr>
<td>40–50</td>
<td>7</td>
<td>8.3</td>
<td>82.1</td>
</tr>
<tr>
<td>50–60</td>
<td>4</td>
<td>4.8</td>
<td>86.9</td>
</tr>
<tr>
<td>60–70</td>
<td>8</td>
<td>9.5</td>
<td>96.4</td>
</tr>
<tr>
<td>70–80</td>
<td>1</td>
<td>1.2</td>
<td>97.6</td>
</tr>
<tr>
<td>&gt;80</td>
<td>2</td>
<td>2.4</td>
<td>100.0</td>
</tr>
<tr>
<td>Totals</td>
<td>84</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

# Cumulative percentage

*Indicates a significant ($p < 0.05$) difference in threshold accuracy according to a one-way ANOVA with Bonferroni’s correction applied.

### Table 2. Absolute mean differences, standard deviation for 42 participants for automated compared to manual audiometry.

<table>
<thead>
<tr>
<th>Frequency (Hz)</th>
<th>Hearing thresholds</th>
</tr>
</thead>
<tbody>
<tr>
<td>250</td>
<td>500</td>
</tr>
<tr>
<td>1000</td>
<td>2000</td>
</tr>
<tr>
<td>4000</td>
<td>8000</td>
</tr>
</tbody>
</table>

For air-conduction audiometry:

- **Right**
  - AMD: 7.86, SD: 6.26
  - Left: 9.17*, SD: 8.69

- **Bone**
  - Right: 10.83*, SD: 9.23
  - Left: 10.60*, SD: 7.42

For bone-conduction audiometry:

- Right: 6.79*, SD: 12.58
- Left: 8.93*, SD: 9.41

### Table 3. Real mean differences and standard deviation for automated audiometry compared to manual audiometry.

<table>
<thead>
<tr>
<th>Frequency (Hz)</th>
<th>Hearing thresholds</th>
</tr>
</thead>
<tbody>
<tr>
<td>250</td>
<td>500</td>
</tr>
<tr>
<td>1000</td>
<td>2000</td>
</tr>
<tr>
<td>4000</td>
<td>8000</td>
</tr>
</tbody>
</table>

For air-conduction audiometry:

- **Right**
  - RMD: 3.33*, SD: 9.54
  - Left: 5.60*, SD: 11.38

- **Bone**
  - Right: 6.79*, SD: 11.38
  - Left: 8.93*, SD: 9.41

### Figure 1. Flow diagram of patient recruitment and drop-outs.

- 54 patients recruited
- Patients excluded due to: incomplete test = 1
  - Device failure = 5
  - English as second language = 2
- Exclusion based on poor reliability on manual test judged by clinical audiologist (4 patients)
- 42 patients with re usable data who completed all assessments

AMD: Absolute mean difference (in dB); SD: Standard deviation; –: not measured.
*Indicates a significant ($p < 0.05$) difference in threshold accuracy according to pairwise t-test.
Table 4. Difference distribution of air-conduction hearing thresholds for mid-frequencies (500, 1000, 2000, and 4000 Hz).

<table>
<thead>
<tr>
<th>dB Diff</th>
<th>500 Hz</th>
<th>1000 Hz</th>
<th>2000 Hz</th>
<th>4000 Hz</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>right</td>
<td>left</td>
<td>right</td>
<td>left</td>
</tr>
<tr>
<td>0</td>
<td>11</td>
<td>11</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>5</td>
<td>18</td>
<td>21</td>
<td>21</td>
<td>18</td>
</tr>
<tr>
<td>10</td>
<td>6</td>
<td>13</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>15</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>25</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>30</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>35</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Totals</td>
<td>42</td>
<td>42</td>
<td>42</td>
<td>42</td>
</tr>
<tr>
<td>% within 5 dB</td>
<td>69.0</td>
<td>57.1</td>
<td>73.8</td>
<td>66.7</td>
</tr>
<tr>
<td>Total % ≤5 dB</td>
<td>67.8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% within 10 dB</td>
<td>83.3</td>
<td>88.1</td>
<td>88.1</td>
<td>87.5</td>
</tr>
<tr>
<td>Totals % ≤10 dB</td>
<td>86.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>% within 15 dB</td>
<td>98.4</td>
<td>95.2</td>
<td>95.2</td>
<td>89.5</td>
</tr>
<tr>
<td>Total % ≤15 dB</td>
<td>94.8</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 2. Scatterplot showing individual absolute mean differences (AMD) for air-conduction between automated and manual audiometry against participant age at testing \( (R^2 = 0.11) \).

Age differences and presence of hearing loss

To calculate the effect of age differences and the presence of a hearing loss on the accuracy of automated audiometry each participant had an individual average absolute mean difference (individual AMD) calculated using frequencies tested for air-conduction (250–8000 Hz) and bone-conduction (500–4000 Hz) on both manual and automated audiometry. This created a summary score of audiometric variance between manual and automated audiometry for each participant to enable analysis. Sixteen participants had a hearing loss of ≥26 dB in both ears. Bilateral hearing loss was not significantly associated with increased variation in individual AMDs in thresholds between automated and manual audiometry for air-conduction \( t(40) = -0.523; p = 0.604 \), or bone-conduction \( t(40) = 1.251; p = 0.218 \). The mean age of participants with a bilateral hearing loss was 56.18 (SD 18.93) compared to 46.08 (SD 14.43) for those without a hearing loss, this difference was marginally not significant \( (\beta = 10.10 [95% CI -0.706, 20.913]; p = 0.066) \).

A statistically significant difference was found between age \((≥55 \text{ years, } n = 14)\) and hearing threshold accuracy for air-conduction \( t(40) = 1.599; p = 0.014 \), but not for bone-conduction \( t(40) = 1.334; p = 0.190 \). However, whilst linear regression showed a slight upward trend of increased individual AMDs (i.e. decreased accuracy) with age \( (R^2 = 0.11) \), the relationship was not statistically significant when analysed independently \( (\beta = 0.019 [95% CI -0.037, 0.074]; p = 0.504) \) (See Figure 2) or once adjusted for presence of hearing loss \( (\beta = 0.025 [95% CI -0.034, 0.083]; p = 0.398) \).

Excluded participants

Details of excluded participants are provided in Table 5. The mean age is slightly higher than the included population and there were equal numbers of normal hearing and hearing-impaired participants excluded.

Discussion

This study presents data relating to the accuracy of automated audiometry in a clinical population, using consecutive series recruitment. The present study also examined a range of patient-related factors that may affect accuracy between manual and automated audiometry. The study cohort represents a wide range of type and severity of hearing losses, which has been highlighted as a key limitation of previous studies (Mahomed et al, 2013).

Accuracy

According to the current ISO standard 8253-1:2010, the standard variability for determining a hearing threshold level at frequencies below 4000 Hz is 4.9 dB, in a sound-treated environment, without masking and assuming no other uncertainties. To account for uncertainties, the standard acceptable variability in audiometry is an absolute difference of 10 dB, representing the typical ±5 dB test-retest criteria that is practised widely by audiologists and present in most audiological standards (ASHA, 2004). Participants in this study had their automated audiometry thresholds tested in a room that was not sound-treated, whilst the manual audiometry testing was performed in a sound-treated environment, potentially introducing further accuracy variation from previous studies (Maclellan-Smith et al, 2013; Storey et al, 2014). The placement of the bone-conductor also differed, with mastoid placement used for the reference test and forehead placement used for the index test; and ambient noise levels for the reference test also exceeded the recommended ISO 8253-1:2010 standard. These factors may therefore have contributed additional variation between the reference and index tests.

Therefore, a number of potential influences may have affected variability for this study, with standard test variability, inter-tester differences for manual audiometry, calibration differences, and the test environment all likely to influence variation in addition to that caused by automation. Considering the compounded variation of these variables, the variability due to automation appears acceptable, with 86.5% of four-frequency thresholds (500, 1000, 2000, and 4000 Hz) within the accepted ISO absolute variation of 10 dB and 94.8% of participant’s thresholds within a further 5 dB of this (i.e. absolute mean threshold difference within 15 dB, real mean difference of ±7.5 dB). This shows that for approximately 95% of participants in this study the additional variation introduced was within an additional 5 dB absolute difference (±2.5 dB relative difference) of ISO standards. It is also comparable to previous
clinical validation studies of the KUDUwave which showed 91% accuracy (Swanepoel et al, 2010) and 92% accuracy (Storey et al, 2014) of obtaining hearing thresholds within a 10 dB absolute difference in sound-treated and 40 dB A multi talker background noise environments.

Whilst the additional 5 dB variation above the ISO standard may be considered low once the confounding factors are accounted for, the accuracy of automated audiometry in this study is lower than previous studies which have fallen within the ISO standard threshold variation limits (Eikelboom et al, 2013; Margolis et al, 2011; Swanepoel et al, 2010). Use of the same KUDUwave audiometer in manual-and automated mode, as in Swanepoel et al, 2010, eliminates the calibration differences that were introduced in the present study, and could easily account for some of the increased differences in hearing thresholds. The use of sound-treated environments in previous studies (Eikelboom et al, 2013; Margolis et al, 2011; Swanepoel et al, 2010), or artificial background noise environments (Storey et al, 2014) can also help account for the slightly better accuracy estimates in these studies. This study therefore provides clinical validation and accuracy levels for automated audiometry in environments with variable noise levels, typical of an outpatient clinic or health professional’s office.

There were a number of outlying hearing thresholds in the study cohort, with six individual threshold differences of 30–35 dB occurring across five different participants, which may have skewed mean differences given the size of our study sample. We have found no clear reason for these outliers but future studies with increased sample sizes may be able to better account for such variation. Whilst previous studies have excluded participants as outliers (Storey et al, 2014), the outliers in the present study were specific to thresholds, not individuals themselves. Inter and intra-tester variability scores were not possible to assess in this study, but previous studies have indicated that the average level of inter-tester variation can be 2.3–6.0 dB for air-conduction thresholds and 2.9–7.9 dB for bone-conduction thresholds (Margolis et al, 2010). This does not include clinician-specific variation around certain frequencies such as 4000 Hz (Margolis et al, 2013), which may also have increased variation in the results. Data from participants with an outlying threshold have not been removed as it could be argued that their results give a clearer picture of the potential issues facing automated audiometry, namely identifying participants who would benefit from manual audiometry due to patient-related factors that may potentially affect reliability, such as memory, attention, reaction times, and physiological aspects of ageing (e.g. ear canal structure) (Landry & Green, 1999).

To date, the analysis of results comparing automated and manual devices has been predominately descriptive; here we have included analysis of the pair-wise relationship between automated and manual audiometry results in conjunction with an analysis of variance approach. We have also included independent tests of the effect of age, presence of hearing loss on accuracy using a multivariate regression model and used simple linear regression to examine the influence of age on accuracy (Figure 2). The results from these statistical tests are mixed. For the pair-wise tests of automated versus manual audiometry accuracy, the low and mid frequencies for both air and bone conduction showed statistically significant differences with one another. However, at high frequencies for air conduction (4000 and 8000 Hz) and bone conduction (2000 and 4000 Hz) there was no significant statistical differences between the two testing methods. These results from pair-wise tests at high-frequencies were unanticipated. There is a recognized positive skew in manual thresholds referred to by Margolis et al (2015) as the ‘Good enough’ bias which is believed to be the result of manual testers not acquiring accurate thresholds below 0 dB because this is deemed to be a sufficient level of hearing. Automated audiometry does not have this basis and may therefore introduce variation. There are also recognized calibration discrepancies at 4000 Hz for bone conduction that may account for additional variation, but in this case, the influence of these differences were not significant on pair-wise analysis (Margolis et al, 2013). For the ANOVA analysis, a significant difference was seen for left ear air-conduction thresholds but not right ear air-conduction thresholds. These results highlight the random error (as opposed to systematic error) associated with behavioural testing and order effects. We therefore suggest that automated audiometry include a number of pre-assessment trials to familiarize the patient with the automated procedure before the hearing threshold assessment begins.

For air-conduction, 8000-Hz thresholds presented the most variability according to absolute mean difference, although these differences were not statistically significant according to pair-wise tests. It has been established that high-frequency audiometric testing at or above 8000 Hz is more susceptible to variation from differences in the coupling of headphones or earphones and individual physiological differences, with additional variation differences of up to 10 dB (Gössing & Richter, 2003). It is therefore possible that the use of insert earphones (KUDUwave) compared to manual audiometry (supra-aural headphones) may have introduced additional variation at higher frequencies.

Age was examined as a potential source of variability in this study. We found statistically significant differences in threshold accuracy between manual and automated audiometry for participants aged ≥55 years using air-conduction audiometry when using a t-test approach. However, this effect was not significant using linear regression or once adjusted for the presence of hearing loss. Therefore, any effect of age on threshold accuracy using automated audiometers (see Figure 2) was either very weak or non-existent (i.e. a Type I error or ‘false-positive’). However, future studies examining this variable would be beneficial. Despite the variation that may have been introduced by a heterogeneous clinical population, we detected no significant association between the presence of hearing loss and accuracy of automated audiometry.

### Table 5. Data from excluded participants (XP).

<table>
<thead>
<tr>
<th>Participant</th>
<th>Gender</th>
<th>Age (years)</th>
<th>4FA (dB)</th>
<th>Reason for exclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>XP1</td>
<td>M</td>
<td>73.59</td>
<td>70.00</td>
<td>Spoke English as second language</td>
</tr>
<tr>
<td>XP2</td>
<td>F</td>
<td>49.62</td>
<td>13.75</td>
<td>Spoke English as second language</td>
</tr>
<tr>
<td>XP3</td>
<td>F</td>
<td>25.37</td>
<td>7.50</td>
<td>Audiologist reported poor reliability</td>
</tr>
<tr>
<td>XP4</td>
<td>F</td>
<td>69.80</td>
<td>41.25</td>
<td>Audiologist reported poor reliability</td>
</tr>
<tr>
<td>XP5</td>
<td>F</td>
<td>37.86</td>
<td>18.75</td>
<td>Audiologist reported poor reliability</td>
</tr>
<tr>
<td>XP6</td>
<td>F</td>
<td>n/a</td>
<td>n/a</td>
<td>Device failure</td>
</tr>
<tr>
<td>XP7</td>
<td>F</td>
<td>61.23</td>
<td>50.00</td>
<td>Audiologist reported poor reliability</td>
</tr>
<tr>
<td>XP8</td>
<td>M</td>
<td>61.83</td>
<td>n/a</td>
<td>Incomplete test</td>
</tr>
<tr>
<td>XP9</td>
<td>F</td>
<td>50.85</td>
<td>n/a</td>
<td>Device failure</td>
</tr>
<tr>
<td>XP10</td>
<td>F</td>
<td>21.05</td>
<td>n/a</td>
<td>Device failure</td>
</tr>
<tr>
<td>XP11</td>
<td>F</td>
<td>56.38</td>
<td>n/a</td>
<td>Device failure</td>
</tr>
<tr>
<td>XP12</td>
<td>M</td>
<td>75.72</td>
<td>n/a</td>
<td>Device failure</td>
</tr>
<tr>
<td>Mean age</td>
<td></td>
<td>53.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SD age</td>
<td></td>
<td>18.46</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
It could be argued that whilst statistically significant differences have been identified at certain frequencies, the variation in hearing thresholds may be of minimal clinical significance. Whilst the absolute mean variation in thresholds exceeded the ISO standard by 5 dB in this study, this is still the minimal measurable difference in most conventional audiometers and the clinical implications of the difference may be similar to equivalent inter-tester, environmental or calibration differences. Future research should focus on isolating key variables that increase threshold variation as optimal test conditions and patient factors deteriorate and investigate the clinical, rather than statistical, significance of audiometric variation and the effect on audiometry interpretation due to automation.

Conclusion
This study has described the clinical validation of automated audiometry in an unselected, clinically heterogeneous population, without the use of a sound-treated environment and with numerous manual audiometry testers, potentially introducing a high degree of inter- and intra-tester variability. Considering this, the difference in hearing thresholds is low, with 86.5% of 4FAs within 10 dB and 94.8% within 15 dB. This study did however reveal that in these least optimal conditions for automated audiometry, the majority of automated hearing thresholds were statistically different to manual thresholds, with the exception of high frequency air-conduction (4000 and 8000 Hz) and bone-conduction (2000 and 4000 Hz) frequencies, and this should be considered when interpreting audiograms produced via this method. However, whilst this variation was statistically significant, future research is needed to ascertain the clinical significance of such variation.

Acknowledgements
We would like to acknowledge all the participants in this study for the generous use of their time. We would also like to thank all the medical, nursing, and support staff from the Department of Otolaryngology, Sir Charles Gairdner Hospital, and audiologists Amie Grandidge, Jonathan Constantine, Susan Tegg-Quin, Varsha Matthews, and Joanne Sahdeo.

Declaration of interest: The authors report no conflicts of interest.

References

Automated audiometry in a clinical population 513
Diagnosis of hearing loss using automated audiometry in an asynchronous telehealth model: A pilot accuracy study

Christopher G Brennan-Jones¹,², Robert H Eikelboom¹,²,³ and De Wet Swanepoel¹,²,³

Abstract

Introduction: Standard criteria exist for diagnosing different types of hearing loss, yet audiologists interpret audiograms manually. This pilot study examined the feasibility of standardised interpretations of audiometry in a telehealth model of care. The aim of this study was to examine diagnostic accuracy of automated audiometry in adults with hearing loss in an asynchronous telehealth model using pre-defined diagnostic protocols.

Materials and methods: We recruited 42 study participants from a public audiology and otolaryngology clinic in Perth, Western Australia. Manual audiometry was performed by an audiologist either before or after automated audiometry. Diagnostic protocols were applied asynchronously for normal hearing, disabling hearing loss, conductive hearing loss and unilateral hearing loss. Sensitivity and specificity analyses were conducted using a two-by-two matrix and Cohen's kappa was used to measure agreement.

Results: The overall sensitivity for the diagnostic criteria was 0.88 (range: 0.86–1) and overall specificity was 0.93 (range: 0.86–0.97). Overall kappa (k) agreement was ‘substantial’ k = 0.80 (95% confidence interval (CI) 0.70–0.89) and significant at p < 0.001.

Discussion: Pre-defined diagnostic protocols applied asynchronously to automated audiometry provide accurate identification of disabling, conductive and unilateral hearing loss. This method has the potential to improve synchronous and asynchronous tele-audiology service delivery.

Keywords
Automated audiometry, audiometry, hearing loss, tele-audiology, accuracy

Date received: 22 November 2015; Date accepted: 29 February 2016

Introduction

There is a global shortage of audiological services that is not exclusive to low and middle-income countries.¹ In developed economies, people living in rural areas have been shown to have higher rates of hearing loss with lower uptake of interventions²,³ and as many as 25% of children may require diagnostic audiometric testing to identify hearing loss associated with recurrent ear disease.⁴ Telehealth models of care utilising automated audiometers have therefore been proposed as a potential means to increase access to hearing services in underserved regions.⁵-⁷

A number of automated audiometers have recently been clinically validated, including the AMTAS⁸,⁹ and KUDUwave¹⁰ and the consensus across studies is that automated audiometry is a suitable alternative to manual audiometry.¹¹ However, these approaches rely on the availability of an audiologist to interpret the automated audiometry results. Pre-defined diagnostic protocols that can correctly identify patients with a hearing diagnosis that requires further examination may have the potential to increase the efficiency of audiological services. In synchronous models of tele-audiology, pre-assessment using diagnostic protocols applied to automated audiometry could reduce the number of audiologist-administered audiograms required per session and decrease the time to diagnosis and referral in asynchronous models of care.

Classification systems for audiometry results have been developed in an effort to standardise the reporting of severity, type and configuration of hearing loss.¹²-¹⁵

¹Ear Science Institute Australia, Western Australia
²Ear Sciences Centre, The University of Western Australia
³Department of Speech-Language Pathology and Audiology, University of Pretoria, South Africa

Corresponding author:
Christopher Brennan-Jones, Ear Science Institute, Australia Suite 1, Level 2, 1 Salvado Road, Subiaco, Western Australia 6008, Australia.
Email: chris.brennan-jones@earscience.org.au

J Telemed Telecare OnlineFirst, published on April 6, 2016 as doi:10.1177/1357633X16641552

RESEARCH/Original Article
However, these are general descriptors and not specific to a diagnosis. As such, previous studies have expressed the need for a consensus on diagnosing the site of lesion for a given hearing loss (i.e. is the hearing loss conductive or sensorineural?). There has been a lack of formal guidance on this issue. However, a number of professional bodies have issued relevant diagnostic classification criteria that will be applied in this study.

Margolis and Saly have previously focused on improving agreement between audiologists by standardising the classification and configuration of audiograms by creating the AMCLASS software which incorporates 161 audiogram classifications. However, the AMCLASS is currently only available for use with the AMTAS automated audiometry software. This study will therefore utilise simple, freely available, diagnostic guidelines issued by professional bodies that can be applied to any type of manual or automated audiometer and interpreted by tele-health facilitators with minimal training. The primary focus of the diagnostic criteria used in this study was to identify patients suitable for interventions, either medical or audiological. Disabling hearing loss, conductive hearing loss and unilateral or asymmetrical hearing loss are common reasons for referral to specialist medical and audiological professionals. However, use of diagnostic criteria can streamline the referral process and limit unnecessary medical and audiological referrals. Patients with a bilateral sensorineural hearing loss will, generally, be considered initially for audiological intervention (e.g. hearing aids) without the need for a medical referral. For patients with a significant conductive hearing loss, both medical intervention (to assess whether any hearing can be restored through surgical or non-surgical intervention) and audiological intervention will typically be required. Patients with a significant unilateral or asymmetrical hearing loss will require a medical and often a radiological referral to exclude acoustic neuroma in addition to an audiological referral. The aim of the study was to examine the accuracy of standard diagnostic criteria applied to automated audiometry to identify a number of key audiometric characteristics which can guide further referral to specialist services.

**Methods**

**Participants**

**Study population.** We recruited 42 study participants (20 male, 22 female) from a publicly-funded combined audiology and otolaryngology clinic at Sir Charles Gairdner Hospital, Perth, Western Australia. Inclusion criteria were: 18 years or over, no known cognitive disorder, spoke English as a first language, both ears suitable for hearing assessment. Ethics approval was granted by The University of Western Australia.

**Participant recruitment and sampling.** Patient recruitment was by consecutive series, with all patients attending the clinic offered enrolment in the study, subject to exclusion and inclusion criteria. Recruitment was not based on presenting symptoms (except where they contra-indicated audiological assessment) or results from previous audiometry. No incentives were offered to participants involved in the study.

**Test methods**

**Reference test: Manual audiometry.** Manual audiometry is the gold standard assessment of hearing thresholds in adults and children over five years of age and served as the reference test for this study. Manual audiometry was conducted by a audiologist within a sound-treated room (mean ambient noise level 37 dBA) according to ISO8253-1:2010 using Acoustic Analyser AA30 audiometer (Starkey Hearing Technologies, Minnesota, USA), calibrated to ISO389-1:1998 and TDH-39P (Telephonics, North Carolina, USA) supra-aural headphones and Radioear B-71 bone-conductor (Radioear Corp., Pennsylvania, USA), calibrated to ISO389-3:1994. Recording the patient’s relevant clinical history, otoscopy and tympanometry using a GSI 38 Auto Tymp (Grason-Stradler, Minnesota, USA) preceded audiometry testing. The reference test was administered by one of five clinical audiologists who were all full members of Audiology Australia.

**Index test: Automated audiometry.** Automated audiometry was conducted using the KUDUwave (eMoyoDotNet, Pretoria, South Africa) a mobile Type 2B screening, diagnostic and clinical audiometer (IEC 60645-1/2) using the ascending method according to ISO8253-1:2010. The KUDUwave audiometer utilises double attenuation with insert earphones and circumaural earcups, continual noise monitoring, enabling accurate testing down to 0dB with an ambient noise level of up to 59dB sound pressure level (SPL). Ambient noise levels are monitored in octave bands through an external microphone on each circumaural earcup. The noise monitoring function of the KUDUwave uses low-pass (<125Hz), even single octave band-pass (125, 250, 500, 1000, 2000, 4000, and 8000 Hz) and high-pass (>8000 Hz) filters to separate the incoming sound. The filters have a stop-band attenuation of 90dB and pass-band ripple of 0.003dB. The outputs of these filters are monitored in real-time and the peak value passes to the user interface software (eMOYO) every 100ms, which is visually represented within the software (Figure 1). Placing insert earphones down to the bony part of the ear canal also reduces the occlusion effect allowing for bone-conduction evaluation with occluded ears using insert earphones. Air-conduction and bone-conduction thresholds were masked where applicable. See Brennan-Jones et al. for a detailed description of the audiological test parameters. The audiometry procedures were automated and recorded on a laptop using the eMOYO (v3.6.7) interface developed by eMoyoDotNet. The diagnostic criteria were calculated post-testing and
were not programmed into the eMOYO software. Testing took place in a quiet room that was not sound treated (mean ambient noise level when there was no outpatient clinic in progress was 46 dBA). The researcher gave standard instructions, placed the insert earphones, bone-conductor and headset on the participant and monitored the progress of the test in case of malfunction or patient discomfort.

Hearing thresholds were presented in dB hearing level (dB HL). Participants were tested at air conduction frequencies of 250, 500, 1000, 2000, 4000 and 8000 Hz and bone conduction frequencies of 500, 1000, 2000, 4000 Hz for both the reference standard and index test. The index test had lower maximum testable sound level limits compared to the reference standard (KUDUwave limits for air conduction were 95 dB for 250 Hz, 100 dB for 500, 1000, 2000, 4000 Hz, and 90 dB at 8000 Hz; sound level limits for bone conduction were 55 dB at 500 Hz and 70 dB at 1000, 2000 and 4000 Hz). In cases where both the index test and the reference standard reached maximum testable limits before a participants hearing threshold was established, a predicted threshold at the index test’s maximum level was imputed. The time interval between reference test and the index test being conducted was less than 60 min for all participants, as patients proceeded directly to the next test, or after a short break if requested.

**Diagnostic criteria**

All participants were examined to identify whether they had normal hearing, a disabling hearing loss, a conductive or sensorineural hearing loss and whether their hearing loss was bilateral, unilateral or asymmetrical, except where individuals had missing data that prevented the diagnostic criteria being calculated (n = 3). Normal hearing in this study was defined according the World Health Organisation (WHO) criteria of $\leq 26$ dB for hearing loss (i.e. normal hearing is $<26$ dB) in the better ear, with hearing thresholds averaged across 500, 1000, 2000 and 4000 Hz.\(^{14}\) Disabling hearing loss in this study was defined according the WHO criteria of $\geq 41$ dB hearing loss in the better ear, with hearing thresholds averaged across 500, 1000, 2000 and 4000 Hz.\(^{14}\)

Conductive hearing loss was defined in this study as an air-bone gap of $\geq 20$ dB at two or more adjacent frequencies out of 500, 1000, 2000 and 4000 Hz.\(^{17}\) The individual frequency bone-conduction thresholds were subtracted from the individual frequency air-conduction thresholds for both the index and reference test to obtain the air-bone gap.

Unilateral or asymmetrical hearing loss was defined in this study as a $\geq 20$ dB difference between the left and right bone-conduction thresholds at two or more adjacent frequencies out of 500, 1000, 2000 or 4000 Hz.\(^{17}\)

**Blinding**

The audiologist administering the reference standard (manual audiometry) was blinded to the results of the index test (automated audiometry). The researcher administering the index test was not blinded to the results of the reference test as the index test was automated and therefore could not influence the results.

![Figure 1. KUDUwave automated audiometry screen display during testing.](image-url)
Statistical methods

The diagnostic accuracy analysis used air-conduction thresholds for 500, 1000, 2000, 4000 Hz and bone-conduction thresholds of 500, 1000, 2000 and 4000 Hz for both manual and automated audiometry. Sensitivity and specificity along with positive and negative predictive value were calculated for each diagnostic category.

Agreement between automated and manual audiometry across the diagnostic categories was assessed using Cohen's kappa statistic. The Landis and Koch recommendations for agreement classification were applied, with $k < 0$ indicating no agreement, $k = 0–0.20$ indicating 'slight' agreement, $k = 0.21–0.40$ indicating 'fair' agreement, $k = 0.41–0.60$ indicating 'moderate' agreement, $k = 0.61–0.80$ indicating 'substantial' agreement and $k = 0.81–1.00$ indicating 'almost perfect' agreement.

Results

Participants

The mean age of participants was 49.93 years (standard deviation (SD) = 17.35, range 19.33–92.55 years). Hearing loss was not the primary complaint of all participants and many had more than one ear or hearing-related disorder at the time of testing. An additional 11 patients who did not complete both hearing assessments ($n = 7$) or had reliability questioned (i.e. suspected attention or cognitive ability issues) by the clinical audiologist ($n = 4$) were not included in the analysis.

Diagnostic accuracy of clinical protocols for automated audiometry

Table 1 presents $2 \times 2$ tables of the diagnostic data for detecting disabling, conductive and unilateral hearing loss using pre-defined clinical protocols applied to automated audiometry. Three participants had incomplete data that prevented the calculation of diagnostic criteria (one patient for disabling hearing loss, two for unilateral hearing loss). Table 2 presents sensitivity and specificity results; overall sensitivity was 0.88 (95% confidence interval (CI) 0.75–0.95) and overall specificity was 0.93 (95% CI 0.88–0.97). The highest sensitivity and specificity was for identifying disabling hearing loss (1.00 and 0.97 respectively), and the lowest was for identifying a normal ear (0.86 and 0.86 respectively). There were generally greater 95% CIs for sensitivity than for specificity. The positive predictive values (PPVs) ranged between 75–86% across the diagnostic criteria and negative predictive values (NPVs) ranged from 92–100% (Table 2).

Agreement

Cohen’s kappa agreement ranged from ‘substantial’ to ‘almost perfect’ across the diagnostic criteria, $k = 0.80$ (95% CI 0.70–0.89); $p < 0.001$.

<table>
<thead>
<tr>
<th>Diagnostic criteria</th>
<th>HL Absent</th>
<th>HL Present</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test positive</td>
<td>4</td>
<td>12</td>
<td>16</td>
</tr>
<tr>
<td>Test negative</td>
<td>24</td>
<td>2</td>
<td>26</td>
</tr>
<tr>
<td>Totals</td>
<td>28</td>
<td>14</td>
<td>42</td>
</tr>
<tr>
<td>Disabling HL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test positive</td>
<td>1</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Test negative</td>
<td>36</td>
<td>0</td>
<td>36</td>
</tr>
<tr>
<td>Totals</td>
<td>37</td>
<td>4</td>
<td>41</td>
</tr>
<tr>
<td>Conductive HL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test positive</td>
<td>3</td>
<td>18</td>
<td>21</td>
</tr>
<tr>
<td>Test negative</td>
<td>60</td>
<td>3</td>
<td>63</td>
</tr>
<tr>
<td>Totals</td>
<td>63</td>
<td>21</td>
<td>84</td>
</tr>
<tr>
<td>Unilateral HL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test positive</td>
<td>2</td>
<td>11</td>
<td>13</td>
</tr>
<tr>
<td>Test negative</td>
<td>26</td>
<td>1</td>
<td>27</td>
</tr>
<tr>
<td>Totals</td>
<td>28</td>
<td>12</td>
<td>40</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Diagnostic criteria</th>
<th>L95%CI</th>
<th>U95%CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.86</td>
<td>0.56</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.86</td>
<td>0.66</td>
</tr>
<tr>
<td>PPV</td>
<td>0.75</td>
<td>0.47</td>
</tr>
<tr>
<td>NPV</td>
<td>0.92</td>
<td>0.73</td>
</tr>
<tr>
<td>Disabling HL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>1.00</td>
<td>0.40</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.97</td>
<td>0.84</td>
</tr>
<tr>
<td>PPV</td>
<td>0.80</td>
<td>0.30</td>
</tr>
<tr>
<td>NPV</td>
<td>1.00</td>
<td>0.88</td>
</tr>
<tr>
<td>Conductive HL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.86</td>
<td>0.63</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.95</td>
<td>0.86</td>
</tr>
<tr>
<td>PPV</td>
<td>0.86</td>
<td>0.63</td>
</tr>
<tr>
<td>NPV</td>
<td>0.95</td>
<td>0.86</td>
</tr>
<tr>
<td>Unilateral HL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.92</td>
<td>0.60</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.93</td>
<td>0.75</td>
</tr>
<tr>
<td>PPV</td>
<td>0.85</td>
<td>0.54</td>
</tr>
<tr>
<td>NPV</td>
<td>0.96</td>
<td>0.79</td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sensitivity</td>
<td>0.88</td>
<td>0.75</td>
</tr>
<tr>
<td>Specificity</td>
<td>0.93</td>
<td>0.88</td>
</tr>
<tr>
<td>PPV</td>
<td>0.82</td>
<td>0.69</td>
</tr>
<tr>
<td>NPV</td>
<td>0.96</td>
<td>0.91</td>
</tr>
</tbody>
</table>

HL: hearing loss; L95%CI: lower 95% confidence interval; NPV: negative predictive value; PPV: positive predictive value; U95%CI: upper 95% confidence interval.
Discussion

This is the first study to examine the accuracy of pre-defined diagnostic criteria applied to automated audiometry in a telehealth context. The aim of the study was to examine whether standard diagnostic criteria applied to automated audiometry is an accurate method of identifying a number of key audiometric characteristics which can guide further referral to specialist services. The primary use of this method would be to increase access to audiology services in remote areas as part of a tele-audiology programme or increase efficiency of audiology practices in urban areas by facilitating patient triaging. The overall agreement ranged from substantial to almost perfect\(^2\) and was highly significant \((p < 0.001)\), indicating that pre-defined diagnostic criteria may be an accurate and effective method of identifying hearing disorders using automated audiometry. However, there was a wide range of variance in agreement \((0.70-0.89)\) and further research beyond a pilot study with additional participants will enable a more precise range of variance to be calculated. Sensitivity, specificity, positive and negative predictive values appeared to be high across conditions, indicating that automated audiometry may be an accurate diagnostic measure when using pre-defined diagnostic protocols. However, there are no similar published studies that allow us to ascertain whether the values obtained truly represent high sensitivity and specificity. Screening audiometry in unselected populations perhaps offers the truly represent high sensitivity and specificity. Screening audiometry are not directly comparable to diagnostic audiometry in this study.

The diagnostic accuracy of audiometry will often be expected to provide conclusive evidence to support adopting pre-defined diagnostic protocols. However, it demonstrates proof of concept and shows that this method warrants further investigation as it has significant potential for improving audiology service delivery. Sensitivity and specificity for identifying normal hearing (the largest clinical group) was 86%. We would therefore expect future studies to have a similar level of accuracy, albeit with more precise confidence intervals.

Role of diagnostic protocols in a telehealth framework

With over 360 million people around the world with disabling hearing loss,\(^1\) and a shortage of hearing health specialists,\(^2\) new models of diagnosis and care are required to provide equitable access to healthcare. Automated audiometry is becoming increasingly sophisticated and has recently been validated for use without a sound treated environment,\(^2\) and in a clinically heterogeneous population,\(^2\) meaning it has the potential to overcome some of these obstacles. However, the results of a hearing test still need to be interpreted to determine the next step in the clinical pathway. Diagnostic protocols such as those applied in this study have the potential to address this, for synchronous, asynchronous and hybrid tele-audiology services.\(^2\) An example of a possible clinical pathway using diagnostic protocols for automated audiometry is shown in Figure 2. Applying these protocols to automated audiometry could enable local health workers to triage patients with ear and hearing disorders according to their type and level of hearing loss. In particular, the identification of participants with normal hearing levels will directly reduce the workload of audiologists and enable more patients to be seen or allow more time spent on complex cases, whether synchronous or asynchronous methods are used.

Clinical relevance and further research

This study presents a method of identifying patients with hearing disorders that can be applied to automated audiometry within telehealth practices. Assistants currently working to facilitate both synchronous and asynchronous tele-audiology services would be able to apply, conduct and interpret results from automated audiometry using pre-defined diagnostic protocols using automated audiometry, reducing the time and administrative burden placed on clinical audiologists. Classification systems do have their limitations.\(^6\) Classifying hearing loss by severity does not provide any medical information (conductive or sensorineural hearing loss) or an understanding of the patient’s perceived experience of their hearing loss (i.e. low-frequency versus high-frequency hearing loss) or quality of life.

The diagnostic accuracy of audiometry will often be strengthened by interpretation by an experienced audiologist. However, clinical protocols that adhere to these diagnostic criteria could allow basic treatment and referral decisions to be made based on the results of automated audiometry conducted in a primary care or tele-audiology setting where experienced audiologists may not be available. In addition to complementing tele-audiology models of care, the diagnosis of these conditions could be an effective method of triaging appropriate referral to audiologists and otolaryngologists in secondary and tertiary care.

It is expected that there will be disparities in clinical guidelines locally, nationally and internationally which would affect the definitions used for the diagnostic protocols. However, this study serves to provide a proof of concept and numerous iterations of pre-defined diagnostic protocols are possible for identifying target conditions. Current automated audiometers can accommodate different diagnostic criteria, but the use of alternative guidelines will vary the sensitivity and specificity. It is therefore
suggested that a formal analysis of alternative diagnostic definitions should be conducted before implementation.

Further research should also consider patient perceptions of automated audiometry testing as part of a telehealth service. Patient non-acceptance of telehealth applications in audiology is decreasing, but has ranged from 9–30% in previous studies.29,30 Quality assessment of automated audiograms is also an area for future research, in order to identify patients for whom automated testing may not be suitable.

Conclusion

The current study demonstrates significant potential for the use of diagnostic protocols applied to automated audiometry to complement current telehealth models of care in audiology. Pilot data demonstrate that pre-defined diagnostic protocols applied to automated audiometry asynchronously are sensitive, specific and have substantial to almost perfect agreement for identifying disabling, conductive and unilateral hearing loss. We have demonstrated proof of concept for this method to be used in a telehealth model of care to improve synchronous and asynchronous tele-audiology service delivery. However, further studies utilising a greater number of participants with the target conditions are now required.

Acknowledgments

The authors would like to acknowledge all of the participants in this study for the generous use of their time. They would also like to thank all the medical, nursing and support staff from the Sir Charles Gairdner Hospital, Department of Otolaryngology, and audiologists Amie Grandidge, Jonathan Constantine, Susan Tegg-Quinn, Varsha Matthews and Joanne Sahdeo.

Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: CGBJ received financial support from an Australian Postgraduate Award (APA) scholarship administered through The University of Western Australia.

References


Asynchronous interpretation of manual and automated audiometry:
Agreement and reliability

Christopher G Brennan-Jones¹,², Robert H Eikelboom¹,²,³, Rebecca J Bennett¹,², Karina FM Tao¹,² and De Wet Swanepoel¹,²,³

Abstract

Introduction: Remote interpretation of automated audiometry offers the potential to enable asynchronous tele-audiology assessment and diagnosis in areas where synchronous tele-audiometry may not be possible or practical. The aim of this study was to compare remote interpretation of manual and automated audiometry.

Methods: Five audiologists each interpreted manual and automated audiograms obtained from 42 patients. The main outcome variable was the audiologist’s recommendation for patient management (which included treatment recommendations, referral or discharge) between the manual and automated audiometry test. Cohen’s Kappa and Krippendorff’s Alpha were used to calculate and quantify the intra- and inter-observer agreement, respectively, and McNemar’s test was used to assess the audiologist-rated accuracy of audiograms. Audiograms were randomised and audiologists were blinded as to whether they were interpreting a manual or automated audiogram.

Results: Intra-observer agreement was substantial for management outcomes when comparing interpretations for manual and automated audiograms. Inter-observer agreement was moderate between clinicians for determining management decisions when interpreting both manual and automated audiograms. Audiologists were 2.8 times more likely to question the accuracy of an automated audiogram compared to a manual audiogram.

Discussion: There is a lack of agreement between audiologists when interpreting audiograms, whether recorded with automated or manual audiometry. The main variability in remote audiogram interpretation is likely to be individual clinician variation, rather than automation.

Keywords
Automated audiometry, audiometry, audiology, eHealth, hearing loss, tele-audiology, telehealth

Date received: 12 July 2016; Date accepted: 23 August 2016

Introduction

The current, conventional method for assessing hearing in adults involves a clinician manually performing pure-tone audiometry in a suitably sound-treated environment and interpreting the results on-site.¹ However, higher rates of hearing loss and lower rates of intervention uptake in rural and remote populations, coupled with the shortage of audiological services in these areas, has been of significant concern in both developed and developing countries.²,³ Telehealth solutions and the automation of audiometry have been proposed as a potential means to increase access to hearing services in underserved populations.³–⁵

A number of automated audiometers have recently been clinically validated, including the AMTAS⁶,⁷ and KUDUwave.⁸ The consensus across studies is that automated audiometry is a suitable alternative to manual audiometry,⁹ although some studies have identified an absolute mean difference of up to 10 dB in automated air conduction thresholds compared to manual hearing thresholds, with further increased variance of approximately 15 dB for bone-conduction thresholds.¹⁰ Automated audiometry has been validated for use without a sound-treated environment, and in a clinically heterogeneous population, meaning it has the potential to overcome some of the obstacles associated with testing in rural and remote areas.¹⁰–¹²

¹Ear Science Institute Australia, Subiaco, Perth, WA, Australia
²Ear Sciences Centre, School of Surgery, The University of Western Australia, Subiaco, WA, Australia
³Department of Speech-Language Pathology and Audiology, University of Pretoria, Pretoria, South Africa

Corresponding author:
Christopher Brennan-Jones, Ear Science Institute Australia, Suite 1, Level 2, 1 Salvado Road, Subiaco, WA 6008, Australia.
Email: chris.brennan-jones@earscience.org.au
Synchronous or “live” tele-audiology assessment, where the clinician administers and interprets the hearing assessment simultaneously, may not be possible due to connectivity issues in many rural and remote areas or due to limited clinician time. An alternative may be assessment with automated audiometry and remote interpretation of the results in an asynchronous telehealth model. This offers benefits such as greater coverage for difficult-to-access and transient populations, such as some Indigenous Australian communities, and allows for opportunistic assessments performed by local health workers which may be more efficient than scheduled appointments in some populations. It may also offer benefits to clinicians, reducing travel and enabling flexible working environments (e.g. working from home). The remote interpretation of test results is common and has been validated in a number of areas of medicine to facilitate telehealth services; including interpretations of retinal images, radiography, echocardiograms and otoscopy. However, comparisons between remote interpretations of manual and automated audiology have not been reported previously.

The aims of the present study were to examine the agreement between remote interpretations of manual and automated audiograms by audiologists and whether the potential variation in hearing thresholds introduced by automated audiology would affect the clinical decisions made by audiologists.

Methods

This study compared the intra- and inter-observer agreement between remote interpretations of manual and automated audiology. Results of agreement studies are intended to provide information about the amount of error inherent in any diagnosis, score or measurement.

Participants for this study were five audiologists recruited from the Ear Science Institute Australia. Audiologists who were more than three years’ post-qualification and maintained more than one day per week of clinical audiology practice were invited to take part in the study. The audiologists analysed existing data collected from patients in a validation study of automated audiology which recruited 42 adults (>18 years) presenting with suspected hearing loss at public audiology and otolaryngology clinics at Sir Charles Gairdner Hospital, Perth, Western Australia; see Brennan-Jones et al. for further details of the study population and test procedures. This study was approved by the University of Western Australia’s human research ethics committee.

Equipment

Manual audiology was conducted within a sound-treated room (mean ambient noise level 37 dBA) using Acoustic Analyser AA30 audiometer (Starkey Hearing Technologies, MN, USA), calibrated to ISO389-1:1998 and TDH-39P (Telephonics, NC, USA) supra-aural headphones and Radioear B-71 bone-conductor (Radioear Corp., PA, USA), calibrated to ISO389-3:1994. The bone-conductor was placed on the patient’s mastoid for manual testing. Patient history, otoscopy and tympanometry using a GSI 38 Auto Tym (Grason-Stradler, MN, USA) preceded audiometry testing. Five audiologists were involved in performing the manual audiology assessments. Automated audiology was conducted using the KUDUwave (eMoyoDotNet, Pretoria, South Africa) — a mobile Type 2B screening, diagnostic and clinical audiometer (IEC 60645-1/2) using the ascending method according to ISO8253-1:2010. The automated assessments were facilitated by authors CGBJ or RHE.

Procedure

Manual audiology, automated audiology, tympanometry and participant demographics (age and gender) were extracted and standardised using an audiogram generator so that the manual and automated audiograms could not be distinguished (manual audiology results were originally recorded by hand, whereas automated results are recorded electronically). The audiograms (and accompanying clinical information) were anonymised, randomised by allocation to a unique, randomly generated 4-digit number and then sorted in ascending order for interpretation (see Appendix 1).

The five audiologists participating in this experiment, blinded to whether manual or automated audiology was used, independently interpreted the audiograms, together with the other available information (age, gender and tympanometry; a full patient history was not available to participating audiologists). Audiologists were aware that some of the audiograms for interpretation were obtained by automated testing and that these would be compared to manual audiograms. However, they were not aware that they would be interpreting matched pairs of manual and automated audiograms. They were asked to provide a determination of: 1) the level and type of hearing loss; 2) a management plan for the patient given their audiometric results; and 3) their judgement of the reliability of the audiogram. There are no universally agreed standards for determining the type and level of hearing loss, management plan or reliability of the audiogram; these are normally the result of professional training and local clinic protocols.

For the purposes of this study, 1) the options for level of hearing loss were normal hearing, slight hearing loss or significant hearing loss requiring intervention and the options for type of hearing loss were normal hearing, sensorineural, conductive or mixed hearing loss; 2) the management options were no intervention, referral for hearing aids, referral for medical treatment or other; and 3) audiologists were given the option to add comments on the reliability of the audiogram. To measure this outcome a “no” answer was considered to imply that an audiologist had no issues with reliability and a “yes” answer combined with a comment questioning the reliability of the audiogram was considered to indicate that an audiologist questioned the reliability of the audiogram (see Appendix 1).
**Data analysis**

Firstly, Cohen’s Kappa was used to calculate the intra-observer agreement of the audiologist’s asynchronous interpretations for determining the type and severity of hearing loss, and the recommendation for patient management. This was to examine whether clinicians agreed with themselves for manual versus automated interpretations.

Secondly, Krippendorff’s Alpha and Q-statistic was used to calculate the inter-observer agreement of remote interpretations for determining the type and severity of hearing loss, each ear separately, and the recommendation for patient management. Krippendorff’s Alpha allows comparison of agreement between multiple coders, in this case audiologists, and was, therefore, used to compare whether clinicians agreed with each other when interpreting both manual and automated audiograms. The Q-statistic represents the probability of reaching $\alpha > 0.6$ (substantial agreement).

Thirdly, the main outcome variable, agreement between audiologist’s recommendation for patient management between the manual and automated audiometry test, was examined using Cohen’s Kappa and $p$-value for intra-observer agreement and Krippendorff’s Alpha and Q-statistic for inter-observer agreement. The paired-samples $t$-test was also used to determine the correlation coefficient between manual and automated audiometry for management decisions between audiologists.

Finally, the audiologist-related accuracy of audiograms was examined using McNemar’s test; that is, odds ratios (ORs) for paired nominal data, in this case, dichotomous audiogram reliability scores.

The Landis and Koch (1977) recommendations of agreement classification were applied to Cohen’s Kappa and p-value for intra-observer agreement and Krippendorff’s Alpha analyses, with $a < 0$ indicating no agreement, $a = 0–0.20$ indicating “Slight” agreement, $a = 0.21–0.40$ indicating “Fair” agreement, $a = 0.41–0.60$ indicating “Moderate” agreement, $a = 0.61–0.80$ indicating “Substantial” agreement and $a = 0.81–1.00$ indicating “Almost perfect” agreement.

**Results**

Intra-observer pooled agreement for clinician’s interpretations for: (i) the level of hearing loss for manual versus automated audiograms ranged from moderate to almost perfect agreement ($\alpha = 0.637$ [95% confidence interval (CI) 0.452 to 0.822]; $p < 0.001$); (ii) the type of hearing loss ranged from fair to substantial agreement ($\alpha = 0.407$ [95% CI 0.207, 0.613]); $p < 0.001$; and (iii) management outcomes ranged from moderate to almost perfect ($\alpha = 0.693$ [95% CI 0.521 to 0.865]; $p < 0.001$), see Tables 1 and 2.

Inter-observer agreement varied from moderate to substantial in regards to the level of hearing loss, type of hearing loss and management for both automated and manual audiometry (Table 3). However, inter-observer agreement was not significant for interpretation of type of hearing loss bilaterally for either manual or automated audiometry. Inter-observer agreement was significant for both right and left ear interpretations of level of hearing loss using manual audiometry ($\alpha = 0.692$ [95% CI 0.593, 0.778]; $q = 0.029$ and $\alpha = 0.696$ [95% CI 0.569, 0.790]; $p < 0.001$).

**Table 1.** Individual clinician (intra-observer) reliability for manual versus automated audiogram interpretations measured with Kappa agreement.

<table>
<thead>
<tr>
<th>Management outcomes</th>
<th>$\alpha$ [95% CI]</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinician 1</td>
<td>0.810</td>
<td>0.638, 0.982</td>
</tr>
<tr>
<td>Clinician 2</td>
<td>0.771</td>
<td>0.618, 0.924</td>
</tr>
<tr>
<td>Clinician 3</td>
<td>0.659</td>
<td>0.481, 0.837</td>
</tr>
<tr>
<td>Clinician 4</td>
<td>0.574</td>
<td>0.386, 0.762</td>
</tr>
<tr>
<td>Clinician 5</td>
<td>0.651</td>
<td>0.472, 0.829</td>
</tr>
<tr>
<td>Clinician agreement (pooled)</td>
<td>0.693</td>
<td>0.521, 0.865</td>
</tr>
</tbody>
</table>

Cl: confidence interval.

<table>
<thead>
<tr>
<th>Type of hearing loss</th>
<th>$\alpha$ [95% CI]</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinician 1</td>
<td>0.433</td>
<td>0.244, 0.622</td>
</tr>
<tr>
<td>Clinician 2</td>
<td>0.343</td>
<td>0.127, 0.560</td>
</tr>
<tr>
<td>Clinician 3</td>
<td>0.449</td>
<td>0.226, 0.672</td>
</tr>
<tr>
<td>Clinician 4</td>
<td>0.429</td>
<td>0.251, 0.607</td>
</tr>
<tr>
<td>Clinician 5</td>
<td>0.379</td>
<td>0.185, 0.578</td>
</tr>
<tr>
<td>Clinician agreement (pooled)</td>
<td>0.407</td>
<td>0.207, 0.613</td>
</tr>
</tbody>
</table>

Cl: confidence interval.
For automated audiometry, however, only the right ear level of hearing loss interpretations showed significant inter-observer agreement (\(\alpha = 0.680\) [95% CI 0.604, 0.755]; \(q = 0.019\)) (Table 3, Figures 1 and 2).

There was no significant inter-observer agreement for management outcomes for either manual or automated audiometry (Tables 3 and 4, Figures 1 and 2).

The correlation coefficient between manual and automated audiometry for management decisions between audiologists was high and significant (0.823, \(p < 0.001\)) using a paired-samples \(t\)-test. Concurrence for discharge/no treatment was highest (93.8%), followed by audiological referral and medical referral (respectively 81.8% and 81.4%), whereas combined medical referral was lower at 56.8% (Table 4).

### Table 3. Overall clinician (inter-observer) reliability across the five remote interpretation outcomes using Krippendorff’s Alpha and probability \((q)\) of reaching \(\alpha > 0.6\).

<table>
<thead>
<tr>
<th>Interpretation outcomes</th>
<th>Manual audiometry</th>
<th>Automated audiometry</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(\alpha)</td>
<td>[95% CI]</td>
</tr>
<tr>
<td>Right-type HL</td>
<td>0.582</td>
<td>0.519, 0.642</td>
</tr>
<tr>
<td>Left-type HL</td>
<td>0.551</td>
<td>0.490, 0.615</td>
</tr>
<tr>
<td>Right-level HL</td>
<td>0.692</td>
<td>0.593, 0.778</td>
</tr>
<tr>
<td>Left-level HL</td>
<td>0.696</td>
<td>0.569, 0.790</td>
</tr>
<tr>
<td>Management</td>
<td>0.569</td>
<td>0.505, 0.630</td>
</tr>
</tbody>
</table>

CI: confidence interval; HL: hearing loss
Table 4. Distribution of management decisions for remote interpretations of manual and automated audiometry.

<table>
<thead>
<tr>
<th>Automated audiometry management</th>
<th>Manual audiometry management, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Discharge</td>
</tr>
<tr>
<td>Discharge</td>
<td>45 (93.8)</td>
</tr>
<tr>
<td>Aud ref</td>
<td>2 (9.1)</td>
</tr>
<tr>
<td>Med ref</td>
<td>6 (16.2)</td>
</tr>
<tr>
<td>Aud + Med ref</td>
<td>2 (2.0)</td>
</tr>
<tr>
<td>Other</td>
<td>1 (100.0)</td>
</tr>
</tbody>
</table>

Audiologists were more likely to question the reliability of automated audiograms than manual audiograms (OR = 2.848; p < 0.001) (Table 5).

Table 5. Subjective clinician-rated reliability of automated versus manual audiograms using McNemar’s test.

<table>
<thead>
<tr>
<th>Audiometry reliability</th>
<th>OR [95% CI]</th>
<th>X²</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinician 1</td>
<td>2.091</td>
<td>0.898</td>
<td>.118</td>
</tr>
<tr>
<td>Clinician 2</td>
<td>3.222</td>
<td>1.296</td>
<td>.064</td>
</tr>
<tr>
<td>Clinician 3</td>
<td>1.111</td>
<td>0.020</td>
<td>.238</td>
</tr>
<tr>
<td>Clinician 4</td>
<td>7.364</td>
<td>3.454</td>
<td>.006</td>
</tr>
<tr>
<td>Clinician 5</td>
<td>0.951</td>
<td>0.051</td>
<td>.999</td>
</tr>
<tr>
<td>Pooled reliability</td>
<td>2.848</td>
<td>6.532</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

OR: odds ratio; CI: confidence interval.

Discussion

The present study shows that agreement for management decisions for participants remains relatively high when automated audiograms are interpreted remotely, with intra-observer agreement ranging from moderate to almost perfect. There was no significant inter-observer agreement for patient management decisions for manual or automated audiometry, indicating that management decisions differed between clinicians regardless of whether manual or automated audiometry was used.

There was significant, moderate to substantial intra-observer agreement for audiologists when determining the level and type of hearing loss using automated compared to manual audiometry. However, for inter-observer agreement, only decisions relating to the level of hearing loss were significant, but this was true for both manual and automated audiometry. Our results show a lack of agreement between audiologists when determining the type of hearing loss and management decisions when interpreting manual audiometry. This highlights that the main source variable in the agreement between decisions made based on remote audiogram interpretation is likely to be individual clinician variation, rather than automation.

Determination of the type and level of hearing loss was deliberately subjective in this experiment with no set quantitative criteria given to the audiologists. Whilst standard criteria exist for classifying the level of hearing loss, the applicability of these arbitrary cut-offs to clinical practice has been questioned. Therefore, the audiologists were presented with options of clinical significance (i.e. whether referral for intervention was necessary, and if so, what type). With a lack of definition between sensorineural, conductive and mixed hearing loss, the classification of patients into these groups was diverse and agreement was poor between test method and clinicians.

Despite good intra-observer agreement and being blinded as to which audiograms were automated and which were manual, audiologists were 2.8 times more likely to question the reliability of an automated audiogram (Table 5). The specific wording and scoring of the question used in this study may have influenced the reporting rates of reliability. Whilst the question prompted audiologists to consider the reliability of the audiogram (rather than a spontaneous comment) the need for confirmation with a specific comment may have resulted in under-reporting. However, it has also been argued that the use of automated audiometry may actually limit bias; that is, increase reliability in audiometric assessment. The tester bias associated with audiometry is well-documented and many audiologists may consciously or sub-consciously alter hearing thresholds to adhere to certain rules or an expected pattern. This perception of poor reliability may be, in part, influenced by the fact that automated audiometry does not perform these adjustments. When audiometric results are shown that do not fit the expected pattern or conventions, such as when a bone-conduction threshold appears worse than air-conduction thresholds, this may be interpreted as an unreliable assessment. However, it is also recognised that for some patients, audiometry can be a difficult task to perform and the reliability or accuracy of the audiogram can be compromised with automation. Therefore, the capacity to provide synchronous tele-audiometry assessment for patients suspected to have poor reliability would be beneficial to complement a predominately asynchronous model.

Audiometry is a key part of any test battery for the assessment of hearing and its impact on daily function and quality of life for patients. The remote interpretation of audiometry potentially offers significant efficiency and financial savings for telehealth programs and potential to improve access to services using automated technology, without the need for a clinician to travel to remote areas. However, the divergent presentation of patients with hearing loss means that audiometric assessment alone is often not sufficient and that the effect of hearing loss on daily function and quality of life should be ascertained before clinical decisions on patient treatment and management are made based on remote audiogram interpretation.
The findings from this study suggest that an asynchron-ous tele-audiology model, where automated audiometry is performed remotely with results forwarded for management by audiologists or medical personnel, is practicable. This could facilitate much wider coverage of ear and hearing services, streamlining metropolitan specialist services and reducing the need for specialists to travel to rural and remote regions to administer services. Future research should investigate the addition of further clinical, contextual and quality of life information that may improve both the inter- and intra-observer reliability of remote interpretations.

Potential limitations of this study include the use of subjective diagnosis criteria and the reliability index used to measure treatment and reliability outcomes. The present study also examines only one automated audiometry device in an adult population. Future studies examining different devices in different populations would be beneficial.

Conclusion
Remote interpretation of automated audiometry appears to be a reliable approach for diagnosing hearing loss and identifying appropriate interventions. Clinician interpretations vary significantly, both for manual and automated audiograms. It is thought that this variation is not exclu-sive to remote interpretation of audiometry in a telehealth context; rather, it is reflective of the diverse needs of patients with hearing loss and a clinician’s personal experience. The findings from this study highlight that the use of remote interpretations of automated audiometry as a method for assessing hearing ability has equivalent agreement to audiologists interpreting manual audiometry and is, therefore, feasible in the context of a comprehensive tele-audiological program.

Declaration of conflicting interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding
The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

Appendix 1