UNIVERSITY OF BELGRADE

SCHOOL OF MEDICINE

Snezana D Andrić Filipović

CORRELATION ANALYSES OF THE SEASONAL CHANGES AND QUALITY OF LIFE WITH THE CIINICAL PRESENTATIONS OF OTITIS MEDIA IN CHILDREN 3-8 YEARS OF AGE

Doctoral Dissertation

Belgrade, 2015

УНИВЕРЗИТЕТ У БЕОГРАДУ

МЕДИЦИНСКИ ФАКУЛТЕТ

Снежана Д Андрић Филиповић

АНАЛИЗА ПОВЕЗАНОСТИ СЕЗОНСКИХ ПРОМЕНА И КВАЛИТЕТА ЖИВОТА СА РАЗЛИЧИТИМ КЛИНИЧКИМ ПРЕЗЕНТАЦИЈАМА ЗАПАЉЕЊА СРЕДЊЕГ УВА КОД ДЕЦЕ УЗРАСТА 3-8 ГОДИНА

Докторска Дисертација

Београд, 2015

Mentor: Prof. dr Jovica Milovanović, University of Belgrade, School of Medicine

Comentor: Prof. Mark Haggard, University of Cambridge, UK

Committee members:

1. Prof. dr Vojko Djukić, University of Belgrade, School of Medicine

2. Prof. dr Tatjana Pekmezović, University of Belgrade, School of Medicine

3. Prof. dr Borivoj Babić, University of Belgrade, Faculty for special Education and Rehabilitation

ACKNOWLEDGEMENTS

This doctoral dissertation is the result of the team work, so it is only illusion that that behind the title is only my name. I would be making a serious mistake if I did not mention at the first place my dear parents, Dusan and Milosavka, who have always striven to give me intellectual values and who were my first teachers. I need to extend my warmest appreciation to my co-mentor Prof M. Haggard in guiding questions and methods, teaching the importance of scientific team work, seeking for scientific truth using the best possible methods and available analyses, developing scientific curiosity and the role of self-criticism in evaluating achieved results and their correct interpretation. I am also deeply grateful to my mentor, Prof. J. Milovanovic for mentoring and guiding, giving initiative encouragement and support through difficult times.

Special appreciation is due to the clinical heroes of the large and informative Balkan sample in the Eurotitis 2 study: Dijana Radovic, audiologist of the Department of audiological rehabilitation of Clinical centre of Serbia; Ivan Baljosevic, ENT specialist at Institute of мother and child of Serbia "Dr Vukan Čupić" New Belgrade; Sonja Stojanovic ENT specialist in the General hospital in Leskovac; Ognjen Jovicevic, ENT specialist in Clinical centre of Montenegro and Branka Damjanovic, nurse in Clinical centre of Montenegro. Special appreciation to the stuff of the Department for audiological rehabilitation and the Pediatric ambulatory care division of the Clinic of ENT and Maxillofacial surgery at Clinical Centre of Serbia. They helped through several long years of technical support printing, distributing questionnaires, collecting and filing data.

I am thankful to all collaborators in Eurotitis 2 study: to the clinicians in the non-Balkan centres of Eurotitis 2 study, and in the core team to statistician Helen Spencer, and to PhD student, Mary Zhang, for technical help in bringing the present of this work up to a uniform standard of checking.

I am also grateful to all my young patients and their parents or caregivers for their reliable responses and patient during questionnaires filling and their contribution to the achievement of high-quality scientific evidence within Medicine.

Finally, but no means least is my appreciation to my family: my husband Branko and sons Dusan and Damjan, without their understanding and support this work would not have had the good context and purpose that it has had.

In Belgrade,

27.12.2015

ЗАХВАЛНОСТ

Ова докторска дисертација је резултат тимског рада, зато је илузија да је иза учињеног само моје име. У изражавању захвалности, погрешила бих ако на прво место не бих ставила своје драге родитеље, Душана и Милосавку који су увек тежили да укажу на значај интелектуалних вредности и који су били моји први учитељи. Најтоплију и неизмерну захвалност дугујем коментору Проф. М. Haggard-у у усмеравању и подучавању научном и тимском раду, тежњи научној истини користећи најбоље методе и начине анализе, развоју научне радозналости и самокритичности у евалуацији постигнутих резултата и њиховој правилној интерпретацији. Такође сам дубоко захвална ментору Проф. Ј. Миловановић-у за менторство и сарадњу, као и подстицај и истрајност у остваривању овог рада.

Посебну захвалност дугујем херојима балканског узорка у пројекту Eurotitis 2 студије: Дијани Радовић, сурдоаудиологу Одсека за аудиолошку диајгностику Клиничког центра Србије, Београд; др Ивану Бањошевићу, ОРЛ специјалисти у Институту за здравствену заштиту мајке и детета Србије „Др Вукан Чупић“, Нови Београд; др Соњи Стаменковић, ОРЛ специјалисти Опште болнице Лесковац; Огњену Јовићевићу, ОРЛ специјалисти у Клиничкот центру Црне Горе и Бранки Дамјановић, медицинском техничару у Клиничком центру Црне Горе. Хвала свим сарадницима из Одсека аудиолошке рехабилитације и особљу Клинике за ОРЛ и МФХ, Клиничког центра Србије који су помагали у припреми материјала, дистрибуцији, прикупљању и чувању.

Такође се захваљујем свим сарадницима у Eurotitis 2 студији: сарадницима из не-балкансих центара у пројекту Eurotitis 2, затим једном од главних чланова тима, статистичару Helen Spencer, као и студенту последипломских студија Mary Zhang на техничкој помоћи око израде докторске дисертације.

Захвалност дугујем и свим мојим малим пацијентима и њиховим родитељима/старатљима на помоћи око израде балканског узорка европске студије о запаљенским променама срењег ува. На њиховом стрпљењу и доприносу у развоју високо квалитетне научне истине у медицини.

На крају, али тиме не и мању захвалност дугујем својој породици, супругу Бранку и синовима Душану и Дамјану без чије подршке и разумевања мој научни рад не би имао свој пуни смисао.

У Београду,

27.12.2015

CORRELATION ANALYSES OF THE SEASONAL CHANGES AND QUALITY OF LIFE WITH THE CIINICAL PRESENTATIONS OF OTITIS MEDIA IN CHILDREN 3-8 YEARS OF AGE

Snezana D Andrić Filipović

ABSTRACT

Otitis media (OM), in both the acute (RAOM) and chronic (OME) forms affecting young children, plus the developmental consequences, are a large burden on families' well-being and on healthcare systems. The burden comes from the large numbers with these very common conditions, rather than extreme individual cases. Past research on OM has mostly not well addressed the need for precise assessment of severity and statistical handling of case flow, via optimum referral and treatment decisions. The OM8-30 (with its short form OMQ-14) questionnaire is being standardised for Europe through the international clinical study Eurotitis-2. I contributed nearly one quarter of the cases (22.6%) and was regional coordinator for about a further 8.8%, making an overall contribution of one third of the cases (31.4%), so was given access to the whole database under agreements in place, for answering my research questions.

The first study moved on from the known seasonality of incidence of respiratory infection plus OM and used the severity measures from OM8-30 on different aspects of OM presentation. It asked: in cases all with OM diagnosis, do the measured severities of the several disease facets vary systematically through the year, after the typical autumn respiratory virus spread triggers an annual cycle? And do the timings of maximum severity in the single-consultation cross-sectional data reflect the cause-related sequence from upstream respiratory infection, through ear infection score (ESS) and measured and reported hearing (HL and RHD) to downstream developmental outcomes? To answer this question with maximum precision, I worked with the study statistician to develop a method for locating annual peak severity which is labour-intensive but robust: fitting a series of 27 lagged sinusoid functions I also developed a control framework fitting, as possible additive confounders, the background determinants of severity (particular centre, age, socioeconomic status (SES), length of history and particular diagnosis; sex was usually not significant). The novel results with these methods were clear for the upstream stages. The cross-sectional data show distinct annual cycles of case severity, with the ordering URTI→(ESS, HL)→RHD, and even close correspondence of estimated absolute delays with those in other true time-series (longitudinal) data. Season (or date) thus has to join the list of adjuster variables in the control framework underpinning standardisation.

Downstream results were less clearly related to plausible causal sequence as confirmed elsewhere by structural equation modelling; a possible reason is that the fact of two main causal pathways (through hearing and through physical health) diffuses the annual peak making no single characteristic delay, and further reasons and alternative approaches are briefly explored.

The second study addressed measures of hearing in OM and two main ways in which they might be similar or different. One objective measure is used widely, audiometric threshold (HL), and the study adds a previous, but not widely used, mapping into an HL scale of tympanometric measures reflecting the function of the middle ear. In addition I have added the precision scoring of four OM8-30 questions on parentally reported hearing difficulties giving RHD-4, short enough for routine clinical application. The latter two are more convenient and less costly in terms of clinical time, leading naturally to the practical extension of Study I into Study IV. The first part of the research question compares the determinants, now including season, of these three measures, and the second (connected) part concerns the interrelations of the measures and how their related-ness can be exploited in research or practice by totalling for reliability or by imputing HL-scaled tympanometry for (frequently missing) HL. Results showed general similarity of determinant pattern across all three measures, but specifically greater contribution to RHD severity from the length of history and to some extent from SES and season. Lower determination of HL than of the other two undermines the assumption of useful clinical meaning of this most accepted and precise measure, probably because of its known short-term fluctuations. Aggregation of all three measures emerged as acceptable for reliability and generality, partly on the basis that scaled tympanometry makes a direct contribution to RHD, not just through HL. Distributional and correlational data showed the degree to which substitution of missing HL with scaled tympanometry could be acceptable in diagnosed OM.

The third study examined criterion validity of the three factor scores in OMQ-14 (roughly corresponding to ESS, RHD, and downstream developmental impact) against the corresponding facet scores (i.e. for discreet item sets) in OM8-30, more fully sampled by about twice the number of items (depending on particular facet). This involved first developing a new corresponding impact measure within OM8-30, which has so far not been scored in this way. The corresponding scores were found all to have good distribution properties, and criterion validity correlations with OM8-30 measures above 0.90 are reported for all three OMQ-14 measures and for the total.

Study IV grew out of Study II, exploring the suggestion of quantitatively combining RHD with scaled tympanometry in a low-cost assessment or screening measure, for settings where audiometry (HL) might not be available. It used their complementary relationships in terms of objectivity, range and continuous gradation of score values, first in quantitative prediction of HL as most powerful examination of relationships. Then via predicting four HL dichotomies, e.g. ≥ for 20 dB HL versus lower it simulated possible screen cut-offs. In terms of standard screen performance parameters of sensitivity and specificity, it was found that the combination could acceptably service screen cut-offs of ≥ 20 or ≥ 25 dB HL (Specificities over 69% at 90% sensitivity). These cut-offs would not be treatment criteria, but by convention and considering the fluctuation of hearing in OM, they would be appropriate low cut-offs for further assessment, given the clinical objectives.

All four studies have contributed to the validation of the standardisation and control framework for Eurotitis-2 and offer practical suggestions for better clinical assessment in otitis media.

KEY WORDS: Otitis media, season, hearing measures, questionnaires, developmental impact, OM screening

АНАЛИЗА ПОВЕЗАНОСТИ СЕЗОНСКИХ ПРОМЕНА И КВАЛИТЕТА ЖИВОТА СА РАЗИЧИТИМ КЛИНИЧКИМ ПРЕЗЕНТАЦИЈАМА ЗАПАЉЕНА СРЕДЊЕГ УВА КОД ДЕЦЕ УЗРАСТА 3-8 ГОДИНА

СНЕЖАНА Д АНДРИЋ ФИЛИПОВИЋ

САЖЕТАК

Акутне и хроничне запаљенске промене средњег ува највише погађају децу млађег зраста, а развојни проблеми који при томе настају су додатно оптерећење за породицу и целокупни здравствени систем. Оптерећење настаје због великог броја болесне деце са просечном тежином симптома пре него појединачних случајева са тешком клиничком сликом обољења. Досадашња истраживања нису на најбољи начин дефинисала потребу за прецизном евалуацијом тежине клинчке слике и статистичке контроле броја деце упућене на даље лечење и врсте примењеног терапијског третмана. OM8-30 упитник (и његова краћа форма OMQ-14) је стандардизован у Европи кроз интернационални пројекат, Еуротитис 2. Ја сам допринела једној четвртини од укупног броја испитника (22.6%) и била регионални координатор за наредних 8.8% испитаника, што укупно чини 31.4% укупног узорка и стога добила приступ, уз сагласност свих ауторитета, бази података за проналажење одгвора на постављене циљеве истраживања.

Прва од студија се бави сезонским аспектима болести и она је границу испитивања померила даље од уобичајеног утицаја сезоне на инциденцу респираторних инфекција и запаљења средњег ува ка испитивању утицаја сезоне на тежину различитих облика клиничке слике болести (аспеката) измерених применом OM8-30 упитника. Циљ ове студије: да ли код све деце са постављеном дијагнозом запаљења средњег ува тежина различитих аспеката болести систематски варира током године након почека циклуса вирусних респираторних инфекција током јесени? Такође, да ли време максималне тежине симптома испитаника током студије пресека рефлектује узрочну везу са вирусним инфекцијама респираторних путева преко скорова упитника за симптоме средњег

ува (ESS) и измереног и регистрованог прага слуха (HLи RHD) па до исходних, развојних, проблема деце? Да бих одговорила на ово питање са максималном прецизношћу радила сам са студијским статистичарем и развила метод за лоцирање годишњег скока тежине симптома праћењем скорова упитника, што је било напорно али вредно: фитинг серије од 27 сукцесивних синусодиних функција. Ја сам такође развила контролни фиттинг за важне ко-варијабле у моделу, позадинске детерминанте клиничке слике (центар, старост, социоекномско стање-SES, историја болести и посебно дијагноза; пол обично није био значајна варијабла). Ефекти су били јачи за клиничке симптоме на почетку канонског, узрочног пута. Подаци из ове студије пресека указују на систематско кашњење тежине симптома у складу са прихваћеним каузалним ланцем URTI→ (ESS, HL) →RHD, и штавише преклапања са апсолутним временским интервалима из других студија лонгитудиналног типа. Сезона постаје важан детерминатор тежине клиничке слике и стога мора бити строго контролисана са осталим детерминантама у оквиру рада студије стандардизације. Исходни фактори или аспекти болести су имали слабију везу са узрочним варијаблама у каузалном ланцу што је и показано коришћењем структурног модела, а разлог је постојање два каузално узрочна пута (преко физичког здравља и преко слуха) који расипају годишњи максимум тежине симптома на више, не на само један, одређени годишњи максимум уз разматрање накнадних приступа и алетрнативних разлога.

Друга студија се бави облицима слушања и начинима њиховог мерења код деце са запаљенским променама средњег ува и издваја два главна облика или мере слушања које имају доста сличних, али и различитих карактеристика. Једна од објективних мера је широко примењивана у пракси и односи се на праг слуха измерен тоналном аудиометријом (HL) и друга, која није у широкој примени, представља импутиран HL применом тимпанометријског записа (ACET метода), рефлектујући функцију средњег ува. Као додатак ја сам користила прецизно скорован упитник од четири OM8-30 питања намењених родитељима (RHD4), довољно кратак за рутинску клиничку примену. Задње две мере слушања су погодније и мање захтевније у погледу клиничког времена што је природно допринело проширењем друге студије у четврту студију. Прва група циљева је

поређење утицаја детерминаната, сада и сезоне, на ове три мере слушања и друга (у наставку прве групе) проучавање односа између ових мера слушања и како се тај однос може применити у истраживању и пракси коришћењем све три мере ради поузданости или коришћењем импутираног прага слуха на основу тимпанометријског налаза у одсуству аудиометријског записа. Све три мере слушања су имале исте битне детерминанте, али величина ефекта историје болести имала је најјачи утицај на износ скора RHD упитника, а нешто слабији ефекат на RHD имале су SES и сезона. Утицај детерминаната на HL био је слабији него на остале мере слуха што подрива опште прихваћено мишљење о клиничкој важности најприхватљивије и прецизније мере слуха, због честих флуктуација. Обједињење све три мере слушања показало се прихватљиво због поузданости и применљивости и то великим делом јер примена тимпанометрије у процени прага слуха има директан, не само индиректан, ефекат на RHD. Дистрибутивни и корелацијски подаци показали су у коме степену је могућа клиничка примена кодиране тимпанометрије у супституцији одсутног прага слуха код деце са дијагнозом запаљења средњег ува.

Трећа студија се бави испитивањен валидности скорова три главна фактора упитника OMQ-14 (у глобалу подударних скоровима ESS, RHD I развојном утицају), њиховим поређењем са скоровима подударних аспеката OM8-30 упитника који има више него двоструко питања и који служи као критеријум за процену валидности. То је захтевало стварање нове, подударне мере развојног утицаја или исходне мере унутар OM8-30 упитника који до сада није скорован на тај начин. Скорови су имали добре дистрибутивне карактеристике и критеријуме валидности изнад 0.90 забележених за све три OMQ-14 мере и PC (principal component) тотал.

Четврта студија је произашла из друге студије истражујући могућност комбиновања RHD са импутираним прагом слуха у ниско буџетном скринингу слуха где аудиометрија може бити недоступна. Она се ослања на присутну комплементарност између објективности, распона и континуиране градације скорова у квантитавној предикцији HL као најмоћније мере односа. Предвиђањем четири дихотоме вредности прага слуха, на пример ≥ 20 dB HL насупрот нижег,

симулирали смо скрининг одређеног прага слуха. Комбинација специфичности и сензитивности, уобичајених термина перформансе скрининга, показала се добра у предвиђању ≥ 20 dB HL или ≥ 25 dB HL (специфичност преко 69% при сензитивности od 90%). Ови пресеци скрининга не би били критеријуми за примену терапије, али у складу са конвенцијом и узимајући у обзир флуктуацију прага слуха код запаљенских промена средњег ува, нижа вредност критеријума за скрининг слуха је пожељнија за задате клиничке циљеве.

Све четири студије заједно су допринеле валидности стандардизације и контроле радног оквира Еуротитис 2 студије и пружају практичне савете за бољу клиничку процену запаљенских стања средњег ува.

КЉУЧНЕ РЕЧИ: Запаљење средњег ува, сезона, мере слушања, упитници, развојни исход, скрининг деце са запаљенским променама среЉдњег ува

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0. General Introduction

1.0.1. Clinical Presentation, Terminology and Definition of Otitis Media

Otitis media is the most often diagnoses seen in paediatric populations all over the world. Under the term otitis media there is a wide spectrum of disease forms: Acute otitis media (AOM), Recurrent acute otitis media (RAOM), Otitis media with effusion (OME), Chronic otitis media with effusion (COME), and Eustachian tube dysfunction (ET dysfunction). Every one of these diagnostic entities is more or less precisely defined by academic paediatric and otorhinolaryngology societies, and frequently updated according to scientific research and clinical trials. Even now we have some ambiguities in establishing right diagnoses when symptoms of some entities overlap, are present together, or have been smoothly changed through pathophysiological processes, leading to transitional, simultaneous existence of symptoms of two or more diagnostic entities. The most common diagnosis in the paediatric population is AOM. Established diagnosis implies presence of symptoms, objective signs and confirmation of fluid presence in the middle ear (AAP, 2013). Recently it has become agreed that a diagnosis of AOM should not be made without middle ear effusion (MEE) conformed by pneumatic otoscopy or tympanometry. Pneumatic otoscopy is helpful for tympanic membrane evaluation; appearance, colour, translucence and mobility. Even in the presence of objective signs of the disease, the sensitivities of these signs are not the same; for some of them are very high (impaired mobility), while very low for others (e.g. slightly red; Shaikh et al., 2009). In the new guideline of American Academy of Pediatrics (AAP, 2013) the importance of the visualisation of the tympanic membrane in the presence of symptoms with conformation of middle ear fluid is also considered essential in confirming diagnoses. The symptoms related to AOM are: ear pain (described as tugging/rubbing/holding), excessive crying, irritability, difficult sleeping, decrease appetite and fever (Shaikh et al., 2009). The most common symptom is ear pain but it is only present in 50-60 % of children with AOM (Rothman, Owens & Simel, 2003). It is interesting that of children with AOM, 28-41% did not have any ear symptom (Heikkinen & Ruuskanen, 1995). Fever is present in children with AOM in 42-69% (Heikkinen & Ruuskanen, 1995), but also in most cases with URTI. Three or more episodes of AOM occur on average by age 3, 5 and 7 years in 50, 65 and 75% of

children (Casselbrant & Mandel, 1999). Of children with AOM 10-20% develop RAOM at 1 year of age (Rovers et al., 2004) but only 4.4% at age 2 (Bhutta, 2014). The diagnosis of RAOM is defined as on the number of AOM episodes confirmed in last 6 months being ≥ 3 or in the last year ≥ 4. After resolution of acute symptoms in AOM, 10-25% of children have persistent middle ear effusion (MEE) for 2 weeks to 3 months and duration of MEE shows exponential decay being the longest at age 2 years (Zielhuis, Rach & van den Broek, 1989). Presence of MEE over three months without presence of symptoms of acute ear disease establishes a new diagnostic entity, persistent OME (in some literature termed COME) (Table 1.0.).

OME peak incidence happens typically after AOM outbreaks, with OME being the last downstream stage, according to the canonical pathway of the aetiology of OM (Bhutta, 2014). Also OME can develop without AOM after upper respiratory tract infection (URTI) and Eustachian tube dysfunction, but the proportion of these is only about one quarter of all OME (1.3% of 5.5%). OME can also be the basis for re occurrence of AOM development although this sequence is a reversal of the canonical. Differentiation between persistent MEE and OME is mainly based on disease history, i.e. duration of fluid, requiring information from parents or medical documentation not often available. This transition of persistent MEE into OME can be interrupted by another AOM accumulating into RAOM within MEE fluid present between attacks of acute disease. This raises the question: what should then be the preferable diagnosis in relation to (further) treatment: RAOM or OME with super-added RAOM? If the three months of MEE persist between each attack in the year should we call it OME or persistent MEE? Should we think of a separate entity: COMBINED (i.e. OME plus superadded RAOM)?

OME is the presence of middle ear fluid without acute symptoms and typical AOM signs; often the only sign of symptom is hearing loss. Hearing loss in OME children could be expressed in range between 10-45 dB (Dempster & MacKenzie, 1991; Sente & Sente, 2000). Hearing assessment in children can be very demanding and time-wasting hence expensive for the healthcare system. Relevant history data from proxy responder, parents or caregivers, could offer a good guide for hearing screening (Study II and IV). The literature findings regarding parents' concerns about hearing are inconsistent. The

majority of articles did not find that parent's responses could predict hearing loss (Rosenfeld, Goldsmith, & Madell, 1998), while others found that parents could best predict OME (Maw & Tiwari, 1988). This is why a streamlined system of referral for diagnostic assessment is necessary. Within that, OME diagnosis can be established using pneumatic otoscope and tympanometry. Tympanic membrane colour and appearance is not necessary for establishing diagnoses but in most cases the colour is pale or blue and tympanic membrane slightly atelectatic. History data and follow-up are necessary for good treatment decisions in OME.

Chronic otitis media (COM) refers to a group of symptoms and signs as a result of long-standing inflammation of the middle ear. The chronic form of the disease is the result of untreated acute infection and it has high incidence in populations with low socioeconomic status and poor access to primary care. In the literature there is no consensus regarding what the definition and duration of the symptoms should be for COM diagnosis. Patients who have ear discharge from 6 weeks to 3 months despite medical treatment are diagnosed as chronic suppurative otitis media (CSOM; Acuin, 2004). Establishing the various diagnoses within otitis media is not easy, and it is especially demanding when we are talking about this chronic form of the diseases.

Classifications of otitis media entities vary according to the approach to the disease taken by their authors. As one of the most common diseases, OM has received attention of clinicians and scientific societies over many centuries. However attempts to approach the disease's pathology, outcomes, impacts, definition of susceptible groups and treatments systematically have not resulted in very great consensus on OM classification. Even current publications in the field of OM differ in the criteria used for classification and terminology of these common middle ear problems. While some attempt to use the pathophysiology substrate as a fundamental basis for classification (Paparella et al., 1990; Bhutta, 2014), others emphasise relatively observable anatomical abnormalities such as tympanic membrane perforation as chief distinguishing feature for COM (Nelson, 1988; e.g. for chronic suppurative otitis media – CSOM; Table 1.0.), while others base the main distinctions on anatomical locus, e.g. distinguishing the tympanic membrane and middle ear cleft (middle ear, Eustachian tube and mastoid), and holding chronic middle ear conditions to be essentially extracranial (intratemporal)

complications of the acute otitis media disease state (Bluestone & Klein, 2007). It is hard to judge any one such basis of distinctions as intrinsically better than the others, because the disease is both distributed spatially and diverse in its evolution.

Clinicians and scientists have diverse views on the pathological picture of the disease; whilst some mainly reserve the term 'chronic' for pathological findings corresponding to chronic tympanic membrane perforation and chronic otthorhea, others attempt to emphasize middle ear pathology even with an intact tympanic membrane. Using the perspective of the pathophysiological substrate, ear inflammations can be classified into two major categories: acute and chronic. The distinction between these is not sharp and a disease continuum from acute towards a chronic phase is expressed through modification of biochemical and histology changes in the mucosal layer and type of middle fluid present (Senturia, 1963). Inflammatory cells, enzymes and other inflammatory proteins and products defined the type of middle ear fluid, making the distinction between an exudate and an effusion. According to Light's criteria the effusion is an exudate if one of the following criteria is present: i) effusion protein/serum protein ratio is > 0.5; ii) effusion lactate dehydrogenase (LDH)/serum LDH is > 0.6; iii) effusion LDH greater that ⅔ of the laboratory's reference range of serum LDH (Light et al., 1972; Paramothayan & Barron, 2002). The old name of persistent OME, Secretory Otitis media (SOM) is still in use in some centres and invokes appropriately this distinction between two types of fluids, effusion and exudate (Senturia, 1963).

The acute form of OM is fast-developing and short-term usually self-limiting, with neutrophil cell predomination in middle ear and clear clinical signs and symptoms of disease (fever, pain irritability, etc). The other chronic form is slow, silent, with lymphocyte/monocyte domination and with more tissue destruction (Kumar, Abbas & Aster, 2014). The extent of tissue destruction is not the same in all chronic forms of OM and accordingly OME longer than 3 months is usually noted as persistent OME or COME (Table 1.0.), thus separating that entity from other chronic form of the disease.

Bluestone's classification of otitis media into four categories has survived 4 decades of use relatively well: Acute otitis media, Otitis media with effusion, Otitis

media without effusion and Eustachian tube dysfunction. In this classification chronic otitis media is under intra-temporal, but extra-cranial complication that any of these varieties of chronic ear condition may have (Bluestone & Klein, 2007). Just after the Second World War, the tendency was to classify middle ear disease into two groups: supportive and non-suppurative. This type of classification is still present in the International Classification of Diseases (ICD). It is notable that diagnostic naming serves a functional purpose some sub-varieties of chronic otitis media have come to the fore since the introduction of numerous surgical techniques for chronic ear treatment and reconstruction (Harkness & Topham, 1998). Under the umbrella of the UK's National Health Service Centre for Coding and Classification, a new classification was developed as a part of the British Clinical Terms Project. The advantage in this NHSCCC scheme is that OM is divided into acute and chronic forms in respect of differing prognoses. In this way the adhesive otitis and retracted forms are included in the chronic forms. Chronicity could possibly be defined according to agreements at the 2015 in the Symposium on Recent Advances in Otitis Media, but this has not been formally published. In summarising the evolution of these multiple systems it is fair to say that the classical approach to classification of middle ear disorders being acceptable by any scientist or physician for practical purposes of deciding on the surgical approach and expectation of outcome has led to excessively numerous synonyms, and that these have led to more confusion than help in standardisation of OM terms and conditions.

The terminologies most often used and recommended by societies and individuals are presented in Table 1.0. There is agreement on definition of acute forms of disease overlap considerably between societies and authors, but the diversity is greater when it comes to chronic forms. MEE and persistent OME have different courses and prognoses. It would be valuable to find ways of making a clearer distinction between them. The literature does not currently offer clinicians a practical way to separate MEE from OME. While MEE is a transient form, persistent OME is a long lasting middle ear problem with documented inflammatory mucosal changing and a tendency to recurrence.

Table 1.0. Terminology and definitions of disease entities, differences between authors and societies but also showing several agreements

Terminology	Definition and references
Acute otitis media (AOM)	❖ Rapid onset of signs and symptoms, such as otalgia and fever, of acute infection within the middle ear (Bhutta, 2014) ❖ The rapid onset of signs and symptoms of inflammation of the middle ear (Rosenfeld et al., 2013) ❖ The rapid onset of signs and symptoms of inflammation in the middle ear (AAP, 2013)
Recurrent acute otitis media (RAOM)	❖ Variable: ≥ 3 episodes of AOM in the preceding 6 months, or ≥ 4 episodes in the preceding 12 months (Bhutta, 2014) ❖ Three or more well documented and separate AOM episodes in the preceding 6 months or 4 or more episodes in the preceding 12 months with at least 1 episode in the past 6 months (AAP, 2013; Rosenfeld et al., 2013)
Middle ear effusion (MEE)	❖ Fluid in the middle ear from any cause but most often from OME and during, or after, an episode of AOM (Rosenfeld et al., 2013) ❖ Fluid in the middle ear without reference to aetiology, pathogenesis, pathology, or duration (AAP, 2013)
Otitis media with effusion (OME)	❖ Fluid in the middle ear without signs or symptoms of ear infection (Bhutta, 2014; Rosenfeld et al., 2013) ❖ Inflammation of the middle ear with liquid collected in the middle ear; the signs and symptoms of acute infection are absent (AAP, 2013).
Chronic otitis media with effusion COME (persistent OME)§	❖ OME persisting for 3 months or longer from the date of onset (if known) or from the date of diagnosis (if known) (Rosenfeld et al., 2013; Bhutta, 2014).
Chronic otitis media (COM)*	❖ Chronic otitis media is the term used to describe a variety of signs, symptoms, and physical findings that usually result from long-term damage to the middle ear by infection and inflammation (Tsilis et al., 2013). ❖ Chronic otitis media (COM), *e.g.* 'glue' ear is characterized by

	middle ear effusion and conductive hearing loss. (Preciado et al., 2010) ❖ COM is defined as a chronic inflammation of middle ear mucosa and associated mastoid cells with tympanic membrane perforations (Nelson, 1988).

Notes:

**Chronic suppurative otitis media (CSOM) is not mentioned as a separate entity as it is above all a common variety of COM, mostly present in countries with low SES and multiple predisposing factors.*

§ Very often in literature the term COME is set out to convey Persistent OME making some distinction from other chronic form of the diseases (CSOM).

1.0.2. Functional anatomy of the ear

General view - Anatomical integrity of outer, middle and inner ear are elementary precondition for normal hearing and developing speech and language. Ability of hearing is thus essential for communication and human social integration. Every part of the complex anatomical structure of hearing system has a unique function, as a result of characteristic anatomical features. The anatomy and physiology are best described together. The anatomy of the outer, middle and inner ear are not the topic of this thesis and will not be discussed in details but the physiology of the middle ear will be explained under middle ear transfer function under the subheadings 1.0.6 in the Introduction.

1.0.3. Pathophysiology of Otitis media

The etiopathogenesis of OM is multifactorial. Classification and models of causes of underlying pathological processes leading to OM should be parsimonious and wide in their scope, at the same time explaining the incidence of the diseases, possible cofounders, mediators of the disease and facets severities. The most common factor in developing middle ear infection is Eustachian tube (ET) dysfunction. Any cause of ET dysfunction leads to loss equalising pressure between middle ear space and atmospheric pressure, impaired clearance of middle ear space and transduction of fluid into middle ear space. The most common reasons are viral and bacterial infections, but also

tumours, upper respiratory tract malformations, injuries, toxic, endocrinological and metabolic disorders. The most important factors are upper respiratory tract viral infections triggering complex host –response reactions resulting in sequential realising a lot of factors, before all interleukins responsible for inflammation processes in ET and middle ear (Heikkinen & Chonmaitree, 2003). Most common viruses found in MEE are influenza A and B, parainfluenza, rhinoviruses, adenoviruses and Respiratory Syncytial virus (RSV; Chonmaitre & Heikkinen, 1997). URTI virus infections initiate immune responses leading to inflammation of the nasopharynx and of the ET mucosal layer. Swelling and impaired mucocilliary function induce ET obstruction and impaired clearance of middle ear space. ET obstruction leads to middle ear air resorption and intratympanic depression resulting in middle ear fluid transduction. Retention of the inflammatory products in the ET and spreading virus infection to middle ear space causes inflammation events, making a good basis for bacterial superinfection. Viral infection of middle ear mucosa results in denudation of the mucosal layer thus changing receptor protein expression and defensins in the middle ear space. All these changes help floating planktonic bacteria to be easily attached to the epithelial cell surface, leading to colonisation of the middle ear. Most often the bacteria found on the various epithelial surfaces of the middle ear are mainly those present in the adenoid biofilm (Streptococcus pneumoniae, Moraxella catarrhalis and Haemophilus influenza; Hoa et al., 2009). These bacterial species are equally those most often found in the middle ear effusions of children with RAOM. Interactions among bacteria and between bacteria and host innate immunity cells and their products form a complex set of dynamic processes, frequently resulting in a special form of bacterial existence, called biofilm. The complex relation of host and bacteria results in DNA skeleton formation composed by virus and host DNA particles (Thornthon et al., 2013). Inside these, communities of commensal bacteria very often exchange genetic information, forming new bacterial types with more virulent characteristic or resistant to systemic and local antibiotic treatments (Thornthon et al., 2013).

1.0.4. Aetiology

The development of OM contains many risk factors which cane for simplicity divided into two groups: Intrinsic and extrinsic.

Intrinsic factors are those responsible for genetic variations in expressing phenotypic characteristics and immunology responses of innate and adaptive immunity and thus favourin pathogenesis of otitis media. In this group we thus could place all syndromes with craniofacial malformations, deformity and dysmorphic cranial structures which could lead to otitis proneness due to dysfunction of Eustachian tube (ET). RAOM and OME incidence are higher in children with cleft palate than in normal children (Flynn et al., 2009). Abnormal innate immunity affects many subsystems: mucosal cells, neutrophils, macrophages, mast and fibroblast. These are collectively responsible for recognising pathogens and presenting to other cells responsible for adaptive immunity. The cells involved in innate immunity responses detect pathogens using pattern recognition receptors (PRR). The PRRs are presented as Toll like receptors (TLR), cytoplasmic-nucleotide-binding-oligomerization-domain (NOD)-like receptors (NLR), retinoic acid inducible gene (RIG-1) and C type lecithin receptors (CLRs; Mittal et al., 2014).

Extrinsic factors are numerous and their activity often co-presents with intrinsic factors. Extrinsic factors are: socioeconomic factors (SES), number of siblings, day care attendance, tobacco or air pollutants exposure, bottle feeding, using pacifiers, season of the year (Heinrich & Raghuyamshi, 2004). Whilst there are pathogenetic bases for specific direct effects, these can also be mediators in the cascade processes in development of the disease, because they have raised risk factor prevalence within the subpopulation mostly susceptible to the disease. Susceptible populations are mainly those with present intrinsic risk factors but also very young children (< 2 years of age), children in day care attendance, from families with a high number of (elder) siblings and low SES. When applying risk factors check lists to children with the disease, rare to observe only one RF and having more than three is quite common. The peak incidence of OM is in infants, 6-12 months (Biles, Buffler & O'Donell, 1980; Hoberma et al., 2002).

The raised incidence of otitis media in very young children is due to several crucial characteristics of anatomy and physiology at that age; position, size and length of ET are different in very young children, especially in infants and toddlers. ET in very young children is short, horizontal and open (Cayé-Thomasen et al., 2013). The Cartilaginous

part of the Eustachian tube is not as strong as in adults and the peritubal muscles are composed of fast fibres called 'white muscle fibres' with anaerobic glycolysis which could be problematic in infants (Tomoda et al., 1984). Problem in anaerobic glycolysis in infant's fast muscle fibres is the result of lower glycolytic enzymes capacity (phosphofructokinase and lactate dehydrogenase) than in older children and adults leading to decrease anaerobic power (Kaczor et al., 2005). The point of the insertion of tensor palatine muscle (MTVP) to the ET is considered the narrowest part of ET (Rood & Doyle, 1978). The angle between MTVP and longitudinal axis of ET differs between infants and adults. The level of surfactant on the mucosal surface of ET in infants is lower than older children. Open and wide ET allows fast spreading of microbes from nasopharynx to middle ear. Before age 3-4, immunity is not mature, making local defence mechanisms unable to localise pathological process (Sharma & Pichichero, 2013). The sutures between bones are still fibrous and the nervous system immature with dominant hyper-reactivity of autonomic nervous system (Landrot et al., 2007). Thus infants with OM could easily develop systemic reactions and be more prone to complications than older children and adults.

In these complex pathogenetic sequences, extrinsic factors could mediate the extent and frequency of events by causing switching between two phenotypic forms of the disease, coexisting sometimes together with overlapping symptoms of both disease forms. Knowing all possible influences in the complex pathophysiology sequence offers better understanding of all aspects of presentation of the disease. In AOM, the resolution of symptoms is relatively fast but inflammatory exudate in the middle ear can be present for few weeks after. In RAOM, the sequence of events is relatively fixed but the number of AOM during the year is > 4. From the pathohistological view the exudate are composed by inflammatory cells and its products and planktonic bacteria. This form is often present in AOM. The effusion (Introduction 1.0.1) is a histologically distinctive formation with commensal bacteria in the matrix of DNA nets from the viruses and neutrophils with other inflammatory products in.

1.0.5. Epidemiology

General descriptive epidemiology - Diagnostic criteria, seemingly precise in definition according to guidelines based on literature in fact do not correspond precisely to presenting disease. The most common diagnosis in primary care is AOM. In some children, the course of the disease is prolonged and so justifies the description as RAOM or OME, and in yet others as CSOM. Therefore OM covers a spectrum of diseases phenotypes where each form presents as a stage of causally related pathophysiological processes (Bhutta, 2014). The incidence of each varies widely depending on presence of previously mentioned risk factors (RF) and for sure the incidence of the driving events: URTI converting to AOM. The sequence of pathophysiological processes is explained in the Study I where the of conversion form one stage in the chain to the next is introduced. Thus each phenotypic form of the disease in causal sequence does not exclude the previous form (Bhutta, 2014) and in many instances even implies it (Figure 1.0.).

If the probability of occurrence of any state at a late stage in the causal chain is not to be very low, then the probabilities of conversion to the next stage must all be fairly high. Proneness (Howie, Ploussard & Sloyer, 1975), which we would now refer to as extreme RAOM, refers to not to a single risk factor but to the aggregate or average value of many risk factors (RF). In addition to the classical RF mentioned in the next subheadings, the incidence depends nowadays also on the level of immunisation for Streptococcal diseases (Vaccine PV7, PV10 or PV13) and on the organisation, of the Public Health System (i.e. the availability of preventive services linked to appropriate health care for denied cases). We can expect in that way differences in disease profile presentations between countries and centres.

Different aspects of otitis media and their relationship - Causal cascade relationships between different pathologic forms of OM are presented in the section on pathophysiology but here extended to all aspects of the disease and consequences including downstream ones naturally viewed as ‘outcomes’ of the disease. The OM affects child’s and parent’s QoL and that influence can be measured by describing and quantifying various aspects of the disease. Physical health and ears symptoms are

dominantly present in URTI, AOM and RAOM, but hearing loss (HL) in children with MEE and OME (Study III-SEM model). Symptoms of the physical health, ears and HL affect child sleep pattern, behaviour, speech/language and schooling (Bennett, 2001). Measures of the severity of the disease can be used as outcome measures in routine service monitoring or in more formal clinical trials. Hearing loss for instance could be measure of severity, but also RHD (reported hearing difficulties) could be measure of hearing severity, presenting only one of the overall aspects of hearing. The aspects of OM are measures which collectively provide a disease profile. We need to define them, proved measures of severity of each facet and use them in prediction of the outcomes but also for prevention and treatment decisions. Large studies using psychometrically developed questionnaires with well-selected and appropriately scaled and weighted items can give a comprehensive view of all the disease aspects and its relationships. One such questionnaire is the OM8-30 with its younger short form OMQ14 (Timmerman et al., 2007; Timmerman et al., 2008).

RFs can usefully be seen as mediators and catalysts of the pathologic processes leading from one form of the disease to another, influencing severity of the aspects of the disease and its outcomes. The causal cascade is started with URTI (influenced by seasonal climate changes connected with other RFs, virus transmissibility, antigen drift and host susceptibility among others on. One of these could be lack vitamin D, closed space with few air changes per hour due to heating, and other physical triggers present in school communities. The causal cascade is explained in pathophysiological terms, but later we can use these stages to describe all profiles of the disease. Some of them are closely interrelated (HL, RHD) and their severity influences child behaviour, speech and language. RAOM and physical health due to URTI affect child sleep pattern and parents' quality of life.

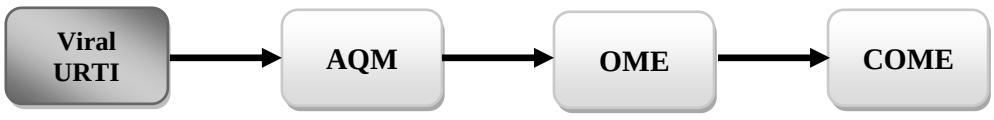

Figure 1.0. Canonical Pathway of the most common forms of OME (Bhutta, 2014)

Incidence and prevalence - The overall cumulative incidence of AOM for children < 3years of age is 50- 80% (Teele et al., 1984; Casselbrant & Mandel, 1999). The incidence in the age group 0-8 years is around 40% (Biles, Buffler & O'Donell, 1980). The same authors found that incidence of OMA is much higher in the younger age group and significantly differ between 0-1 year of age and 5-6 years of age group. This could be the result of starting the school year and thus spreading the disease vectors, viruses and bacteria in the school communities. The incidence in children < 12 months is more than 90% (Biles, Buffler & O'Donell, 1980; Bardach et al., 2011). Children who experienced AOM before 13 months of age are more otitis-prone (14%), having four or more episode per year, than children who had first AOM episode after 12 months of age (5%) (Biles, Buffler & O'Donell, 1980; Daly et al., 1999). AOM incidence and prevalence reduce with age, after about the 2nd year of age. Approximate prevalence of OM for children from 1 year to 7 years of age is 30-65%, but RAOM prevalence according to the same author was only 5-12% (Teele et al., 1989). Using a mathematical model for AOM and RAOM estimated predictions and fitting to the data of the two large studies, the prevalence rates were higher (40-80%) for AOM, but lower (3.5-5.5%) for RAOM than in other reported studies (Bhutta, 2014). It is thus possible that publication bias in smaller studies, coupled with difficulty of follow up studies and definition of observation period according to event occurrence has given a higher impression of the frequency of RAOM than is justified. This prediction model also matches the age function with its two peaks; one around the end of the first year of life (2.16) and second around school entry, 5-6 years (2.46) (Bhutta, 2014) agreeing with other findings (Biles, Buffler & O'Donell, 1980). The prediction model for RAOM prevalence done by same author showed the same trend, higher in the first year of life and before school entry. The prevalence of OME in Italian children is estimated around 14 % and younger children are more prone for bilateral OME (Marchisio et al., 1998). Other authors emphasize that the OME prevalence in the first year of life is 3.2% and adjusted for number of births per month is 4.4% (Rovers et al., 2000). This prevalence is the highest in April (late winter) and lowest in October emphasizing the importance of season when consider incidence of the disease forms of the unique canonical pathway of OM (see Study I). Prevalence prediction from mathematical model suggested that

percentage of OME preceding AOM is 4.2% (Bhutta, 2014), while 1.3% does not have AOM in past history.

General stratifiers (age, SES, season, gender, country) - The incidence and prevalence of different pathological AOM forms depend of the presence of RFs. Effect of those factors listed in the next sub-headings are of differing strength and generality, or may have differing amounts of information available, so I allocate them differing amounts of space. The most influential RFs besides age are: country (culture, latitude, climate, customs), season, history of OM and SES. Explaining incidence and prevalence without considering age is unrealistic for almost any disease. OM facet severities differ between countries. These differences are results of different latitude and climate, culture and availability of health care. Therefore we see large differences between countries in described incidence and prevalence of the disease and even differences between different parts of the same country (Ting et al., 2012). The SES effect on incidence and prevalence may be modest in size but can be very important in modulation some aspects of the disease explaining sometimes higher incidence with low SES. High SES families recognise OM symptoms and seek health care on time, and in many countries this means private physician clinics in most instances. Effect of SES on incidence can be explained partly with low income and poor life conditions. Its effect on incidence and prevalence is not constant and not often significant, but in understanding the differences between countries should be considered and controlled (Kong & Coates, 2009; Zhang et al., 2014). For convenience and non-intrusiveness, SES is very often documented approximately via level of the mother's education. Gender differences in the population are seen with boys being more susceptible, but this difference is not always present (Casselbrant & Mandel, 1999). The effect of gender seems more influential in younger children < 5 years of age than in school-aged children (McFadden et al., 1985). I later show that other OM facets, RHD, behaviour, school, HL, have distinct seasonality in forms that can be readily incorporated in a model of causal sequence of OM impact. Seasonal effects started with rising URTI incidence in late autumn, initiating the canonical pathway and logically expected sequential chronological unfolding of succeeding OM aspects (Figure 1.0. and Study I).

Risk factors for incidence and prevalence - The number of reported risk factors on incidence of OM has risen during the past decade. The novel genes are discovered responsible for the syntheses of proteins and glycoproteins involved in innate immunity or adaptive immunity in pathogenesis of OM. We need better solution for longitudinal and cross sectional studies of paediatric populations and close follow-up of OM demographic characteristics and impacts (Timmerman et al., 2007; Timmerman et al., 2008; Rovers et al., 2002). The number of intrinsic (host) and extrinsic (environmental) risk factors rise every year. The most important are: children with altered host defences and underlying diseases (craniofacial abnormalities, including cleft palate, children with atopic constitution, autoimmune disease, immunodeficiencies), specific gene abnormalities (TLR 4, FBX011), race/ethnicity, gastro-esophageal reflux, perinatal factors (low birth weight, LBW: < 2500 gr, prematurity < 37 gestation weeks), age, history of ear infection, using pacifiers, season of the birth, season, exposure to tobacco smoke or air pollution, SES, number of elder siblings, day care attendance, insurance coverage (Hoffmann et al., 2013; MacIntyre & Heinrich., 2012; Macarthur et al., 2014). The influence of some of these factors is stronger than for others but their influence also depend upon presence other risk factors. With improving technology, diagnostic precision improves, requiring new statistical approaches with better study design. If properly executed the research programme built on new technologies can widely improve knowledge of the disease parameters, prediction of the outcomes, screening, prevention and treatment.

Children with immunodeficiency, craniofacial anatomical abnormalities, allergy, and chronic diseases are more at risk for AOM, RAOM and OME than other children (Leach & Morris, 2001).

Race and ethnicity influence the prevalence of OM, but the results are not uniform and could be wrongly interpreted because most of the ethnic minorities in established societies able to undertake the research have lower SES. The currently prevailing weather patterns, closed household lifestyle with more elder siblings in the family, and lack of accessibility of health care may all raise incidence and prevalence of the disease. Despite these differences, there are some close communities with higher incidence of the disease. The prevalence of AOM, OME and CSOM was very high (25%, 42% and

15% respectively o in some Australian Aborigine children communities age 6-30 months (Morris et al., 2007). Inuit children have more AOM but not CSOM than other children (Julien et al., 1987).

Age is a very strong risk factor for OM. The peak of incidence of AOM is in the first year of life, between 6-12 months (Biles, Buffler & O'Donell, 1980). RAOM and OME follow the same trend, with the incidence of RAOM also higher in first two years of life than in older children (Teele, Klein & Rosner, 1989). The reason of the age influence is explained in more detail in the pathogeneses study.

Premature, high risk infants and low weight infants have higher prevalence of OM for 20% than normal infants (Engel et al., 1999). The apparent basis of raised risk for OM in preterm infants is lower antibody titer and immune immaturity than in normal infants.

Children born at the second half of the year have more likelihood of for a RAOM but this appears not to hold for all- AOM (Biles, Buffler & O'Donell, 1980). A lower educational level in the mother and attendance of day care for 13-23 months increase the risk for RAOM, while higher age, breastfeeding > 6 months decrease risk for RAOM (Hoffman et al., 2013). The incidence rate of AOM and RAOM is the highest in January and February (Biles, Buffler & O'Donell, 1980) and the incidence of RAOM rises with recurrent URTI infection which are more frequent during the winter months (Hoffman et al., 2013). Low SES increase risk for URTI, RAOM and OME (Zhang et al., 2014). Low SES can predict susceptibility to URTI and more often is associated with short telomere of T lymphocytes (Cohen et al., 2013) probably responsible for recurrent URTI infection. The finding is not consistent between authors and mainly investigated in adult population.

Reflux of oesophageal acid in infants rise incidence of AOM causing ET dysfunction and obstruction (Tasker et al., 2002).

A positive history of OMA before age of 13 months is a strong predictive factor for RAOM (Biles, Buffler & O'Donell, 1980). The number of repeated ear infections at very young age (0/1 years of age) could be reason for hearing problem in older children

reported by parents (6/7 years of age; Yiengprugsawan, Hogan & Strazdins, 2013) and these children usually signified as 'otitis-prone' (Howie, Ploussard & Sloyer, 1975). Children who use a dummy (pacifier) are at raised risk for more than three episodes of AOM than other children. The pacifier user younger than 2 years of age have more than 3 episodes of AOM in 29.5% of children while non users in 20.6% children (Niemelä, Uhari & Möttönen, 1995).

The causal sequence of OM started with URTI. URTI incidence has strong seasonality with late autumn peak and spreading till late winter. URTI infection are usually initiated by varies viruses which are listed in previous study under pathophysiology. They often cause AOM during the winter triggering the sequence of inflammatory processes leading to bacterial superinfection, MEE and /or OME. Some of them, particularly younger children, develop RAOM. Seasonal changes responsible for URTI differ according to the temperature, latitude and humidity. Transmissions of the viruses and their virulence are results of seasonal changes (surface antigen drift). Season does not only influence incidence and prevalence of the disease but affects all other aspects of the OM and its severity. This different form of seasonality is taken up in detail in the next study.

Burden of Otitis Media - Otitis media is the one of the most common diagnoses in paediatric visits and reason for antibiotic prescription (Grijalva, Nuorti & Griffi, 2009). Despite the more conservative recent recommendations in guidelines from medical and scientific societies, antibiotic prescription for AOM and RAOM is still high in general physicans' and generalist paediatricians' practices. Antibiotic over-prescription resulted in risen number of penicillin resistant serotypes of Pneumococci and βlactamase negative penicillin-resistant *Haemophilus influenzae* species (Pichichero & Casey, 2007; Hotomi et al., 2006). Shift in middle ear pathology resistant serotypes has led to more complications and chronic forms of the diseases. The serotypes shift is indeed present in *Streptococcus pneumoniae* after introducing pneumococcal vaccines. The levels of antibiotic prescription and thus of bacterial resistance differ between countries, and the number of complication still high in some ethnic groups and communities such as indigenous Australian, and indigenous American (including Inuit Children. The global cost of OM cases in US has been calculated as approximately $3-5 billion,

including physician visits and indirect costs such as parent sick leave (O'Brien et al., 2009).

The chronic forms of OM, especially OME, are the more frequent reason of hearing loss in young children at the age most sensitive for speech and language development, social integration and schooling (Bennett et al., 2001). The effect of HL on child development depends on its duration or history, age of the child, intrinsic risk factor presence and SES. Presence of these risk factors adds to and may modify the overall hearing impact on child quality of life. For better understanding of the complex effects of poor hearing on overall child wellbeing, requiring good reliability and validity of the hearing measure. But the literature contains little discussion of what precisely is meant by 'hearing'. To address this, we can use along with HL two other forms of the hearing: ACET and RHD. Each of these measures can reflect one aspect explain part of hearing not present in the other form and vice versa. As the hearing is not static, the cumulative effects of poor hearing over time will be better observed by using more than one measure. This point about importance of the duration of the history is well illustrated in longitudinal series where reported ear infections in preschool age reflect hearing problem in subsequent years (Yiengprugsawan, Hogan & Strazdins, 2013). Small studies and/or the short follow-up period are not enough for observing the wider OM impacts and the influence of severity of the disease upon them.

The influence of reduced HL on child development and behaviour is not the only pathway to OM impact in children, despite an almost exclusive emphasis on this pathway in the literature. In OME and super added RAOM there can also be sleep disturbance, with consequent effects on the child's scores and the parents' quality of life. Lack of sleep can also affect a child's cognitive performance and hence behaviour and schooling in other ways. However these parallel pathways to developmental impact may have a somewhat shared origin in ENT disease. Frequent URTI infection and RAOM coexist with infection/obstruction pattern of child's nasopharyngeal adenoid tissue producing obstructive sleep apnoea (OSA). OSA leads to low oxygen saturation, tiredness, loss of daily concentration and school report. RAOM and frequent ear pain disturb sleep patterns and so affect child behaviour. The presence of risk factors affects not only prevalence of disease but HL, RHD, speech delay, schooling and general

impact (Vergison et al., 2010). Thus appropriate outcome measurement in research, and to an extent comprehensive assessment in clinical practice, requires not only a defined disease profile but some form of profile of the impacts also. In the long term it is a reasonable hope that knowledge bases in these fuller assessments will pay off in quality of health care and particularly in the effectiveness of treatments, for example by using particular measures of severity in one or more facets as basis for prompt treatment.

1.0.6. Middle ear transfer function

General view - The preceding anatomical introduction sets a background for the transfer of acoustical energy leading to hearing, and the transfer mechanisms with which middle ear disease interferes, so causing hearing loss. This is chiefly through impedance loading of the eardrum by the fluid behind it, although there are other detailed mechanical consequences of the general pathology and its variants. The detailed physical processes of middle ear sound conduction have been understood for some decades, but an understanding of them is necessary to appreciate the justified basis of the measures used in the present work, of which the focus is hearing in otitis media. Middle ear transfer functions explain influence of middle ear pathology on acoustic absorption and reflection by changing physical parameters in the system. The outer ear canal and middle ear of adults and children (> 1 year) can be considered as a linear mechanical system. Linearity means that the output acoustic pressure response is proportional to the input pressure (Rosen & Howell, 1991), and that the output/input relation does not change with the absolute level of the stimulus. For any acoustic travelling wave, the reflected wave stands in constant proportion in linear mechanical systems. The resulting acoustic pressure in the ear canal is then the sum of the forward and reverse travelling wave. The ratio of the amplitude of reverse travelling wave to the amplitude of forward travelling wave is called pressure reflectance $|R(f)|$ (Robinette & Glattke, 2007). Similarly energy reflectance is defined as the squared magnitude of the pressure reflectance $|R(f)|^2$ and is the ratio of output (reflected) energy and input energy. Incidental energy (input energy) is equal to the sum of reflected and transmitted energy. The ratio of any output (reverse wave components) signals or variables to input signals (forward wave components) or variables is called the Acoustic Transfer Factor (ATF). Acoustic admittance, acoustic impedance and acoustic reflex test are all immittance

measures and also measures of the ATF. Acoustic admittance Y(f) is defined as the change in acoustic volume velocity per unit pressure at the tympanic membrane (TM), or in other words, the ratio of the total volume velocity to the total acoustic pressure. The opposite of acoustic admittance is acoustic impedance Z(f) which represents the change in acoustic pressure per unit change of acoustic volume velocity. According to these ATF is defined by five equations (Rovers et al., 2004):

$$ATF = output/input \quad (1)$$

For the middle ear considered as a single acoustic model the transmission of acoustic energy is defined by the three measures: mass, elasticity and friction. When we change the acoustic pressure in the outer ear canal, and hence the pressure drop across the eardrum, then each of these elements acts in a different way and the overall resulting transmission is given by the sum of responses for each of these three impedance components. Total Impedance is the sum of the impedances of the mass [Xm], stiffness [Xe] and resistance [R] (friction). The impedance of the mass, Xm is proportional to the mass of the system and frequency, but the stiffness impedance (Xe) is inversely proportional to frequency (Bess & Humes, 2003).

$$Xm = 2\pi fm \quad (2)$$

$$Xe = 1/2\pi fE \quad (3)$$

$$Z = [R^2+(Xm-Xe)^2]^{0.5} \quad (4)$$

We can say that the total impedance is proportional to the square resistance (R) and reactance (X = Xm-Xe). Admittance is not a scalar but a vector quantity with two orthogonal elements, conductance [G] and suseptance [B]. Total admittance is the sum of conductance [G] and suseptance [B]. Suseptance has two dimensions, susceptanace of the mass [Bm] and susceptance of the stiffness [Be]. Total admittance is therefore:

$$Y = [G^2+(Bm-Be)^2]^{0.5} \quad (5)$$

The response of the frequency-dependent forces and frequency-independent forces is the vector value, and the angle between them phase angle [ϕ], in effect a delay as

energy is stored and released by the mass and stiffness components. Susceptance of the stiffness and of the mass are 180^{o} out of the phase. The total susceptance [B] as the sum of the Bm and Be determines wheather the middle-ear transmission system is mostly stiffness-controlled ($> 0^{o}$) or mass-controlled ($< 0^{o}$). When the two components of B are equal, then the B is 0 and the system is in resonant state. Below the resonante frequency the middle ear is described as stiffness-controlled and above the resonante frequency as mass-controlled. An increase in the stiffness (as with otosclerosis or negative middle ear pressure) decreases transmission of low frequencies. Increase in mass raises mass reactance which opposes transmission of high frequency. This explains why in mild otitis media (OM) with eustachian tube (ET) dysfunction, or change of aeroplane cabin pressure, the middle ear pressure drop attenuates chiefly the low frequencies; however, with advent of middle ear fluid, the extra mass and friction loading effects all frequencies and slightly more the high frequencies. In consequence the average audiogram in OM has a very shallow peak at 2 kHz and dips above and below.

The sign of an admittance Y is always positive and its amplitude reaches maximum at the point where Be and Bm are equal. In normal ears this is usually between 800 and 1200 Hz (Shanks & Shohet, 2009). In standard tympanometry, the single frequency of 220 Hz is a compromise but in children of two years and over, it is a useful one for summarising the energy transfer into middle ear.

Since the late 1960s the assessment of middle ear function has been greatly aided by the development of somewhat standrad tympanometry, which gives incomplete but essential information in the two main forms of OM. Acoustic immitance tests tell us what amount of acoustic energy is absorbed into the middle ear, but provide limited information about hearing ability, because processes in the inner ear (cochlea) and central nervous system are also involved. Furthermore the cochlea is not a linear system and a precise understanding of the complex acoustics of cochlear transmission and hair cells transduction is still obscure in some of its details. I return to this divergence between measured hearing and middle ear mechanics in Study II. Even considering the most basic aspects of tympanometric shape, the relation to middle ear pathologyis not simple and direct [x] (Smith et al., 2006). Different middle ear pathologies can have the same form of tympanometry and also different forms of tympanometry can accompany

particular middle ear pathology (Sichel et al., 2003). The sensitivity and specificity of tympanometry in evaluation of middle ear status in hearing depends on parameters used for evaluation, age, history and calibration. The most important tympanometric parameters in practical assessment of middle ear condition are: Ear canal volume (Vec), Acoustic admittance (Ytm), Tympanometric peak pressure (TPP), Tympanometric width (TW), Phase angle (ϕ).

Classification system - The tympanometry classifications by Jerger (1970) are made with combinations of the TPP values, Y and shape of tympanometric curve into A, B, C, D and E form of tympanograms. The set of types captures the non-linear relationship between severity (as measured by Hearing Threshold-HL - in a large relevant sample) and TPP and between severity and Y. This fact explains their wide clinical use. In turn, the clinical familiarity and ability to check against other data for inconsistencies explains why in the Eurotitis-2 study, without intensive paid research support data were gathered in this categorical form rather than in terms of the underlying continuous physical parameters (see General Method). Without this appreciation of simplicity for reliability in data acquisition, the re-conversion of categories into continuous estimates of HL would seem perverse.

Type A form: Y in the normal range (0.2-2 cm^3), TPP distribution between -50 to +50 daPa in adults and -150daPa to +150daPa in children. Two variants are sometimes distinguished: As (Y = 0.2 cm^3) and Ad (> 2.0 cm^3);

Type B form: Completely flat tympanogram or Y< 0.2 cm^3;

Type C form: Y > 0.2cm^3 and TPP < -50 daPa in adults and < -150 daPa in children;

Type D form: Normal tympanogram with a notch in the top;

Type E form: W shape of tympanogram.

The above classification did not consider effects of wide range pressure difference in the middle ear, mostly within the C form of tympanograms, which Fiellau-Nikolajsen et al. (1977) offered some years later by distinguishing C1 and C2:

A: Y in the normal range and TPP between +100 and -100 daPa;

B: Flat tympanograms or Y < 0.2 cm^3 and/ or TPP < -300 daPa;

C1: TPP between -101 and -200 daPa (Y > 0.2 cm^3);

C2: TPP between -201 and -300 daPa (Y > 0.2 cm^3).

In recent years a type F has been added to provide a physical criterion for a non-closed tympanic membrane for use clinically to reflect: the patency of ventilation tubes (tympanostomy tubes) or presence of a perforation (Higson & Haggard, 2008)

We use modified Jerger classification due to its simplicity and possible different influences of large TPP values on hearing threshold, as explained in the later section on 'ACET' (MRC Multi-centre Otitis Media Study Group, 2009).

Tympanometry as a measure of estimation middle ear transmission capacity - Even tympanometry is only a rough surrogate measure of hearing, it captures only middle ear sound transmission, but this is the main determinant of hearing threshold in a population not assumed to have other major determinants so is useful in OME and to an extent in AOM Using appropriate test tone and pressure direction, adjustments for age, and with careful interpretation of the data, tympanometry is still very objective, reliable and quick, and so clinically irreplaceable in diagnostic assessment, also in specific decisions, treatment approach, differential diagnoses in marginal cases and postoperative follow up. In children > 3 years of age tympanometry can be done with test tone of 220 Hz as in adults because the outer and middle ear could be considered stiff control system. The younger end of this range as well as some particularly anxious older children may not well accept the probe or dislike the pressure sensation with dulled hearing as ear canal pressure is varied. In suspect cases for OME, it is logical and economical to avoid a battery approach and to reserve audiometry for cases failing some tympanometric criterion such as (B, C2), but acceptability in these young cases may limit the general applicability of such case triage. For infants, the 1000 Hz test tone is preferable until about 4-6 months of age when the outer ear canal walls are no longer collapsed.

Using tympanometric parameters for estimating hearing threshold-Air conduction estimated threshold – ACET - The modified Jerger tympanometry offers two subcategories of type C i) C1 with TPP between -100 and -200 daPa and ii) C2 more negative values TPP between -200 and -300 daPa. This modification is more appropriate for clinical practice because slight negative middle ear pressure is quite common. Ears with C1 tympanometry nearly always have normal hearing threshold (HL < 20 dB), and usually less, similar with those with A tympanometry, while those with C2 can be slightly over 20 dB. The distinction is prognostic as well as metric. In both categories we have relatively normal Y and it means that the values is equal or above the 0.2 cm^3. The hearing threshold in the ears with B tympanograms can be as low as 10 dB HL but is usually above 20 dB with a maximum in the region of 45 dB HL (Sente & Sente, 2000). These generally appreciated relationships were formulated into a means of predicting HL, imperfect but useful, known as ACET by the group conducting the UK Trial of Alternative Regimens in Glue Ear Treatment (TARGET; MRC Multi-centre Otitis Media Study Group, 2009) on a sample of well over 1,000 cases having complete data in two variables. Binaural hearing prediction from formula based on 4 categories of tympanogram per ear and adjusting for age accounted for 49% of the variance in HL, equivalent to r = 0.696 between the predictor (ACET) and the true HL. The ACET formulation is explained in more detail in General Method where a sufficiently close replication of this relationship (r = 0.598) is reported for Eurotitis-2.

Among cases referred chiefly for suspected OM-associated hearing loss, the tympanometric states and the HL values on the two ears are strongly but not perfectly associated. The development of ACET formula revealed an aspect of monaurality/bnaurality deserving further consideration. This interaction or conditioning effect from the middle ear status of one ear to the hearing level on the other with flat, B tympanograms, showed 6-9 dB HL worse hearing when the 'other' ear was also a type B. This conditioning or interaction effect could explain away some of the classical noted wide range of HLs for a B tympanogram (unilaterally) as a failure to think in whole-child terms. This interaction probably not the involvement of any true synergistic mechanism, but just the selection of a sub-population with a more serious physical effect on hearing or going through a more severe stage which both ears reflect to some extent , but not in a perfectly correlated way.

For further studies of hearing ability and developmental sequel in OME thinking beyond monaural terms is entirely welcome, beyond measurement considerations in tympanometry but the evidence that much deeper insight and accuracy is obtained by going beyond binaural averages is not strong. The ACET formula can be implemented in clinical practice as a fast approximation to binaural hearing status when the equipment for audiology testing is not available, as well lack of staff and/or time. Together with RHD questionnaires and history (Study II) could be used in diagnostic algorithms, in hearing screening and in estimating OM impact in large-sample epidemiology where intensive measurement would be impractical and costly. The present work in the large Eurotis-2 sample has this general practical aim as a long-term goal. The challenge is to propose specific uses and arrange objective evaluation studies of the advantages that its use offers.

1.1. General background to the need to study seasonality

The seasonality of disease has been almost entirely a phenomenon of infectious diseases, although recently research on gene expression has powerfully extended seasonality analysis to diseases not thought of primarily as infectious ones (Dopico, 2015). Seasonality mostly implies a periodic process with a peak around the same time in the annual cycle, but as periodicity is defined as 'state of recurring at regular intervals' (Stevenson & Waite, 2011) a disease with 2 or even 3 reliably located peaks within a year could still be defined as seasonal. Temperature, as a result of solar angle, changes over the seasons, bringing also differences in relative and absolute humidity. These are the main drivers for seasonality of infectious disease, because of the climatic conditions favourable to survival of microbes or to their transmission (e.g. particular behaviours of hosts and vectors). Although the climates of the two hemispheres are indeed not in exact inverse relation, for gross seasonality patterns I here handle them as though they were (e.g. by simply inverting, i.e. 6-month delay, the data which come from New Zealand). Seasonal patterns can differ in the same hemisphere according to latitude, attitude (Bloom-Feshbach et al., 2013) and humidity (Shaman et al., 2010). The regional differences in the above main climatic factors are often responsible for different onset of seasonal epidemics in the same country, but Japan and some other parts of East Asia have two annual peaks of influenza with the total incidence of disease being higher than in other parts of the world, even parts where the level of humidity is equally high (Chiu et al., 2002). Habitual lifestyle during cold winter weather, decreased solar insolation, melatonin production and vitamin D uptake are further possible reasons for increased host susceptibility and a drop in strength of the immune response. These seasonal environmental changes can influence not only exposure and susceptibility, but also virus virulence through increased virus survival time. The virus antigen shift and drift happen periodically, every year, seasonally (Sangket et al., 2015) and also over longer time intervals (years); the exact mechanism of this change is still obscure.

In temperate latitudes (including the highly populated parts of Europe, North America and East Asia) the seasonal spread of virus infections leading to otitis media happens every year at a fairly similar but not identical time – within about 3-4 weeks around November as the temperatures drop, but in accordance with current weather and

recent virus evolution in population centres. The exact complete list of drivers for this annual cycle is still not entirely clear. We can expect regular seasonal incidence for most viral infections thought causative for OME, but sometimes small variations in the rate of replication can create large changes in disease incidence, resulting in pandemics. As well as seasonal incidence of URTI caused by viruses, we can observe the consequent cycle in OM and its forms and sequelae (the 'canonical causal pathway'). A central tenet of this thesis is that the pathway should be seen as including not just the traditional pathological and pathophysiological states used as signs to define diagnosis but also sets of symptoms corresponding to these signs. The short-term frequency and severity of the various aspects (facets) of OM within diagnosis, and of the consequent impact on the child and family (sometimes referred to as 'outcomes') will therefore unfold in a fashion showing dependences that can be at least interpreted as causal, and seasonal fluctuations will therefore be seen in large samples of such data.

URTIs are the antecedent of AOM and other phenotypic forms of disease. Although RAOM and OME have to be defined over a longer period with some consequent delay from the peak for URTI, it is the seasonal incidence of URTI which drives the subsequent OM incidence, in all its forms. In epidemiological terms, season can be seen as a package of interrelated causal influences: more than low solar angle: as a complex of related physical features, presence of specific risk factors, characteristic of virus virulence and host susceptibility. These will result in an incidence peak of viral infection and hence in incidence of related diseases. Knowing the antigen switching of the virus and the general risk factors is essential for understanding impacts, limiting morbidity, and for the long-term planning of successful prevention. Within this, seasonality carries implications for optimum vaccine schedules, and potentially for seasonally adaptive criteria in clinical assessment.

1.1.1. Seasonality of case incidence versus severity of disease and impact facets within case presentation

Almost all studies of seasonality in infectious disease concern season of occurrence of a small number of related diagnoses, quantified as case incidence. However there are two problems, for fully understanding what seasonality really represents, in using only

case/control status as the dependent variable. (1) Obviously there is the universal problem of who should be considered as controls and whether some of the observed seasonal variation may in fact lie not in aspects of disease but in referral criteria of patients and of primary care for what should be considered a case of a diagnosis; this can be handled in part by imposing a second more narrowly defined diagnosis within the clinic sample but this still leaves the pattern in the population undocumented, does not document the factors in uptake of healthcare (e.g. referral criteria) and reduces feasible sample sizes. Unless special precautions are taken (e.g. use of objective measures, analysis according to multiple definitions) the information contained in an act of diagnosis can be generally confounded by variation in physician criteria or by variation in family criteria for consultation, which degrade general data quality and so degrade study power. (2) As in the analysis for all research problems, the fact that a diagnosis is categorical will generally waste useful information. Even if there is no particular reason to believe these two problems will especially confound analyses of seasonality, it is worth attempting to explore seasonality through the severity patterns of aspects of a disease. The length of this introduction is required to set out this novel and in some ways unconventional approach. The aims would be: (i) to achieve an account which reconciles what is learned with the existing understanding of seasonality based on case incidence, and (ii) to enrich the understanding of clinical presentations and of causal links between these aspects in some detail, not only between, but within diagnostic case-types. Two implications follow which need to be addressed: (a) the findings are no longer about occurrence or otherwise of disease, but about differences within diagnoses in severity or profile, and (b) to achieve this aim we will need: i) large samples; ii) continuous measures of severity in each facet of disease, these being of greater discriminative power than mere diagnostic categories can offer; iii) sufficient data items in each measure for reliability, and iv) the constraint of explicit and stated hypotheses about timing (or more correctly, a plausible narrow range of timings) to protect from imaginative over-interpretation of random patterns in data. This fourth requirement is sometimes referred to as analysis being ‘theory-driven’.

About 30-60% of URTI infection result in AOM for the high-incidence age-range (6 months to 3 years) (Chonmaitree et al., 2008; Revai et al., 2007) indicating high prevalence of the risk factors that convert the first into the second. Viewed

retrospectively, more than 95% of AOM cases have identifiable (i.e. symptomatic) antecedent URTI infection (Chonmaitree et al., 2008). It is well known that peak incidence of URTI in late autumn is followed by peak incidence of AOM in winter (Heikkinen & Chonmaitree, 2003) and incidence of OME (Kontiokari, Niemelä, & Uhari, 1998) in late winter-early spring. The present research goes behind the generality of these preceding statements, linking pathogenesis to profile of OM presentation (i.e. the severities of the various facets of the disease) and onwards into a range of impacts (sometimes described as 'outcomes'). In doing so, it faces two major methodological challenges, one of analysis methods (see later) and one of data validity and reliability. In turn the latter requires appropriate scores made up on either continuously weighted factors of many items or on a facet basis, each with sufficient data items, in all of OM domains (aspects of physical health, parent quality of life − PQoL, developmental impacts). In a large multi-purpose-study of OM, there is no feasible alternative to such scores being based on reports by parents, i.e. responses to questionnaires. This is practicality, not preference: it is very welcome to have specific reference studies also using objective measures, clinical examination and performance tests as grounding for interpreting questionnaire results. The OM8-30 questionnaire was formed by selecting in a psychometrically rigorous way, 32 items from a pool of over 200 generated for the TARGET RCT of which 74 were used in the 'full TARGET' scoring. The 30 in the name comes from the initial formatting with two questions having two parts. The items were grouped into eight a priori facet sub-sets known to relate to OM (URTI, ear infection score, behaviour, speech/language, sleep disturbance etc.), according to item inter-correlation and subsequent analyses. In some instances, later factor analyses (patterns of item inter-correlation) on the present large data base were used to decide on splitting facets (URTI → infection, obstruction) or to combine them for reliability or need for a more aggregated domain score (behaviour, speech-language, parent quality-of-life → 'general impact'). The scored levels for the response categories of each item are separated by scale values, estimated through regressions predicting the best facet score available, a process which went through at least two iterations (for more details see General Method). The OM8-30 has good criterion validity and consistency (Timmerman et al., 2007), has been adopted in several RCTs (e.g. Williamson et al., 2015) beyond TARGET which gave rise to it. It has also been used to map a generic

Quality of Life measure (Dakin et al., 2010), in the course of which it emerges that the QoL correlates highly with the 1st principal component of all items. Thus the signs and symptoms characteristic of OM are accumulated over a narrow range to provide the homogeneous facet scores, but can also be accumulated over a wider range to give domain totals (e.g. general impact) or a grand total. This multi-level scoring allows use of facet scores for specific validity in large samples, and a PCtotal or mapped QoL, or domain scores of intermediate breadth such as impact. This enables a rational decision about the desired breadth of measure in relation to the sample size and the trade between general reliability and specific validity. Specific validity may not be achievable with small samples. Thus the instrument responds to the fundamental constraint, almost a conflict, in measurement, of which many instrument users are unaware, holding a simplistic concept of the quantity measured and the act of measurement.

1.1.2. Seasonal variation in profile at presentation of cases

In investigating the seasonality of OM presentation, facet or domain scores will have to undergo analysis of the general annual pattern. This could be complex, but according to what the available data can support may have to be simple, just identifying the dominant 52-week periodicity, and capturing the amplitude of seasonality, with phase delay summarised as the smoothed estimate for the annual centre of gravity for maximum severity.

One of the major methodological challenges in OM research is the difficulty of longitudinal studies. Over a life cycle, parents and individual cases are willing to sign up for a research session every few years, or in a trial of a few years only, for a follow-up visit every few years. But there are very few finely sampled longitudinal studies in OM due to research burden on parents; one notable example (Hogan & Moore, 2003) was on an unselected sample so gave detailed OM information on only relatively few and mild cases. We are therefore faced with the need to extract temporal information from a cross-sectional study of OM across the whole year. The challenge here is like the one in ‘bucket biochemistry’. Patterns present in the individual will be present in the mass data in an indirect form, and for these to amount to an interpretable temporal signal rather than noise, the characteristic timing must be strong even if not universal.

The measures require a sufficient number of items in each disease facet and a very large sample to accumulate typical patterns across individual variations in the annual pattern, and particularly across the quasi-random variable of where in the cycle of disease the particular family happens to consult (although we see later that this affords an opportunity as well as challenge). History-taking (anamnesis) in diagnosis attempts to escape excessive dependence on this arbitrary timing, but can escape it only imperfectly, because there is no clear evidence-based guidance on how to do so, and the measures and examinations are made on one particular day. Thus while the diagnostic decision AOM/OME is not particularly difficult to make given some specialist training, it may in the individual be more of a statement about time of consultation than about the individual. This realisation leads to the appreciation of two propositions underlying the present work: (a) that the quasi-random timing of consultation may actually be helpful in research by creating a source of possible covariation to document. In this way autumn virus timing is like a pulsatile stimulus in the coherent averaging used in physiological measurement, and the date of consultation is like a random noise excitation, also deliberately used in some circumstances; we have to assume it will drive not an absolute response but a relative one, the seasonal change in the balance between upstream and downstream facet severity, according to timing in the annual cycle. Secondly, (b) we can conjecture that if (a) leads to firm conclusions with sufficient constraint for modelling, it may then be possible to work backwards from the clinical data to a more precise reconstruction of the history. The confirmation of proposition (b) has to remain a future possibility, because the present data source had only one occasion of consultation and a fixed reporting period before it, for capturing a mixture of severity and recent frequency – the last 3 months – with no attempt to locate previous attacks, timings or durations of symptoms. For the present, the aggregation of cases at secondary care seen across the year must reflect maximum population incidence (with some smoothing and delay) but should also reflect the seasonally changing profile of OM.

The case-incidence, and hence the monthly density of cases throughout the year is not uniform but must follow typical seasonal pattern of URTI incidence. This leads to varying grain in the data available; the number of cases in the late autumn and first 3-4 months of the calendar year will be high, giving large numbers to document the associations between upstream variables and the probably differing phase delays. In

contrast the numbers of summer and early autumn consultations, of whom a higher percentage will on average be late in relation to triggering attacks and so have chiefly 'downstream' problems will be smaller. Initially at least, it will be impossible in cases of presentation late in the annual cycle to distinguish between (a) by chance, the first viral infection being unexpectedly late in the year pure; (b) inertia on the part of the parent or healthcare system; and (c) children having all the risk factors against them for particular conversion between stages hence versions of OM, at the various stages in the canonical OM pathogenetic pathway (Bhutta, 2014) but in a particularly slow-acting form. We cannot see a coherent biological basis for expecting (c) so must prioritise the distinguishing of (a) from (b). Conclusions about variables downstream, as to their absolute timing, their associations, and any strengthened causal inference will face greater problems of statistical power, because the fewer cases that get this far will tend to have extreme severity. Within this limit of the lesser certainty about summer cases, the above argument sets out the conjecture that summer-consulting cases are not just like winter cases happening to occur in summer, but just more rare. Rather they are a differing type of case because the annual cycle for the majority of children starts with the late autumn rise in URTI. This chapter addresses the precise ways in which this conjecture can be shown to be true.

1.1.3. Theoretical basis for expecting delays between aspects of disease and impact, hence differing seasonalities

Despite the challenges summarised, there are grounds for optimism that tests of the above conjecture could be successful at least in part. Using a questionnaire with a sufficient number of items in the relevant population (defined for the purposes of applicability of the same set of questions as OM 3 to 8 years of age) over year may be able to capture seasonality of severity for at least upstream aspects of disease and impact. Preliminary results estimating annual shape with sine and cosine terms (hence potentially delay for each aspect's maximum in relation to the start of the season) produced a coherent sequence of delays although weak downstream seasonality - even for a maximally reliable 15-item developmental score (no longer used -- Filipovic et al., 2013). A fuller statement of the theoretical expectation for a more comprehensive test with fuller data is therefore justified.

In causal sequence, ear infection symptoms are followed by MEE with some of these progressing to OME. The delays to this and subsequent developmental impacts (in those seriously enough affected to have them) have not been precisely measured; that is what the present work sets out to do. However we can state a generally expected range of delay, and it is better to do this than to rationalise obtained delays after seeing the apparent results. In a well-sampled study (Zielhuis, Rach, & van den Broek, 1990) reported that a half-life of 3 months (i.e. halving of point prevalence for cases having B tympanograms at the start in 3 months) gave a good approximation to the spontaneous resolution of MEE. The number of children with long periods of auditory deprivation will be small (and hence material developmental sequelae will be fewer than there will exist children having immediate symptoms and consequences of disease). However with a half-life of 3 months, the general time-scale to be considered must span at least 6 months (decay of index cases from 50% still having fluid at 3 moths down to 25 % prevalence at 6 months) and preferably 9 months (decay to 12.5%) of the cases originally qualifying. This is the exponential function common in biological decay processes: for example 9 months is three times the half-life period of 3 months, and so we could expect consequently HL and ACET to show maxima (high scores) reflecting bad hearing with a peak in late winter or early spring. Hearing questions for parents should follow the overall HL in general but with delay because the effect on the functioning level of the child would aggregate over some months, hence continue into the early summer. Some delay between these two aspects of presentations is inevitable if the time with HL is to be accumulated in to a causally dependent functional deficit or impact as reflected in some degree of parental concern at the deficit. We already have outline evidence of this (the stronger fit of the cosine term in the sine & cosine approximation in Study II). The estimate of such exact delays (whether absolute for one facet or as an interval between estimates for two) should not be taken too concretely or over-precisely, as the estimation process could introduce some difference according to the adjusting variables, cases, and centres contributing to one method. Once general coherence is established, greater precision can become a goal.

The further downstream aspects of OM, which result from the aggregate auditory deficit and physical health symptoms can only sensibly be conceived of as separated by several causal stages. We have no precise basis for estimating the time courses

involved, so the term ‘6 months and/or over’ will have to suffice. The literature does not contain sufficiently precise information for producing estimates of characteristic delays. There is one relatively precise source however for the time constants of the accumulation suggested above. Haggard et al. (in preparation) predicted RHD (the 9-item TARGET version) from HL in longitudinal data on the second of two occasions with a 3-month separation. They found approximately equal coefficients for simultaneous HL and HL 3 months before. The implication of this equality is that the mean timing of the two predictors (6 weeks before) must bisect the period of maximum influence so suggests an average 6 week delay for RHD; in turn this implies an at least 6-week delay to peak severity for any further downstream facets which are chiefly influenced by RHD, or may be directly affected but be similarly subject to cumulative auditory deficit, e.g. speech and language. The difficulty arises with this not being the only path to outcomes. Sleep disturbance can also be viewed as a mid-stream mediator of outcomes, but influenced via physical effects where it is not at all evident that there would be an accumulation process, hence not much delay. However an alternative view of effects on sleep could be that they are heterogeneous – influenced by events summarised by several of the other facets shown and perhaps by others not shown – there could be perhaps 4-6 weeks to the maximum annual severity but a very indistinct peak due to multiple influences. For sleep problems, this is speculative, but for behaviour and quality of life, the problem is certainly compounded by their receiving influences from health as well as hearing through pathways where similar time-constants cannot be assumed. These causal sequences are expressed graphically in the figure in Section 4.1.3.

The downstream impacts, given the two pathways of influence are very likely to have a broader peak, and it may be hard to demonstrate any significant seasonality where this is true because of differing delays in two or more pathways. In other words, the several processes in the causal pathway will inevitably lose the sharp synchronisation, and probably also the homogeneity across cases, that is expected for the physical and biological processes upstream. Factors such as age, length of history and SES are also expected to be important influences in the downstream steps; if they are, and if cases also differ in the relative strengths of the influences from the main two pathways, auditory and physical health, then the resulting timing may become so

different between individuals that it cannot be precisely estimated. Thus we would expect the delay effects to wash out towards this downstream end of the annual cycle and the causal cascade. The feasibility of measuring these facet-specific delays may be limited at both ends of the delay spectrum: downstream via this compounding of sources of variability in long-delayed processes, upstream by the shortness of the intervals involved, relative to the annual cycle, and also because the viable time-sampling of short units containing the quantity of data required to be reliable will be limited by the sample size.

In theory, finer sampling than weekly is possible to address the upstream limit (e.g. daily); but in practice with a finite sample size this refinement cannot work beyond a certain limit. With only 2,000 cases, the expected average number of cases is approximately 40 per week (more in winter, fewer in summer), but per day it would be less than 6 cases - too few for reliability. In the event, with weekly sampling we do find seasonal fluctuations, thanks to sinusoidal smoothing.

1.1.4. Estimation technique

An adequate representation of seasonal phase delay was offered in Study II for all three markers of hearing, when including season among a set of predictors of severity analogous to risk factors. This involved a pair of sine and cosine functions with starting point December and using as discrete time quanta each month of the calendar year. This has two slight disadvantages for precise timing and causal inference: (i) the month-level quantisation provides some pre-averaging for stability but can amplify variability where minor sub-peaks may be close to a month boundary; (ii) in theory, where closely similar magnitudes of sine and cosine coefficients are obtained, this entails that the timing of the centre of the annual peak severity is half way between the sine sensitive peaks in the stored values of the sine and cosine functions. In practice where this ambiguity occurs, the individual terms may not be statistically significant. In contrast, where one is strong and the other weak, the estimation of where the exact maximum falls within the flat top of the function will remain uncertain even if receiving a strong estimate.

It is necessary to avoid naive estimation of annual delays from the shape of the monthly figures using a categorical variable with 11 degrees of freedom (*df*), as this is

drastically unparsimonious and so permits instabilities such as false interactions with other adjusters; in contrast sine plus cosine only costs 2 *df*. For more detailed interpretation of time lags, we have used a series of 26 sinusoids starting at the first week of the calendar year and seeking a maximum fit to the actual annual pattern. This second approach is more precise and direct for extracting a phase delay measure, even for weak seasonalities uncomplicated by estimation difficulties when seasonality is weak, although it takes a lot more work. The advantage of having both sine/cosine pair and 26-sinusoid methods available (which give the same general result) is that a broad descriptive or control purpose can be followed or a more precise narrow analytic one as here. The pattern need not be exactly sinusoidal but the smoothing involved will find the annual centre of gravity (i.e. the group phase delay) for any pattern (e.g. a square wave). This has major advantages over exploratory reaction to visually extracting peaks or beginnings/ends in histograms for month or week which demands use of 11 *df* and risks focussing on chance coincidences and not the mass delay within the gross pattern. Visual inspection of monthly figures is additionally open to interpretative biases in the scorer.

1.2. Parental reporting hearing (RHD) and objective hearing measures (HL and ACET): analyses of shared properties and discrepancies/Study II

1.2.1. Measures of hearing and their precision: do they measure the same thing?

The population most affected by hearing loss is young - through the period of speech and language development, schooling onset and social integration. Of this prevalence and incidence, despite the condition being of limited duration in most instances in developed countries, ear infections are the overwhelmingly predominant cause. This has led to numerous guidance criteria for diagnosis and management, based on the distinction between acute otitis media (AOM), recurrent acute otitis media (RAOM) and otitis media with effusion (OME; AAFP, 2004; Rosenfeld et al., 2013; Bhutta, 2014). This diagnostic distinction serves well in practice, but is not perfect, due to various overlaps in pathogenesis, and to the existence both of some hearing loss following acute episodes in main (R)AOM diagnoses and acute flare-ups within main OME diagnoses. Nevertheless, the cumulative incidence of any attack of OME is over 80%, and the major alignment of hearing loss with this OME form makes OME the most common form of acquired hearing loss in childhood (Zielhuis, Rach, & van den Broek, 1990; Bluestone, 2004).

The scientific literature on the impact of OM upon language, cognition and behaviour has concerned primarily the genuineness or otherwise of an effect from the mild and fluctuating conductive hearing loss in OME as needing to be established in a logically prospective design. Except where there is proactive post-OM referral or some system for screening of hearing, many case identifications (i.e. by convention, meaning referral and diagnosis) result not from symptomatic hearing problems but from a delayed referral, and because of an evident impact on language, social development and schooling (Bennett & Haggard, 1999; Roberts, Rosenfeld, & Zeisel, 2004; Williams & Jacobs, 2009; Hall et al., 2014). From a clinical and retrospective point of view, the causal attribution therefore seems far less in doubt. The hearing loss is sometimes described as characteristic of a latent asymptomatic stage. Other pathways than hearing could be involved for the same sequelae, but these have revived less attention; repeated

ear infections also have cumulative effects on behaviour, attention and literacy (Bennett et al., 2001; Paradise et al., 2000). The retrospective causal attribution chiefly to the episodes of OME is not necessarily incorrect but it (a) may overlook the RAOM component and (b) be prone to over-interpretation, and the overlooking of important co-factors such as poor environment, because only the most severe cases arrive at assessment for of these developmental sequelae.

Many cases in the pre-school child population have fluctuating hearing loss and/or long lasting mild conductive hearing loss (MHL); this could nevertheless pose problems in the playgroup or classroom environment because of background noise and reverberation (Smaldino & Flexer, 2008). Prolonged history of ear disease even if (rarely) present in only one ear, can still affect binaural hearing and speech discrimination even after the hearing loss is resolved (Williams & Jacobs, 2009). Speech and language problems are known to be more common in families with adverse socioeconomic factors (Bennett et al., 2001), non-stimulating environment and prolonged history of the disease (Zhang et al., 2014). Without needing here to address the issue of possible synergies in the underlying mechanism, the co-occurrence of the disease with parallel risk factors for its sequelae builds up a picture of a challenging complex, not sharply defined but serious in a minority of cases, where extremes of disease and circumstantial factors co-occur. This analysis leads us to expect findings to be consistent with the importance of cumulative auditory deprivation for sequelae, but poorly controlled for possible comorbidity effects. As experimental deprivation would be unethical, the comorbidity requires careful control in observational studies; it may be hard to show whether such effects are mainly additive or synergistic for generating developmental sequelae. This complicated picture of natural history, presentation and sequelae (Rovers et al., 2004) poses challenges to conventional medical thinking, particularly in respect of (a) multi-aspect presentation, so a difficulty for standardised assessment methods, and (b) the need for case-finding with such methods that should be targeted at the most severe and persistent of conditions, but within the very common diagnostic label ‘OM’.

The professional response to this complexity has been largely one of retreat into a narrow conception of the disease, and into the technological attractions of hearing level

as a single summary measure. The standardisation of procedure, precision, objectivity, and clear scale definition all make the absolute auditory sensitivity (threshold in decibel hearing level – HL) one obvious choice, particularly for OM that is labelled as OME. A cut-off level of 20-25 dB is conventional for distinguishing educationally significant hearing loss (Canadian Task Force on Preventive Health Care, 2001). However any such single-occasion value fails to capture the importance of duration, given the fluctuations in HL that occur in marginal cases. Regularly repeated audiometry is impractical, however, setting one type of limit to the value of HL. This recognition allocates central importance to any measure that would capture cumulative auditory effect over past months and even years, as effect must be proportional to time of auditory deprivation, but the ability of any measure to do this has not been documented. I do not deny the key role of hearing in presentation and path to sequelae; indeed that key role demands that other aspects and measures of hearing be considered along two dimensions: (1) for simplicity and practicality along stages in the ascertainment path between suspicion, confirmation, referral and full specialist assessment and (2) for completeness of relevant information about hearing beyond absolute sensitivity. As example of (2) we note that a speech-in-noise test is a better predictor of benefit from ventilation tubes to hearing level in a randomised design, even better than baseline hearing level is itself (MRC Multi-centre Otitis Media Study Group, 2004). Unfortunately, it is not realistic either to provide speech-in-noise tests widely at present, nor to expect all suggestions for measures used in assessment to be subject to this crucial and most rigorous test: the differential prediction of benefit in a randomised trial. Assessment is not exclusively about candidature for early treatment: prognosis and counselling are involved. Therefore a coherent body of knowledge is required on which to base detection of disease, measurement of its severity in known main relevant aspects including hearing and further mediators of sequelae. Within this schema, aspects typically attracting parental concern over child hearing and development need to be distinguished from aspects not receiving such concern, but still prognostically relevant. Further publications will address the wider picture but this Study II concentrates on hearing, by considering the determinants and interrelationships of three very feasible measures. Two are already widely available: HL, tympanometry and the third, reported hearing difficulties (RHD) is the systematisation of an idea which is widely

acknowledged if perhaps somewhat variably practised: asking the parents about the evidence for their child's poor or good hearing. The next section covers the unfamiliar form in which tympanometry information is used here, and the justification for considering a defined RHD measure, given the unfavourable reports from one singular application of it. The longer-term objective is to rigorously determine the information value of combinations, e.g. complementing tympanometry with RHD in settings where audiometry is not available to give a more complete view of disease severity and profile, and greater resilience to the challenge of missing data, unavoidable in practice and most research.

1.2.2. Tympanometry as an estimator of threshold (ACET)

Some of the measures in clinical assessment are very sensitive but not specific when considered in screening terms, for example for detecting a given degree of hearing loss (Rosenfeld, Goldsmith, & Madell, 1998). In screening and monitoring applications, this requires complementary information which is specific if perhaps not so sensitive, one reason for considering combining tympanometry with RHD. Particular applications of questionnaires with tympanometry could also include selecting sub-populations of children needing fuller HL testing or totalling with other continuous measures.

Due to the correlation between middle ear pressure and compliance, the physiological measures from measurement of middle ear function do not produce linear mapping into severity within the normal-to-mild range, leading to the use instead of summary categories A, C1, C2 along a bivariate dimension. But continuous measures are more powerful so for this range the MRC Multi-centre Otitis Media Study Group provided a general scaled formula for air conduction threshold estimated from tympanometry first scored under the modified Jerger scheme, calling it 'ACET' (MRC Multi-centre Otitis Media Study Group, 2009; Haggard & MRC Multi-centre Otitis Media Study Group, 2009). As defined, the ACET procedure acknowledges that most clinics do not note down data on compliance and pressure but go straight to the conventional categories, so ACET uses these and does not require handling of continuous compliance and pressure values; because of the non-linear patterns in compliance and pressure, accepting this categorical information as input and the

upgrading to a continuous measure loses little information but avoids mistakes. Applications include combining with other continuous measures, justifiably imputing missing HL data in research, or substituting for HL under specified conditions when audiometry is unavailable or unachievable (Milovanovic et al., accepted). The source data remain middle ear function, but the ACET mapping to the HL scale enables it to also be considered a hearing measure and the interval scale properties thus created make it possible to use more powerful parametric tests. More details of the ACET procedure are given under General Method. Briefly, the resulting distribution of values is markedly bimodal, with a low flat peak, a gap and a second high peak. However for data analysis in research, the wide availability of bootstrapping has downgraded the importance of normality of distribution as prerequisite for the powerful non-parametric statistical tests in general used here. It is the scale properties that count.

1.2.3. Reported hearing difficulties – parental questions about hearing

Questioning parents and caregivers in order to obtain a fuller picture of child wellbeing is not new, but the verbal reports forming the main part of case history in paediatric health examinations are not well standardised. For more standardised questionnaire approaches there is even a specialised stream of work involving proxy response (i.e. another child, parent or caregiver) showing acceptable correspondence between self and proxy measures (Calderon, 2000). The movement for measuring Quality of Life (Fidika, Salewski, & Goldbeck, 2013) and other valued outcomes, rather than being limited to measures of organ function, has produced excellent instruments for research, but the clinical application of the short forms has been limited, because of the limited reliability achievable with a very small number of items. In hearing there are inventories with well-established properties applied in the context of hearing aids, cochlear implants and effects of noise (Dettman et al., 2007; Dutt et al., 2002; Filed & Haggard, 1989) but so far no recommended and widely used short form questionnaire for hearing screening in children with otitis media. Brouwer et al. (2007) studied reliability and validity of specific and generic health questionnaires in children with RAOM and both showed good responsiveness and good psychometric qualities (Study III). Timmerman et al. (2007) reviewed of the available questionnaires for RAOM or OME; none received high rating on all their quality criteria, but OM8-30 (in which

RHD-4 as used here is a component) had the best psychometric properties. Data of the same authors using OM8-30 showed it to have good internal consistency between items in the facets, especially for the 4 RHD items (Timmerman et al., 2008).

Proposals to adopt parental questions for hearing face a potential objection from conclusions of several studies with moderate sample sizes, claiming that questioning does not well identify cases of sensorineural hearing loss in asymptomatic populations. In conductive hearing losses due to middle ear problems, Rosenfeld et al. (1998) showed that parents were not able to accurately rate the hearing in a sample of 186 children with middle ear problem, and the change in objective HL was not well associated with change in parents rating after treatment. A possible contribution to a lowered correlation in their study might be the loss of reliability on differencing to get measures of change. In detecting mainly sensorineural hearing loss, Watkin, Baldwin and Laoide (1990) found low level of parental suspicion concerning hearing. However two considerations argue against prematurely closing down the issue of the usefulness of RHD in OM on the basis of these and other unfavourable reports: (a) the circumstance of a child with an ear condition, known in general to be accompanied by hearing loss, is very different from the circumstance of survey questions to parents of asymptomatic children; (b) in the past era of late ascertainment of SNHL, many families had been given false assurance about this rare condition despite suspicions, so the account from past studies is one-sided. The now available literature is much more diverse and balanced in its conclusions than those studies suggested. Nondahl et al. (1998) in their study of self-reporting hearing loss in adult patients found that a single item, the best general rating one, was good enough for surveys on hearing loss. Moving to middle ear disease and its hearing losses, we see very mixed results. Maw and Tiwari (1988) found that middle ear disease overall is more often discovered by parents than by others, but Lo et al. (2006) assessed parental suspicion in 276 children with OME and found it to be very low, incapable of detecting OME in their children. Heidemann et al. (2013) using the OM-6 questionnaire in 435 children found that caregiver's perception for child hearing was less than accurate. Brody et al. (1999) showed in a sample of 115 children that the H-7 questionnaire plus a global hearing rating in a parent survey about child hearing still missed much of the mild hearing loss. Thus proposals to make

systematic use of hearing reports require substantial data to overcome the unfavourable view from these evaluation studies taken together.

The larger more recent studies in OM(E) caution that proposals for using questionnaires need to be accompanied by clear specifications of criterion for application and of sample, to show optimised measurement, and to consider accompaniment of other measures with specific hypotheses about how the deficiencies of each would be offset by some combination of measures. One study has already illustrated this principle. The results of such comparisons may depend on the particular criterion adopted (e.g. whether it is continuous HL or some cut-off such as ≥ 25 dBHL, given that many OME cases fluctuate in their hearing loss over time), and also on the time interval between measurement of predictor(s) and of criterion. The idea of combining different measures within the domain of hearing may seem novel within a clinical perspective; however, epidemiologically, when relating hearing to its possible antecedents or consequences, combination has obvious attractions such as generality and reliability through aggregation (Study II & Study IV). For example, to overcome error due to fluctuations over time and general unreliability, Hall et al. (2014) provided an example of this pooling of variables in otitis media: they used combined OME &HL in a derived variable to predict IQ (both performance and verbal) in 1,155 children up to age 8 years of life. They obtained the expected significant relation between this derived measure and IQ at 4 and 8 years of age ($p < 0.001$).

In summary, for detecting permanent hearing impairments, the literature gives a strong signal that questions to parents are at best under-sensitive (Rosenfeld et al., 1998; Watkin, Baldwin, & Laoide, 1990; Lo et al., 2006). This has probably discouraged further work but need not be the end of the matter, because: (a) there has been very little exploration of optimal or multiple question items, (b) in most studies there was some lapse of time between answering the question(s) and the time of HL testing (Maw & Tiwari, 1988; Watkin, Baldwin, & Laoide, 1990; Lo et al., 2006; Olusanya, 2001) so part of the poor correlation may be due to fluctuation, and hence more to a deficiency in the criterion measure used than in questioning itself; also (c) there are other goals of hearing assessment than detecting undetected hearing loss in families where it has gone undetected so far. We have also been unable to locate any studies that had pre-notified

parents of particularly revealing situations to consider when observing hearing difficulties for subsequent report. Of course it is necessary to document the role of possibly poor parent awareness of hearing behaviours when using these measures, but the literature by no means rules out a role for reported hearing measures. Indeed, as the marginal hearing losses in OM are subject to fluctuations it is arguable that a HL (or tympanometry) on a single occasion could be misleading about overall hearing status, and especially so about that status through the past several months. It could be valuable to capture the recent hearing history for prognosing or interpreting developmental impact, and also (because this predicts persistence into the future) such information may potentially reflect the need for treatment.

1.3. Correlation analyses of OMQ14 scores with seasonal and other demographic characteristics of children with OM/Study III

OM is an extremely common disease of childhood and the leading reason for antibiotic prescription, for which the overall costs exceed 3 billion dollars per year in USA (O'Brien et al., 2009). These facts point to the need for wider evaluation of the symptoms of the forms of OM, of severity, outcome on the child's and parents' quality of life (QoL), at consultation and to document treatment outcomes. The disease impacts upon children, parents and caregivers in diverse ways, but parents are more able to describe child needs and behaviour than others. Younger children have not enough cognitive maturity or descriptive vocabulary to express and rate their social needs or improvements in these properties and so fullest information depends on responses of parents (Eiser, Mohay & Morse, 2000). Thus the parents as proxy responders could rate problems in their child and the results of their rating may therefore depend on some other factors: age, SES, culture, etc. There are numerous psychometrically developed instruments aiming to rate health related quality of life and disease outcomes but relatively few for children and fewer still for OM; these have differing conceptualisations of disease domains and of aspects of QoL. Good psychometric questionnaires should be based both on good empirical evidence (e.g. large samples and recognised techniques) and should also have a good conceptual underpinning (Davis et al., 2006).

A good OM questionnaire would cover several health-related domains and generic quality of life measures: physical symptoms, child development, educational performance, behaviour, general health status, parent's quality of life. In a review comparing 15 OM-related questionnaires, three of them - OM8-30, OM-22 and OM-6 questionnaires - best reflect functional health status (FHS) (Timmerman, 2007). Using general quality criteria for measurement properties of health related quality of life questionnaires (validity, reliability, responsiveness, floor and ceiling effects, interpretability; Terwee et al., 2007), Timmerman et al. (2007) concluded that the OM8-30 showed best psychometric qualities; this could be expected from development for and using the reasonably large reference sample of the TARGET RCT begun in the UK in the mid-1990s. The orientation and early steps of this development had as aim the

measurement of relevant aspects of Quality of Life in OM (Haggard, Smith. & Nicholls, 2003), but also more specific uses: OM8-30 covered 3 domains, via 9 facets and 32 items: physical health (13 items), reported hearing difficulties (4), developmental impact (15) (see General Method). The OM8-30 is designed especially for children with OME and captures parents' answers for OME-related symptoms and signs. Its 32 items were selected from the initially larger pool for TARGET (over 200) using strict psychometric principles, grouping into facets and weighted based on the degree of positive inter-correlation via factor analyses or principle components. The initial development sample was a group of 441 cases in the TARGET trial including the 376 randomised and a further 65 having post-randomisation hearing levels. The best 32 items were selected for a brief yet effective set, out of a more comprehensive and initial 83. The 83 (necessarily giving more reliable totals) were used by Multi Centre Otitis Media study Group (MRC) to provide first set of outcome measures of treatment in the TARGET RCT (Haggard, Spencer & Gregori, 2007; Haggard, Smith & Nicholls, 2003). Validations of the facet scores have accrued at various stages. As an initial example, the behaviour problem domains in OM8-30 correlated well with the scales of the Strengths and Difficulties Questionnaires (SDQ) assessing psychosocial adjustment in children (Timmerman et al., 2008). The facet scores of OM8-30 with and also without HL and domain scores with age and sex produce accurate predictions (mapping) of child Quality of Life (QoL) on the scale of disutility for the HUI3 instrument (Health Utility Index Mark 3 instrument). The domain scores of OM8-30 also well predict disutility for the HUI2 (Health Utility Index Mark 2 Instrument) instrument (Dakin et al., 2010) in the GNOME trial sample. Both HUI versions (Mark 2 and Mark3) are multi-attribute health status classification systems for estimating health related quality of life (HRQOL). This and other types of utility score system for health-related QOL of patients are widely used for estimating cost utility as a quantitative extension to cost-effectiveness analyses (Drummond et al., 1997; Horsman et al., 2003). The psychometric characteristics of OM8-30 in the Eurotitis 2 database were consistent with findings in the TARGET database, with good content, criterion and construct validity and internal consistency (reliability -- Timmerman, 2008). OM8-30 (or its short form OMQ-14) has been adopted as a suitable instrument for estimating OM impact and follow up after treatment in several clinical trials. Of these, two trials (of nasal steroids and auto-inflation) in

general practice in UK are now published (Williamson et al., 2009; Williamson et al., 2015). The Eurotitis-2 study group holds these databases to these two studies having a generic QoL measure, so is in a position to do further development of optimal scoring for predictive mapping of QoL.

The item selection for the OMQ-14 short form used in the second trial (AIRS) was largely influenced by the item prediction of QoL in the first (GNOME), because its main purpose is QoL estimation in children with OM. The selection of its 14 items followed two criteria:

a) The chosen questions considered individually are those that best predict child QoL (CQoL) in TARGET and HUI3Qol in GNOME, i.e. the same data at item level that were summarised at the facet level to produce the OM8-30 mapping (Dakin et al., 2010). The overlap in these two item subsets qualifying as predictive in the respective data sources was 12 out of 14 (86% agreement) so the next most QoL-predictive item was also chosen from the ordering, one from each data source. From the RHD facet of 4 items in OM8-30, three items are kept in the short form (overall hearing rating, mishearing things, and hearing in a group) and only the item of 'asking to repeat things' was eliminated. After this reduction of items from 32 to 14, the 14 items were weighted into three factor scores by Varimax Factor analysis. These were readily interpreted and named as: physical symptoms [general health impact (1) plus, ear symptoms (3) = 4], parent-reported hearing difficulties, RHD (3), and general impact (7), This last is a domain formed with high-loading items drawn from the three OM8-30 facets [behaviour (3), speech and language (2) and parent QoL (2)].
b) The factor scores were then taken as preliminary sub-totals and used as dependent variables in the stage of item scaling. Here the best scale vales for the response level of each item are established by predicting the factor score on which the item loads highest. This process increases factor loadings (and hence Cronbach's alpha). The question formulation was kept the same so the short forms scoring should reflect profile the same as the previous (i.e. OM8-30 form). On the basis that this connection with OM8-30 should also make a good instrument for follow-up to document treatment effects, the OMQ14 is now in use in other studies. In the

second RCT in UK general practice (auto-inflation in OME) the OMQ14 showed good benefit after treatment (Williamson et al., 2015). This 'responsiveness' (i.e. to treatment or other type of change) is sometimes considered one form of validity.

The OMQ14 scoring serves two aims, against which discussion of validity has to be considered: i) summarising general severity scaled as QoL and ii) providing a profile of disease in 3 independent dimensions.

i) For general severity, criterion validity needs to be established. The criterion measure is offered by the more reliable longer form, OM8-30. Criterion validity is simply and sufficiently estimated using the correlation of the PC total for OMQ14 with the corresponding total for a larger form of questionnaire (i.e. the OM8-30). OM8-30 is already accepted and has been confirmed to have good construct validity and reliability (Haggard, Smith & Nicholls, 2003; Timmerman, 2008).
ii) Assessing the validity of the profile of disease offered by the three OMQ-14 factor scores is more complex and requires some discussion of how the scores are derived. Some general construct validity for items and facets is present in the back ground as already discussed for OM8-30 (i.e. the correlation with SDQ and HUI3 (Timmerman, 2008). However the OMQ14 item pool was selected on a generic basis (Dakin et al., 2010) (i.e. predicting QOL) so there is an extra stage of validation required for the profile information, in preliminary demonstrations and in the accumulation of validity information with early applications of the instrument.

Because there would be largely generic information in the total of 14 items (as provided by the principal component (PC), or by the QoL mapping -- they are quite similar), it was decided to go for relative and maximally independent information in the profile. In this study is adopt the general word 'facet' also to describe type of score based on a discrete set of items as in the scoring for OM8-30s, and 'factor' to describe a near-equivalent but with some weighting from each other item in the defined set, here the 14 OMQ-14 items. Factor scores have two especially useful properties: a) all factor methods produce better distribution than facet scores that would use only the small subset of items loading highly on each factor. The use of some contribution from every item may or may not add useful information hence statistical power, but via the

continuous distribution achieved, it certainly enables using the more powerful (i.e. parametric) statistical tests with their great potential to control for confounding; b) the orthogonal rotation of axes (Varimax; SPSS Version 22.0) enforces zero correlation among the factor scores which would otherwise be expected to be somewhat positively correlated (e.g. the worst cases of disease would be expected typically also to be the worst cases of impact). This general severity is already present in the PC total so the derivation is free to formulate the profile in whatever way may be most useful. Using Varimax-based scoring makes it more possible to show specific and distinct patterns of influence, from determinants, or conceivably upon other variables in the health or psycho-social realm. Otherwise, a finding with one factor score is not independent from a finding with another. The reduction in number of items that load highly on profile scores does not affect the scoring of total severity, because the total is in effect a weighted sum of the factors, hence of all 14 items. The Varimax rotation of the 14 items has high precision, reliability and generality because done on a very large sample (N = 2,865).

Validity information for OMQ14 scores takes several forms: (i) ecological validity is assured by the progressive sampling of disease and impact domains through earlier stages of TARGET and OM8-30; (ii) metrical validity is assured by the PC and factor derivations and by item scaling; (iii) criterion validity is available by correlation with corresponding scores from OM8-30 (further details in Method); (iv) construct validity is expected to accrue over time via strong effects for expected patterns of association with determinants or consequences, although to some extent it emerges from the strength of the factor analysis (as the % variance explained), and also factor interpretability; and (v) pragmatic validity and face validity arrive via satisfactory use in specific clinical applications. Types (iii) and (iv) are the subject of this chapter and at time of writing, core Eurotitis-2 collaborators have provisionally agreed to the Eurotitis-3 study which would pursue validity of type (v) as the ultimate aim of this development work.

1.4. Clinical prediction of HL from RHD questions and tympanometry with applications to case referral criteria and hearing screening in OM/Study IV

1.4.1. General epidemiological argument that proactive case-finding is appropriate in OM

Much literature emphasises the importance of adequate hearing input in early life for auditory cortex maturation. The studies are diverse in number of cases, methodology and variables chosen to be of main interest. In recent years one main strand of well-conducted studies has focused on animal models to document the presence of structural and functional differences in auditory subcortical and cortical centres according to experimentally controlled auditory deprivation. For ethical reasons, the strand involving human participants has been observational, although there are a few short term laboratory studies modifying the auditory input. The human studies are harder to interpret because of the inevitable problem of establishing a high degree of control by using only natural variations and statistical adjustments or longitudinal case-control studies. The consensus has been that there is some justification for a preventive approach to all early hearing loss, although conclusive evidence for a large degree of impact (e.g. cognitive, linguistic, behavioural adjustment) in mild and fluctuating loss has not been forthcoming. The consensus has been held together by two types of argument: (1) analogy from sensorineural hearing loss, mostly more severe and with advantage to linguistic and cognitive outcomes with early versus late cochlear implantation in more serious cases (Kral & Sharma, 2012); (2) interpolation with assumed causal linkage between animal studies documenting profound anatomical disorganisation in auditory deprivation and the results of more uncertain and variable human studies. The anatomical findings go back almost four decades. Lack of normal sensory input in children causes morphological changes in auditory brainstem (Folsom, Weber & Thompson, 1983). The absolute and inter-wave latencies were significantly prolonged in children with early recurrent middle ear disease than in children with no history of middle ear pathology. Two more recent findings on the neurophysiological basis for concerns help to justify the conclusion of a recent review (Whitton & Polley, 2011) concluding that early disruptions to auditory input can be harmful. Authors

suggest that sensory signal degradation and distortion during early life otitis media influence cortical brain development and central processing deficit called ‘amblyaudia’. Brief hearing deprivation in early childhood influences maturation of the inhibitory synapsis important in sharpening excitatory receptor fields and stimulus selectivity (Sanes & Kotak, 2011). Conductive hearing loss (CHL) decreases amplitude of spontaneous inhibitory synaptic currents, changes auditory cortex inhibitory transmission what exhibits further worsening of auditory function. Lack of normal sensory input in children causes morphological changes in auditory brainstem (Folsom, Weber & Thompson, 1983) and compromises the neural plasticity important in obtaining complex auditory skills (Myers, Ray & Kulesza, 2012). Impacts of transient hearing deficit on language and cognition are hard to estimate, partly because OM is an early childhood disease, transient in character with lack of evidence of cumulative effects on child quality of life.

In a recent review synthesising the two strands of work in early auditory deprivation specific to OM, Whitton and Polley (Whitton & Polley, 2011; Sanes & Kotak, 2011) concluded that the often expressed scepticism about cognitive or behavioural sequelae of OM is not justified: ‘Overall, physiological and perceptual testing in animal models as well as humans suggests that the connection between OM-associated degradation of afferent signal quality and subsequent neurological impairment is substantially clearer than generally believed’. However they do repeat the classical point that the evidence on degraded input is overwhelming, while that on otitis media is more marginal and has to be qualified. The difference is of course due to the simple fact that in many cases receiving justifiable diagnosis of otitis media at some point close to the outcome measurement, the hearing loss is not necessarily persistent or even ‘moderate’ (e.g. 40-65 dB) by the standards of the full range of audiometric values seen clinically for sensorineural hearing loss. The hearing loss is variable, and in mild cases often fluctuating. Various findings showed that not only sensorineural hearing loss (SNHL), but CHL in critical time of child neural development influenced behavioural, cognitive and expressive language development in children with OM (Bennett & Haggard, 1999; Myers, Ray & Kulesza, 2012; Sanes & Kotak, 2011). Children with prolonged MEE had significantly worse expressive language performance in comparison with short lasting MEE (Zielhuis, Rach & van den Broek, 1989). Despite

the fact that the pathological substrate is at the periphery of the auditory pathway, the effects on hearing are largely central, causing complex disruptive influences on central synapses and lack of auditory inhibition.

A full review of the issue about early sensory input and sequelae in OM is beyond the present scope but the next section illustrates its diversity. The recent review by Whitton and Polley (2011) leaves the argument standing sufficiently strongly that there could be some value in studies of various aspects of case-finding or screening: best method, best timing, appropriate outcome measures implementation details and cost issues relevant to preventive healthcare (i.e. screening). The issue of healthcare gain from any screen is complex because of the two stages involved: (a) the need to show that the screen cost-effectively identifies relevant cases, and then (b) the need in a controlled trial to demonstrate that sufficient improvement in outcome occurs (across some specified population) from treating specifiable cases from among those identified. Historically the rate of introducing new screens slowed down after the 1980s because it was hard to obtain 2-stage evidence (it has to be built up in stages), and existing evidence was much less favourable than had been supposed. The reduced enthusiasm for mass screening in the late 20th century resulted from economic analyses of the whole system considering the benefit from treatment against the cost-per-case amortised over the entire screened population, not just against treatment costs over cases having come through screening as likely to benefit from available treatments (Morrison, 1998). In managerial terms, this implies considering the budgets of the treatment service and the case-finding service together as one. For appropriate use of scarce resources this cautious perspective has also to be shared with hospital diagnostic service and primary care or community preventive services. In the context of OM screening specifically, to assist sharing of this sober perspective, Haggard and Hughes (1988) introduced the simple idea that proposals for a screen should consider not just the sensitivity and specificity parameters of screening technologies (and related expressions of these) but the 'yield' – how many cases appropriate for treatment the screen actually finds.

For specific issues in potential screens for otitis media and associated hearing, as for the sequelae issue, a lack of evidence specific to otitis media demands that arguments by analogy and interpolation still have to be used, so as to integrate the best

evidence there is. For brevity here I use the term ‘screening’ in a wider sense than a universal screen by test; some of the obtained knowledge on case-finding could be equally applicable in at-risk screening or in the decision to refer from primary care (e.g. generalist paediatrician) for an audiological and/or specialised ENT assessment. This flexibility requires that the knowledge take two related forms, mostly proceeding from the first to the second: (i) a more basic knowledge about predictability of the criterion variable on various sub-populations with various possible screening tests and (ii) the simulation or implementation and evaluation of screens, considering the numbers and types of cases that would be referred, and the accuracy with which this can be done. In this chapter I follow the distinction (i) – (ii) by first documenting on the Eurotitis-2 database the predictability of HL, through the entire relevant range, roughly 10 – 50 dB with general linear models. I then examine within a screen simulation, dichotomising the predicted HL, whether the properties of those models are shown with the same relative importance or at all for optimum identification of cases having a certain degree of hearing loss. Knowledge about screening method concerns what low-cost robust form of assessment suitable for community or office setting can well predict a diagnostic measure that is likely to be available to referred cases at secondary care (hospital) and is also relevant to the need for or ability to benefit from treatment in many such referred cases. In otitis media with effusion there remains a debate, which has never been fully and properly held, about whether the appropriate treatment-relevant dimension really is the measured HL on one occasion (or perhaps 2-3), a debate to which the work on OM8-30 and OMQ-14 in Study II contributes. However the convention to base almost all discussion of OM sequelae on hearing impairment has been so dominant for 7 decades that any discussion of alternatives can only be held in terms of their implications for HL. I have therefore accepted that HL must, when proposing alternative arrangements for evaluation, occupy the role of the diagnostic criterion variable.

1.4.2. Illustrative literature underpinning the case for considering active case-finding or screening

Early identification of sensorineural hearing loss (SNHL) accompanied by early aiding and cochlear implantation where necessary has shown benefits for speech and language development and educational achievement (Geers & Nicholas, 2013).

Newborn hearing screening programmes with this condition as their target mostly cover upward of 95 % of children with SNHL in countries which have implemented such screening Cases ascertained by the age of 3 years give a prevalence of SNHL at 1.07 per 1000 and at age 9-12 years 2.05 per 1000 children. Estimates of this have remained remarkably stable over years and in many countries, late onset or mild SNHL is still a challenge for service provision because of the low yield of cases. This requires creative thinking about further purposes (and so cost-savings) that could also be served by any screening programme or proactive programme that might meet this need; hearing loss associated with OM is an obvious candidate (Haggard & Hughes, 1991). Concrete proposals are therefore in order for systems that would deliver preschool identification of SNHL in those unidentified neonatally or progressing postnatally, and also of those children with transient conductive hearing loss, and do so without strong competition between objectives.

Prevalence of CHL due to OM is very high 20-35% in children 2-3 years of age (Zielhuis, Rach & van den Broek, 1990; Ho, Daly, Hunter & Davey, 2002) and the highest in the most sensitive time for speech development, first and second year of life. Whilst the severity and prognosis may not be as poor as for SNHL, the prevalence/incidence of OM entails that the case yield is not a problem. A common framework could serve both aims, testing for conductive hearing loss (CHL) within a screen or other proactive system cannot be exactly the same as for SNHL for several reasons: i) transient character of the disease and high recurrence, ii) quick device OAE screeners are calibrated for pass and refer criteria on the basis of three of five or two of three frequencies of which three dominant are at the range between 3-5 kHz, iii) automatic ABR screeners with a pass/refer cut-off at 35 dB, are insufficient for mild SNHL that would be sought at around 3 years and in children with CHL, iv) the dB HL equivalent loss for a given level of OAE response will differ between SNHL and the CHL of OM. The assessments relevant to consequent development need to be multidimensional in character: quantitative HL estimation but also reported history data, number of recurrences, tympanometry category (coded as ACET if possible) and duration of MEE after any AOM episode, and conducted within a clinical algorithm reconciling comprehensiveness with efficiency and precision. Numerically, the much larger challenge is recognition of those children with transient, CHL caused by OM. In a

combined approach identification of children with OME, RAOM or OME with supra-added RAOM, the case selection would need to have partly shared and partly distinct elements in the diagnostic path, following up after the first one or two general screening stages. In this context we need to explore just how much and at what stage asking about child's hearing could play a role in a combined proactive approach. The fact that concern about hearing is the chief alarm for further hearing examination (ASHA, 1997) suggests that there are natural patterns in parenting that can be built upon.

Evidence of hearing deprivation estimated via pure or sweep tone thresholds is measured at one point in time and even when normally does not reflect past history. We can therefore envisage that a combined approach screening system would have a second HL measurement, with a degree of watchful waiting, within its process flowchart. Multifactorial causes of disease, complexity in causal relation of risk factors, effect of season and socioeconomic factors on disease severity all underline the likely necessity for assessment after the first outreach screening test to be complex, even before the time aspect is incorporated. Proposals need to be detailed and realistic, and to undergo rigorous evaluation.

1.4.3. Best methods and results achieved with previous approaches to screening in OM

Searching Cochrane Controlled Trials Register with the search terms; 'otitis media' with 'hearing' with 'tympanometry' with 'impact', it yielded only one controlled trial in the area highlighted one of 2 articles and Google scholar and EMBASE 22 out of 6,680. Of these 23 articles 9 were closely related to the topic of OME screening; the remaining 14 focusing on prevalence and incidence, and also outcomes and impacts. We do not yet have a single large well-conducted randomised trial showing good justification (i.e. of the two-stage evaluation for introducing OM screening). In the work close to this possibility that has been done, it is accepted that the middle-stage criterion variable used as 'gold standard' for hearing assessment is tonal audiometry, and the initial screening test is tympanometry. The most often-used outcome measure is speech and language development (Butler et al., 2003; Rovers et al., 2000; Paradise et al., 2003). Traditionally the argument has been led by educationists and speech/language

pathologists and centred on language competence and performance, but modern studies of the presentation of OME including that in Study III (OMQ14) suggest that this concentration is too limited. Most studies have found that language performance depends on hearing level and that with improving hearing the language comprehension also did improve, but that the effect of available treatments was not major: there were no significant differences between treated and watchful waited participants in hearing and language performance. The cardinal common finding in these studies is that language performance depends on hearing status and that duration of MEE significantly affects expressive language performance (Zielhuis, Rach & van den Broek, 1989; Rovers et al., 2000). Thus intervention studies agree with correlational studies (Shriberg, Friel-Patti, Flipsen, & Brown, 2000) that it is indeed earlier hearing that plays a crucial precursor role for language development, and the conjunction of experimental control in the first type of study and time-sequence in the second justifies strong causal inference. Lack of improvement in language outcome in a treated (VT) group compared with a watchful waiting group, gave only low cost- effectiveness (Hartman et al., 2001) in randomized trials and in turn this implied little justification for OME screening in the asymptomatic child population. It then seems a paradox to find a lack of language improvement in the treated groups compared to WW, despite a sharp and immediate hearing improvement. However we should note the low determinacy of HL shown in Study II. It was less predictable by the available clinical information, and the present paradox is consistent with that finding. There are probably several ways that this paradox could be resolved, of which three (not mutually exclusive) involve time and appropriateness of outcome measure: (a) it might involve many months of good hearing to make up for a speech/language or other developmental deficit, or such deficits might be only partly reversible (strong form of critical period hypothesis); (b) the language measures may just not be sensitive to the form of underlying improvement that is influenced by recent auditory experience; (c) there may be more than one mediating variable, as suggested by the conceptual schema introduced for Study II. A review of all the possible outcome measures and their mediators in OME is beyond the present scope. However, mediation by other markers of OM severity than HL could be a contributory reason for this paradox of HL improvement not working through to improvement in language. Surprisingly, the baseline hearing level is not a good predictor of the amount

of improvement in HL with ventilation tubes (MRC Multi-centre Otitis Media Study Group, 2012). However performance on a speech in noise test (SiN) appears to be a good predictor of VT benefit to HL in children with OME (MRC Multi-centre Otitis Media Study Group, 2004). SiN tests attempt to measure understanding of speech in the noisy backgrounds often present in the everyday life of children in school and day care so may tap more directly into the underlying need for treatment that is a past history of persistent auditory problems better than HL does. So hearing level may not be the best marker of need for treatment before, nor the best mediating predictor for extent of benefit after treatment. Finally there is an elementary consideration in measurement within the paradox. HL values after treatment with VTs are homogenously low, so the underlying variation in improvement, even after considering the large contribution to it from the baseline value, may be small relative to the measurement error.

Parents concern over their child's hearing is certainly an important alerting signal for further hearing testing. Some other groups of articles focus on parents concern of child hearing using different forms of questionnaires in OM screening and comparing the scores with gold standard, audiometry giving the hearing level (HL). The common feature of the most of these studies is the conclusion that parents cannot estimate correctly child's hearing (Hammond et al., 1997; Rosenfeld, Goldsmith & Madell, 1998; Stewart et al., 1999). Hammond et al. (1997) used questionnaires from the Child and Youth Health preschool screening program in Australia with 10 hearing questions having dichotomous answers. Despite the high cut off level, 30 dB, they found sensitivity only 60%. All these studies used the dichotomous type of answers and gave similar sensitivity/specificity ratio with low PPV (positive predictive value) and NPV (negative predictive value) (Li, Driscoll, & Culbert, 2009). Relatively little can be concluded from this literature, as results cannot be aggregated because of seven major inhomogeneities in the group of relevant studies done: i) the type of questions used for parents' estimation of child's hearing; ii) number of questions iii) type of answers; v) method of item scoring; vi) sample size and vii) approach to the study. The main confounders in interpretation of results were not all observed and controlled: length of history of the disease, educational level of mother, age, diagnoses and the season (Study I). The good common feature of many of the studies is that the starting point for identification of OM children is about the degree of parents' concern; having a

compulsory question in every child's examination would build on the fact that the mothers are the people who most often first suspected mild and severe hearing loss in child (Maw & Tiwari, 1988). Questionnaires would be cheaper as first screen test, but no precise evaluation study has been done on whether they would give good enough effectiveness. One of the first explorations of a questionnaire screening approach to OME was by Hind et al (Hind et al., 1999). Although their point of concentration was different from the one here, their results suggested that there could be effectiveness in answering also questions on hearing. For sensitivity and specificity of questionnaires at a desirable cut-off point much attention is required to item content and scaled quantitative scoring. The cost per child using questionnaires was estimated 6 x lower than using pure tone tests (Davis et al., 1997).

1.4.4. Type of hearing measures for proposed case-finding systems in the light of their apparent place in the pathological cascade between OM acute and chronic forms; Lessons from earlier chapters

Audiometry giving an average hearing threshold at four speech frequencies is the gold standard for evaluating any other form of hearing test. This measure of peripheral hearing threshold reflects ear status at one particular moment or short time interval and does not guarantee any information about its duration, number of episodes with high HL, communication performance, auditory memory, localisation, discrimination, language vocabulary. Thus we have one measure of hearing at a single patient visit. In order to have a deeper look at the disease influence on overall hearing we need other measures to express effect on the ear and auditory pathways and to reflect duration. One of these measures uses tympanometry to estimate hearing level (ACET -- MRC Multi-centre Otitis Media Study Group, 2009), described under General Method. Even the hearing level is within normal limits (≤ 20 dB) the tympanogram can be flat, suggesting that fluctuation of hearing over time is involved in such a case. In Study II we learned that tympanometry scaled as ACET could be considered a useful hearing measure and also that RHD reflected duration of auditory deprivation. From the models of the type developed for exploring determinants of the hearing measures in OM we can see a need for two distinct groups of measures: peripheral hearing evaluation (HL and ACET) and functional hearing performance (questionnaire). Combining these two measures into one

approach with careful optimisation towards a specific screen role could raise both sensitivity and specificity, and so maximise the combination of these two parameters in practice. Choosing the right items, enough number of items, precise item scaling and weighting base done in large samples are necessary, before useful clinical application can result (Bennett & Haggard, 1999).

1.4.5. Clinical and public health application; model proposal

The best approach for identification of OM children should be that which offers best achieved screening effectiveness and efficiency. Subsequent realistic cost–effectiveness evaluation is then necessary to decide whether a particular screen is an appropriate use of resources – usually public sector resources.

Using the audiometry sweep test in preschool hearing screening gave poor cost-effectiveness for two reasons: the sweep test did not show precise results in background school noise conditions and the cost of the test was estimated at 3-4 £(492 RSD) per child tested with a much higher cost per child effectively treated (Davis et al., 1996). OAE and AABR hearing screening cost 35$ (£20 or 27.59 EUR or 3,728 RSD) per child and complete case-finding of a hearing-impaired child $35.00 (£22.056 or 29.953 EUR or 3,727.97 RSD -- Olusanya, 2001). Quick OAE screening devices widely used for SNH screening have a pass criteria more fixed for higher frequencies so the pass criterion does not mean middle ear pathology is absent. The agreement (Kappa index) between tympanometry and TEOAE findings was more moderate than substantial (Ho et al., 2002) and most often TEOAE was present in the 6-36 months age group, even if tympanometry was failed. The suggested OAE approach is expensive, not currently available for wide coverage in public services, so needs further attention in research studies. The long test time, necessity of trained staff, and cost of the equipment for AABR make large demands that cannot reasonably be met in most countries, i.e. in those with low annual per capita income. Questionnaires are imperfect but may offer acceptable cost-effectiveness at only £0.50p/child tested (82 RSD).

If we focus on identifying risk groups of children at primary level and select those who need further ENT assessment, this promise of cost-effectiveness might be kept. I suggest that in the light of the above problem analysis, we use a questionnaire for pre-

screening and immediate tympanometry as the second half of this first stage, to select possible OME children who need a second audiometric stage, with an element of watchful weighting built in via the delay to hearing estimated threshold and other test at secondary and tertiary institutions. The right choice of questionnaire items will depend on previous development in a large sample, selecting the best items while still retaining enough items for reliability and a quantitative scoring system. The RHD questions from the OM8-30 questionnaire potentially offer these features, and two large relevant databases are available on which to follow this screen proposal further. It is based on a perhaps limited but certainly clear and an accepted criterion of best explaining variation in hearing level using multivariable regression. In examining such a proposal expressed as a model, all analyses should be adjusted for age, season, maternal education, and length of history where significant. They should also be adjusted for centre because a public health screen in one city or province would be concerned only with the relation between its catchment population and the referral centre. It would not have to span multiple such relations as seen in the Eurotitis-2 data. Therefore adjustment best approximates the circumstances that would be met in setting up a screen or other case-finding system in one city or province.

2.0. Objectives and research questions

2.1. Objectives of Study I

2.1.1. Establishing peak presenting times for annual score severity maximum of URTI and ESS, and their relation to the corresponding time delay to maximum severity for hearing measures (HL, ACET and RHD).

2.1.2. Estimating item seasonalities for individual items within the RHD facet.

2.1.3. Explaining phase delay in upstream and downstream disease aspects in relation to other knowledge and determining whether downstream impact measures have longer phase delay and reduced seasonality effect size.

2.2. Objectives of Study II

Bearing in mind the need in OME for reference to an enduring condition, and the potential advantages of combined measures for clinical practice, as well as more fundamental measurement issues, I have systematically analysed three types of hearing measure obtained on a large set of data, focussing on three main research questions. These bear on the issue of a discrepancy between reported hearing and objective measures:

2.2.1. Similarities and differences in determinants of the three hearing measures. Specifying in a highly controlled way the determinants of the variance in objective (HL and ACET) and in a reported (subjective) hearing measure (RHD); b) using effect size as a measure of influence, comparing magnitudes of various influences between the measures and explaining any large differences of these influences found between measures as an approach to the basis of any discrepancy.

2.2.2. Inter-relations of measures and potential for totalling or substitution. Examining the degree and form of the inter-correlation between the three measures and its potential to support two ways of making better use of clinical and research data: (a) totalling for aggregate reliability (PCtotal) and (b) substitution for missing data (ACET for missing HL). The form of inter-correlation is expressed in a mediation analysis,

relatively new to otolaryngology, where the question is re-phrased mathematically to allow strong inference about causal sequence.

2.2.3. Designs able to make direct inference about discrepancies between RHD and objective measures. Examining the determinant model for RHD when additionally one, both or neither of the objective variables is fitted alongside the determinants: (a) the altered effect of those determinants (b) ways of reflecting determinants of the difference or discrepancy between RHD and objective measures and hence its nature; also examining the discrepancy seen as a difference between standardised scores for RHD and HL.

2.3. Objectives of Study III

2.3.1. Do these two properties of the factor-based scoring in OMQ14 (all items in each factor, and orthogonal factor rotation) create a problem, limiting criterion validity?

2.3.2. How do these OMQ-14 factor scores relate to, or break down by, other variables in the data base (determinant analyses, construct validity and need for adjustment to avoid confounding)?

2.3.3. How do these scores relate to hearing measures (a special major issue) after control for the main determinants covered in question 2.3.2.?

2.4. Objectives of Study IV

Three questions were posed for this study all answered by GLMs or logistic regressions with HL as dependent variable

2.4.1. What are the roles of a) ACET, b) hearing rating (HR) and c) 3 other RHD questions in predicting HL?

2.4.2. What are the roles of interactions; a) ACET*overall hearing rating question and b) ACET*3RHD items in HL prediction?

2.4.3. Two questions answered by logistics for various HL cut-offs simulating a screen: difference between 20 and 25 dB hearing screening?

3.0. General Method

Many of the methods are in common between the four studies so are stated as general below, leaving those details specific to each study to their special sections later, but some are cross-referenced in this chapter. The seven sections here are: a) Eurotitis-2 study and database; b) ACET coding of tympanometry; c) Scaling and imputation issues for Hearing Level d) Item scaling for precision scoring of questionnaire items; e) Non-questionnaire clinical data available; f) Centres used in the present analyses; g) Statistical analysis strategy. Section 3.0.4. (item scaling) provides the justifying basis for the precision scoring used in this thesis and in so doing presents new data also from the Eurotitis-2 study, but although used, this method is not itself one of the topics of the thesis.

3.0.1. Eurotitis-2 Database structural properties and summary of OM8-30 item content

A summary account of the history, administration and structure of the study is given in the 1st Appendix to the first Eurotitis-2 publication (Milovanovic et al., accepted), on interrelations and determinants of the hearing measures. That account was simplified by using only OM8-30 data, and hence only data from the centres active in Phase 1, chosen because OM8-30 possesses one extra question on reported hearing difficulties, a valuable addition for the research aim. This section gives an up-to-date account of the database available with the larger dataset from all centres.

The thesis as a whole does not discuss in much detail the differences between centres (in effect, international differences) because for most present purposes, the data have simply been pooled across the centres and no claims are made about which centres show more or less of the effects examined. In some places the possibility is considered of an analysis not adjusting for centre differences, as an exploratory tool to understand how the role of other factors might be due in part to the centre differences; however most analyses have been adjusted on the basis that adjusting out the centre differences at the mean gives a sharper general picture. This is one special instance of the central principle in the statistical analysis strategy (Section 3.0.7.) that, in general, analyses adjusted for known influences (and hence for possible confounders) are better from the

point of view both of statistical control, and usually also better for power if the adjusted effects are large.

Awareness needs to be exercised over collinearity generality – an effect of interest being exaggerated or more often diluted by the presence of a correlated independent variable in the analysis; in such circumstances, making a correct interpretation demands that more than one analysis be done. This is the reason for having also conducted certain analyses without centre adjustment as check, but mostly not reporting them. Centre differences, and certain influences carried at centre level will be examined in a later publication using the appropriate technique for the latter – multi-level modelling (MLM). Outline specific-centre data are given in Section 3.0.6. below. It is sufficient to state that the problem is removed for present purposes by fitting centre adjustments as overall (main) effects with M-1 degrees of freedom where M is the number of centres in the analysis.

The items and OM8-30 questionnaire pre-date Eurotitis-2 so will be described first. They originated from a first round of conventional psychometrics on the item pool of the TARGET randomised surgical trial (MRC Multi-centre Otitis Media Study Group, 2012) done by Haggard, Smith and Nicholls S Smith and E Nicholls (2003) shortly after TARGET data were all gathered and checked in 2002. Briefly, facets of OM were defined *a priori* on the basis of coherent item content (face validity) and submitted to successive factor analyses then principal component analysis of defined item subsets, to justify retention of best items on the basis of coherence, a mixture of reliability and construct validity. The properties of the pool and the absence of any opportunity for further large scale data acquisition in a second stage of instrument development meant that not all constructs of importance were well sampled with multiple items in the data available. Where the number of items is low, the reliability of that measurement obviously suffers; an example is speech and language, where degree of validity may also be in question. This arises because there was some difficulty in developing the measure with enough coherent items, and so some items of marginal quality had to be retained. Nevertheless, a review of available measures in the last decade (Timmerman et al., 2007) concluded that OM8-30 was the best available. On the basis of many analyses including the present ones, it can be said that some of its facet scales (like speech and

language) have to be used with awareness of limitations – in this instance that the items are heterogeneous and a hence too few in a 3-item facet score to achieve reliability. Others e.g. the 6-item URTI score seem to have excellent properties. This needs to be borne in mind in interpreting results, particularly marginal or null ones, which may be due to the inadequate sampling of a facet by a small number of items even if those items contribute usefully to a total. Some studies using questionnaires in ENT do not show this awareness of the limitations of sub-scores (Khalfa et al., 2002). The difficulty arises from incorporating two aims which stand in some degree of conflict: of having an overall OM impact score, and of offering a detailed disease profile. The OM8-30 has been successfully mapped (Dakin et al., 2010) to generic quality of life (QoL), a stage undertaken, because there is no sub-score in OM8-30 on generic quality of life for the child (though there is one for the parent). On the same data in the mapping sample, the resulting QoL predictor correlates very highly with the first principle component of variation (PC total), showing that, for the purpose of a total score, the limitation of having few and heterogeneous items on some facets is not a serious limitation. For the aim of defining a clinical assessment profile, it can be a genuine limitation: some facet scores are better than others.

The item pool was originally divided into supportable facets by factor analysis, once the items loading highly on each factor were defined they were then separated and handled subsequently as discrete facet items, providing the facet score to be used. After a process known as 'item scaling' (Section 3.0.4. below) to give precision scoring of the response levels, each score was formulated by taking the first principle component of the set of selected items, multiple factor rotation no longer being necessary. For the short form OMQ-14, rotated factor analysis was used in the formulation of scoring, to make optimum use of the reduced item pool, and this is described under special method for Study III which addresses correspondence between OMQ-14 and OM8-30 as criterion validity for OMQ-14. Table 2.1. gives a summary of the data available on OM8-30 items; because the OMQ-14 items are a subset of those, the table also provides a useful description of the overall structure of the database, reflecting its evolution over time and centres. There are some further detailed administrative omissions of items or variables in particular centres, reflected in footnotes on numbers in tables.

There are only three major deviations from structural uniformity of the database that the reader needs to bear in mind, mentioned here. Some of the deviations from uniformity are directly visible in Table 2.3. in Section 3.0.5. There are always some small variations in numbers available to a particular analysis, around a reference denominator number of cases, and this is due to availability of case with all variables needed. But there are also major differences between Phase 1 and Phase 2. I decided for the Serbian centres and Montenegro centre in Phase 2, with agreement from the principal investigator, to reduce the questionnaire part of the study to the OMQ-14 items (by then defined from Phase 1). This was to ease respondent burden and data entry for large numbers but to allow most of the general aims of the study to continue to be served. However, the Phase 2 data additionally include the 6 URTI items not in OMQ-14 which are used in the study on seasonality. There is a sub-phase 1b involving 163 cases from Clinical Centre of Serbia (CCS- known in the computer coding as 'Old Belgrade' because of the addition of New Belgrade in Phase 2). Phase 1b data with OMQ-14 as a defined instrument do not have URTI, the decision to drop these having been made because URTI items are not very important for QoL (Dakin et al., 2010). The decision was then made to extend the Study to a broader Phase 2, re-including the respiratory infection and obstruction score URTI-6 for its relevance to treatment decision and seasonality. Because administrative in its basis, the reason for information being present or absent can be reasonably assumed to be at random (and so unbiased), except in so far as different centres are involved. It is therefore handled adequately for present purposes by the categorical centre adjustment. CCS numbers are large enough to be divided into dummy centres referring to the sequential time-periods; originally there were three (i.e. Phases 1, 1b and 2) but due to there being only small differences, these were collapsed to two and CCS simply appears as lacking URTI for the transitional cases. Again this was done for administrative reasons so reasonably assumed to be on an unbiased basis, in any analysis where timing and centre are adjusted.

The second major structural deviation from uniformity is analysed in some detail in the hearing measures article (Milovanovic et al., accepted) and concerns the smaller numbers having pure tone audiometry (HL) and tympanometry (here scaled to HL as 'ACET'). The disposition of these across centre is given later in the table for Section 4.2.5. These differences result from local conventions and also from local resource

limitations, channelled through the clinician's judgements about the relative clinical necessity for this information. This pattern of missings is not at random and hence must be presumed possibly to be biased (chiefly but not exclusively towards presumptive diagnoses as OME justifying the extra measurement, where (R)AOM might not, and hence to more impaired mean values). Such possibilities of bias require detailed documentation of sample properties, if good generalisability is to be claimed. The third major structural non-uniformity is the absence of a diagnosis item in the clinical data sheet until a late stage in Phase 1. Except insofar as it may also concern centre, this omission may be assumed to be unbiased. In general this form of missingness is adequately handled by adding a missing category to the categorical variable of diagnosis. With over 1,200 cases having diagnosis information some useful work has been done on diagnosis, presented at meetings but not within this thesis; that work covers the definition of, and the high impact suffered by, cases given a 'combined' diagnosis (i.e. OME with super-added RAOM) and it offers a mapping of OM8-30 facets to the three resulting diagnoses. The immediate implication is that with a lot of diagnosis information missing, the overall statistical significance for 'diagnosis' as an independent variable serves mainly a somewhat agnostic adjustment function, because of the large number of missings and so this is not emphasised here; more specific interpretations can be made from components, i.e. contrasts between particular real diagnoses where present, but such contrasts are somewhat peripheral to this thesis.

Table 2.1. will now be described in detail. Down the main rows are the 32 OM8-30 items with their short internal labels. The first three column heads define three reference denominators of numbers of cases approximated by actual numbers of cases available for analyses, in order: the whole database as of summer 2015, the Phase 1 OM8-30 cases with four items for reported hearing difficulties (RHD), and the number of 'hearing-complete' cases within the latter, those also having both HL and ACET. The column entries are % missing data. In general the percentage of missing is low. The presence of some numbers over 5% in the first column aligns perfectly with absence of the items from OMQ-14 as confirmed by alignment with the entries for OMQ-14 at the very far right. It is not really any deficiency, being completely explained by the administrative phases and the overall database structure, as covered in the preceding paragraph. Apart from being associated with centre (which hence needs to be controlled

for), the occurrence of these large numbers of missings can be taken as random and not due to a bias in clinician or in parent/patient. Within the column for the 'hearing - complete' 1,400 cases, there are only three items with a missing rate over 3%. The practical problem that this leaves is met in composing scores of many items, and in multivariable analysis generally, where drop-out is cumulative; this is because each data item that is added also increases the chance that it may be missing. The solution to this is guarded imputation, discussed under item scaling later. The next two columns simply give the number of levels of response category offered in the question format for each item, and whether or not a 'not sure' (i.e. scalable uncertain response, not missing) is offered within the item's response options.

The next two fields of 10 and then 3 columns refer to OM8-30 and OMQ-14 forms respectively. Below the score name is the 'guarding level', explained at the end of this section. Within these fields, entries define the scores used by their presence, and by a value for the loading on the 1st principal component of the subset of items (OM8-30) or rotated factor (OMQ-14). The G column for PCtotal shows that despite the heterogeneity referred to above, there is only one loading below 0.3 (for the child's condition requiring more effort from the parent). This item also has low loadings in other columns where it appears, suggesting that this is the one generally poor item, and might better have been dropped, although its loading on parent QoL is acceptable at 0.468. The next columns H to K simply show the weightings in their total scores (also done by 1st PC) of four of the facets as discrete item sets. This is a fundamentally differing approach to the formation of scores from that of factors as used for OMQ-14. The four scores are URTI, Ear symptom score (RAOM), sleep disturbance, and RHD-3. The entries in columns L and K align, but not perfectly because the dropping of one of the 4 RHD items for OMQ-14 necessitates a choice between a 4- or a 3-item score depending on numbers wanted. Columns Q to S show that the item of child asking parent or other to repeat is the one that has been dropped. The next three columns, M, N, O, are again straightforward specifications of facet scores: speech/language, behaviour and parent QoL.

The measure in column P with two blocks of factor loadings is somewhat recent and is used in Study III, for validating OMQ14. There has been a previous domain

division of OM8-30, broader than facets, for where aggregation rather than specific profiling was required, but allowing a distinction between upstream physical health and downstream impact, one that is lost in PCtotal. That system is not used here although it has enabled a useful determination of the roughly equal importance of the two major domains to parents (Filipovic et al., 2013). Instead we have addressed the issue of greater inhomogeneity in downstream variables in a fashion responding to the emergence of a general impact factor in OMQ14, sampling speech/language, behaviour and parent QoL as the facets from which QoL-predictive items had originated. This OM8-3O impact variable is in Column R. The 7 highly loading items in that column are a subset of the 14 items in column P in OM8-30. Inspection of details in column P shows that the column R is a fairly representative subset, not just the best loading items from column P. This is because the selection for content OMQ-14 was made on a wider basis: the prediction of generic child QoL, not just on reflecting this set of facets. This impact measure from OM8-30 is therefore a suitable, more reliable and ecologically valid, criterion measure for judging the adequacy of short-form general impact (Study III, OMQ 14). The last three columns give the loadings for rated factors in OMQ-14, showing how low loading items are still incorporated into the factor scores, with advantages for continuity of distribution but disadvantages of multi-item dropout with missing data. The factor solution is satisfactory for what was inevitably going to be an inhomogeneous score, with only two of the subset of high-loading items (child concentration 0.454; parent tiredness: 0.475) loading below 0.5.

Table 2.1. OM8-30 and OMQ-14 items, their PC and factor weights and numbers of missing data on 3 main sample definitions

	% missing					OM8-30 un-rotated PC loadings										OMQ14 rotated varimax factor loadings *		
	2886 cases	**2170 cases**	**1400 cases**	**@ Levels**	**'Not sure'?**	**PCtot**	**URTI**	**ESS**	**Sleep**	**RHD3**	**RHD4**	**Speech/ Lang**	**Beh-5**	**PQoL**	**Impact**	**ESS**	**Impact**	**RHD**
Guarding level^						23/32	3/6	1/3	1/3	2/3	3/4	2/3	4/5	3/5	10/14	10/14	10/14	10/14
Global health	0.7	0.6	0.8	4	No	0.499										0.435	0.286	0.030
Colds	6.9	1.6	2.4	6	Yes	0.321	0.502											
Throat	6.5	0.8	1.1	5	No	0.446	0.563											
Breathe	6.4	0.5	0.5	5	Yes	0.393	0.724											
Blocked	6.2	0.3	0.4	5	Yes	0.435	0.753											
Snoring	6.4	0.4	0.2	5	Yes	0.443	0.717											
Runny	6.4	0.5	0.4	4	Yes	0.367	0.520											
Ear trouble	0.8	0.8	0.8	5	No	0.417		0.875								**0.862**	0.004	0.108
Ear infection	2.1	2.3	3.4	4	Yes	0.369		0.895								**0.871**	0.029	0.002
Eearache	1.6	1.7	2.5	4	Yes	0.408		0.895								**0.857**	0.022	0.112
Hearing rating	0.6	0.0	0.0	4	Yes	0.499					0.809					0.123	0.081	**0.832**
Mishear	0.6	0.0	0.0	4	Yes	0.529				0.888	0.868					0.045	0.160	**0.864**
Hear in group	0.4	0.0	0.0	4	Yes	0.546				0.876	0.861					0.068	0.177	**0.861**
Ask to repeat	24.3	0.0	0.0	4	Yes	0.495				0.882	0.859							
Sitting still	25.2	1.8	2.5	4	No	0.342							0.510		0.485			
Concentrate	1.8	2.1	2.8	5	No	0.300									0.441	-0.025	**0.454**	0.068
Seek attention	1.7	2.0	2.9	5	No	0.417							0.722		0.560	0.094	**0.617**	-0.043
Whine	25.8	2.6	2.8	5	No	0.483							0.812		0.584			

Unhappy	2.0	2.4	3.1	5	No	0.512							0.755		0.585	0.139	**0.553**	0.142
Takeout	25.2	1.9	2.7	4	No	0.361							0.513		0.467			
Mispronounce	0.8	0.4	0.5	4	Yes	0.466						0.849			0.535	-0.089	**0.616**	0.287
Speech behind	0.7	0.2	0.3	4	Yes	0.377						0.874			0.506	-0.139	**0.617**	0.175
Articulation#	25.4	2.1	3.1	4	Yes	0.309						0.823			0.454			
Affect sleep	25.0	1.4	1.9	3	No	0.432			0.680									
Listless	24.4	0.6	0.9	3	No	0.427			0.797									
Listless ENT	25.6	2.3	2.0	4	No	0.450			0.863									
School	25.0	1.5	1.1	3	No	0.495												
Tired	0.9	0.7	0.9	2	No	0.484								0.677	0.507	0.220	**0.475**	-0.036
Attention	0.8	0.6	0.6	2	No	0.517								0.739	0.614	0.153	**0.587**	0.105
Demanding	24.6	0.9	0.9	2	No	0.492								0.714	0.625			
Energy	24.6	0.8	0.8	2	No	0.488								0.782	0.588			
Effort	25.5	2.0	1.9	2	No	0.217								0.468	0.231			

Notes:

In the right half of the table, entries denote the contribution of item (row) to score (column) and of the given relative importance weight

The 2,886 cases are those in age range 36-108 with OM8-30 or OMQ-14 questionnaire. Usual case exclusions apply; this phrase refers to out-of-age-range children, to a handful with inconsistent individual data or suspected misclassification, but chiefly to adjunct samples retained in the database for local projects such as non-ear patients (Maastricht, Budapest) and Maori and Pacific Islanders in Auckland New Zealand.

2,170 cases are those in age range 36-108 and with valid responses to all 4 RHD items. Usual case exclusions apply.

1,400 cases are those in age range 36-108 and with valid responses to all 4 RHD items and valid HL and ACET. Milan is excluded from specialised analyses on hearing as only a small subset of severe cases had hearing level data. Usual exclusions apply.

** Strictly facet names should not be used for OMQ-14 factors, rather 'Factor 1' etc but the dominant loading is shown here to enable quick identification and comparison, and the high-loading items justifying the factor identification and naming have their loadings emboldened*

@ As given the number of valid levels includes all original 'real' OM8-30 categories response alternatives but excluding 'not sure' and missings; the effective number of levels, given the scaling results in the preceding table is therefore this plus two, and 'not sure' is usually informative of a low to middle response.

 # For Target and early centres articulation took only 2 values "yes", "no". For later centres it was changed to 4 levels.

Note RHD4 is the new latest rescaled PCRHD4 variable, and not the OMRHD4 used in the analyses for interrelations of hearing measures.

^ The guarding level shows how many items must be present for a score to be generated and the item not declared to be truly missing, in effect a level-1 form of imputation (using partial data). Mostly this is a clear majority as permitted by small integer numbers, asymptoting at 10/14 (at least 71% present in each individual) for scores with 14-items. But in ESS(ear infection score) and Sleep disturbance it is only 1/3; in practice this does not mean that many more cases had scores generated on a minority of data items being present, as on 2886 cases the number with 2 items missing only one present was only 20, 0.07%, and for sleep disturbance only 2 cases.

3.0.2. ACET coding of tympanometry as an HL estimator

Research often has to develop its own new methods; this and the next section provide a past and a present example of where a new method enables the research and is also one of its products. That specification requires too much detail to sit easily within a report of the research content. Although tympanometry has other uses, the starting point for the present studies involved seeing it as essentially a screening technique. Some screening tests are very sensitive but not specific in detecting hearing loss and this 'over-sensitivity' was an early recognised issue in the suggestion for using tympanometry alone as a screen for OME (Lous, 1987). In screening and monitoring applications, the balancing requires complementary information (i.e. not redundant on the first test) which is naturally specific if perhaps not so sensitive, which can be used in parallel or in series; this classical issue provides a main reason for considering combining tympanometry with RHD (Study II). Particular applications of a questionnaire alongside tympanometry could also include selecting sub-populations of children needing fuller HL testing (case-triage rather than screening) and in such applications totalling with other continuous measures is also possible.

Due to the correlation between middle ear pressure and compliance in standard tympanometry, these mechanical measures do not produce a straightforward mapping by linear regression into severity within the normal-to-mild range. This has led to use instead of summary categories A, C1, C2, and B along this bivariate dimension. But continuous measures are more powerful. To address the issue of possible missing data in the Eurotitis study, recognised when one centre, Milan, could only provide audiograms on a small subset of severe cases, a general scaled formula for air conduction threshold estimated from tympanometry - ACET – was provided, giving an HL-equivalent mapping of tympanometry (MRC Multi-centre Otitis Media Study Group, 2009; Haggard & MRC Multi-centre Otitis Media Study Group, 2009). As defined, the ACET procedure acknowledges that most clinics go straight to the conventional categories, and it does not require handling of continuous compliance and pressure values; because of the patterns that co-occur, accepting this categorical information as input and the upgrading to a continuous measure does lose a little information, so may appear perverse, but this avoids mistakes in data transfer. The generalised expression of the formula is given below. An exact form for implementing calculation is not given because it requires many lines of syntax. For further detail the reader

is referred to the original papers (MRC Multi-centre Otitis Media Study Group, 2009; Haggard & MRC Multi-centre Otitis Media Study Group, 2009).

ACET (in HL) = a (left tympanogram category) + b (right tymp categ) + c (interaction left*right) + d (age)

where each set of tympanogram levels (coded 0, 1, 2) is in 3 levels with a,c1 combined, i.e. (A,C1), C2, B.

Applications of ACET include combining with other continuous measures for a total, justifiably imputing missing HL data in research, or substituting for HL under specified conditions when audiometry is unavailable or unachievable (Milovanovic et al., accepted). The source data remain in the domain of middle ear function, but the ACET mapping to the HL scale enables the resulting predicted value to also be more easily considered as a hearing measure as it is on a recognised hearing scale.

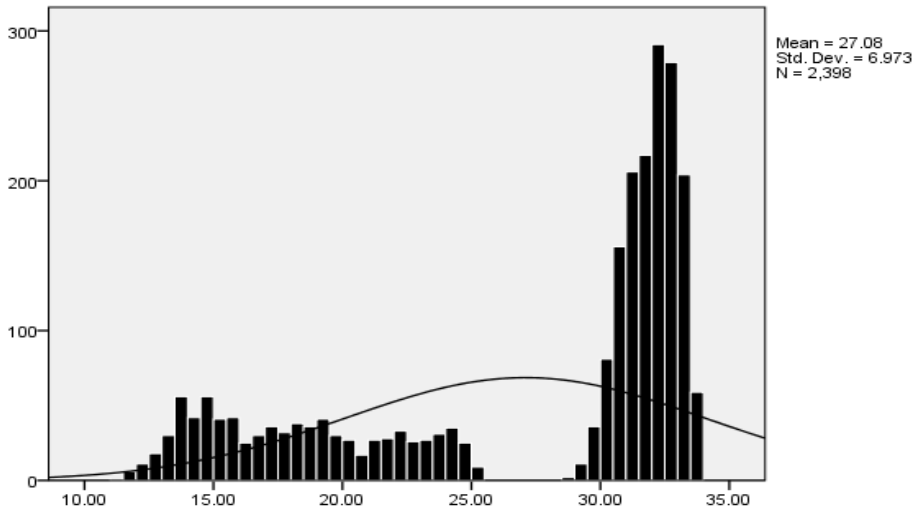

Figure 2.1. The distribution of ACET-values for 2,398 cases having tympanometry data from among the 2,886 within age-range: frequency against ACET-predicted HL value

Broadly speaking, having more than a little fluid (MEE) behind each eardrum, and hence a B tympanogram identifies a case for consideration, but largely fails to resolve severity among such more serious cases. In the values produced by the ACET formula, some slight resolution is seen (Figure 2.1.) in the width of the second peak above 25-30 dB HL, given by the fact that the HL value for a B tympanogram is conditioned by whether the other ear is

also a B (worse if the other ear is also 'B'), and also by child age. The prediction above about 25-30 dB is how ever crude, and the resulting distribution of values is extremely bimodal, with a flat barely peaked table at near-normal values, a gap, and a genuine peak above 25 dB. However for data analysis in research, the now wide availability of bootstrapping (e.g. in SPSS 21, see Section 3.0.7.) has displaced the importance of normality of distribution as prerequisite for applying the powerful non-parametric statistical tests, at least with large sample sizes. The usually ignored, though more important, pre-requisite of being able to assume equal-interval measurement remains, but here this has been specifically met by the measurement operations behind HL and the mapping to these via ACET. This provision of a best-estimate scaled HL-equivalent increases the general usability of the tympanometric information in many ways, making it now incorrect to describe tympanometry as inevitably categorical, and so we now describe ACET as quasi-continuous.

3.0.3. Scaling and imputation issues for Hearing Level

For attributing an effect size to HL, it might appear that the familiar natural dB HL scale was sufficient. However relying on conventional bands of dB HL and range of clinically met dB values as anchors for what constitutes a small or large difference is problematic, because this approach derives mostly from use in a different condition, permanent sensorineural hearing loss, with only rare cases occurring in the upper part of the range. Anchor reference values taken from SNHL are inappropriate for OME, because the variability in HL plus its incomplete mediator status entail that modest differences in HL can accompany large differences in impact. Also, a transformed HL value (for normality of model residual distribution) will no longer be on the familiar dB scale with which we started, so HL numbers emerge from the transforming stage in barely better a position than scores from an unfamiliar questionnaire score start at. Externally defined yardsticks such as the clinically important difference can be used to rescale arbitrary numbers, but such rescaling is uncommon.

The best approach, because of its universality and long history from the 1930s, involves adopting a variability-based metric, with an appropriately defined standard deviation as unit. For RHD, the scale values are unfamiliar and in a sense arbitrary, in that the units are defined internally to the psychometric development; so we require a universal expression of effect size for reasons of clear communication. This communication need goes beyond simple miss-use of p-values as if they were effect sizes. As some effects examined here are categorical and some continuous, we used not SD effect size but partial eta-squared (Cohen, 1988) as

appropriate for comparing the magnitudes of effect within one analysis, even between the two types of variable as the more widely familiar overall R-squared does for whole regression models, the proportion (% /100) of the total variance explained. The obtained effects are by no means all large, but we use verbal magnitude terms in an internally referenced way, i.e. in relation to the spread of those obtained: partial eta-squared under 0.003 is not worth discussion and for present sample sizes also mostly not statistically significant although some down to 0.002 are. We here use 'small' for 0.003 < effect < 0.01, moderate for 0.01 < effect < 0.03, and large for > 0.03. This is slightly more lenient than some recommendations (Lakens, 2013) but not greatly out of alignment with them, as in our usage these values are the lower boundaries for the three level categories, rather than their centres.

We used very little imputation of data for these analyses. Elsewhere we have set an acceptable minimum for the number of items within a facet to be present, and have imputed for a small proportion of missing. However, with the emphasis on properties of RHD here, the total sample is usefully defined as those with all 4 RHD items present (2,170); these are 98.5% of the super-sample (2,202) having an OM8-30 questionnaire at all and in the age-range 36-108 months, a very high rate of data completion for the questionnaire data. For the objective dependent variables, we did not impute for missing ears if only one ear's data were present, and with ACET we did not impute at all. Within HL, both ears had to have at least one dB reading for the case to be used in complete-case analyses; we adjusted for the particular frequencies present (a form of imputation across missing frequencies) by applying a small additive correction constant for each frequency reflecting the shape of the average OM audiogram, which is not perfectly flat. These constants were determined from the shape of the average audiogram in over 2,000 cases complete for HL, as defining the relationship between an individual frequency threshold and the 4-frequency average. Their values are: for 0.5 kHz, -4.17195; for 1 kHz -1.30195; for 2 kHz, 4.64805 and for 4 kHz + 0.82305. Mostly, all four frequencies were present for 0.5- 4.0 kHz, as exemplified by 1,749 (out of the 1,793 with RHD and usable hearing level data) having full 4-frequency binaural audiograms; the difference set of 44 cases (2.5%) with < 4 frequencies present on at least one ear but > 1 on both ears are the ones who required the additive constant adjustment for the particular frequencies actually present to give a 2-ear average HL, the precise definition of the 'HL' used throughout.

3.0.4. Item scaling for precision scoring

The computer codes (0,1) for dichotomies, or (1-5) or (1-7) or various other small ranges of integers used in ordinal category scales are usually, i.e. in simple scoring systems, assumed to give values proportional to (i.e. linearly related to but re-weightable by PC or factor coefficients) underlying equal-interval scale values for the measure of the concept. This gets a process started but it is simply wrong in general: most items have either a sigmoid or simple negatively accelerated response curve when a provisional total for the concept being measured is plotted (y-axis) against the item's response levels (x). The issues remaining are: whether it is wrong to a serious extent, and whether the amount of error introduced into the measurement process is large enough or is biased to make the removal of the introduced error worth the effort. Without undertaking item scaling and comparing magnitudes of correlation of a raw total and of a scaled total each with some relevant at least moderately correlated third variable (or some simpler index of improvement such as factor or PC variance explained) it is impossible to know. This item scaling involves a categorical independent variable for each item (usually separately) regressed against a provisional total. The provisional total can take several forms, from a simple raw sum of dichotomously coded items, through the sum of literally coded (0, 1, 2 etc.) category levels up to a PC formulated on previously scaled items from using a former smaller or differing sample, that may have been judged best until a more recent more relevant scaling exercise. The item scale values output by this process are the category level estimates from this regression; these are then re-weighted in a final facet PC. To provide tables of about 6 scaled values each, for the 32 items in OM8-30 would be excessive. However one example is given (Table 2.2.); it shows the starting scaled values which OM8-30 inherited from the TARGET study, which for reasons of sample size (N = 441) do not very well resolve the underlying dimension of RHD at the highest levels where responses are rare.

The greater sample size in Eurotitis-2, increased almost 7-fold, over the sample used originally to derive these scale values has great potential to improve the scaling. Of this increase in N, about 900 cases have come from Serbia and Montenegro, and so I have contributed either directly by acquisition in clinic or indirectly by co-ordination and data-entry almost half (i.e. about 3-fold of the 7-fold) of the increase that permits more precise scaled values. RHD-4 is a conveniently available illustration and representative, i.e. not having been chosen for great differences in the resulting scale values permitting more

improvement (if the values do not change they cannot make much difference). In general, scaling has been found to improve power often by 10% rarely by over 20%. It can this be seen for the long term as approximately equivalent to adding one whole item of average quality to a 6-item total. Where items start of being intrinsically dichotomous (steep sigmoid item response in y against criterion in x) the gain may not be large. The gain is judged from increased consistency by increasing correlation between item and the set of items defining its PC or factor of high loading and increased % variance explained by the PC or factor. In other words it squeezes noise out of the measure formulation. Often a validation paradigm is also available in which it can further be shown that the scaled version better predicts a third relevant variable of already moderate, preferably high, correlation with each version. All the scaled values in OM8-30 were recently re-derived (Haggard et al., in preparation) on the basis of the greatly enhanced sample size, in all the instances so far examined, too numerous to show here, the factor loadings increased and the % variance explained increased; furthermore, in all those where there was a validation paradigm available, the correlation involved also increased, due to the reduction in measurement error variance after the re-scaling of items.

In Table 2.2., it is shown that for suitable comparisons there is some general similarity between the two versions; this is seen in the estimated intervals for the milder less extreme responses, and so the more common response levels. This similarity has to be considered in ratio terms, and the absolute magnitude difference is unimportant (within and between versions) being finally adjusted in the item re-weighting. The regression is simply a more sophisticated standard way of expressing the differences between the response scale values in terms of the total provisional scores on all items (i.e. selected for the facet) given by those individuals who also give each response level. Choosing as example the first spacing interval relative to zero reference and the first plus the second, as non-extreme non-capped values, we obtain a ratio of 1.68, being 0.62/0.37 = 1.68; likewise 2.12/0.97 = 2.19. These two ratios of intervals are somewhat different, illustrating that to improve on the original scaled values could be possible and worthwhile, but the factor expressing the deviation from shared linearity is only 30%. The replication of such interval comparisons in the earlier stages of psychometric development obtained for the spacing of the response levels were generally closer than this, within about 15-20%, suggesting that the scale values are relatively fixed properties of the items, as conjunctions of item content with the semantic quantifiers in the response wordings. Thus precision and gradation is achieved by the scaling process.

There are also two practically important differences between the existing (Left) and new (Right) sets of values, readily seen as improvements, in addition to the greater precision from the large sample. (1) Particularly at the higher extremes, the response levels in the first version (for the rare responses received) have been pooled, giving a pair of identical values in the first two of the four items shown. This was done to increase reliability and get an estimate with narrow enough standard error to be used. This leads to an inevitable ceiling effect: poorer resolution, particularly in the high severity range. (2) The large sample has provided a justified scaled value for more categories of response including 'not sure' responses that no longer need to be default-imputed to 'no problem' or to the mean (individual item properties had seemed to determine which was more appropriate); also including two types of missing data. Detailed examination of the improvements achieved by using the Eurotitis-2 database in this way will be done in other publications. Study of discrepancy between RHD and objective measures (Study II) used the old values because it was analysed before the rescaling exercise was undertaken, and only needed OM8-30's RHD-4 items. Otherwise all the studies presented in this thesis have benefitted from rescaling on the whole Eurotitis-2 sample in the ways shown in the table, except the study already mentioned in Chapter II on RHD and interrelations to other hearing measures, which uses the original values. In small samples (< 300) it is necessary to test that a derivation has not been locally optimised and hence will give non-generalisable results. At over 2,000 cases, the present study gives no such concern about applying measures derived and optimised on the same sample.

The application of the 'missing' value raises an obvious issue: if all 4 items in RHD-4 were missing, it would need to be flagged, so as not to generate a value based on the four actually missing items, one that could be used as though it were 'real'. The OM8-30 and OMQ-14 instruments therefore divide the problem of imputing missing values for a score into two levels as a way of respecting the principle that as much as reasonable of the obtained data should be used. Missing raw data values are never overwritten, but they are overruled (in a constrained way) in computing the score to which they would contribute. At the first level a 'guard' is postulated, in the form of the number of items for each facet (for example all out of 3 or a majority for scores with more items) which must be truly present if a score is to be generated, by using the 'missing' values in the table for any missing items. If the guard is failed, more being missing, the score is declared truly missing. The ability to impute for a small number of missings in this way is considered to be a property of the instrument, one which has been made available by the re-scaling on a very large sample. Eurotitis-2 has thus

created a useful part of a piece of public intellectual property and these rescaled vales for all items, exemplified in the table, are as much applicable Results of Eurotitis-2 as they are Methods. In contrast, the responsibility for handling of scores declared missing through failing the guard, for example by acknowledged case exclusion, demonstration of missingness at random or some different level of imputation, passes to the statistician advising on the design of any new study using one of the instruments, to be judged as requiring formal effort proportionate to seriousness of the missingness problem.

Table 2.2. Illustration of item scaling for precision scoring in RHD-4 sub-score of reported hearing difficulties from OM8-30

		Old OM8-30 scale	Old PC weight	New OM8-30 scale	New PC weight
How would you describe your child's hearing? (Short name: Hearing rating)	**Normal**	0	0.898	0	0.809
	Slightly below normal	0.37		0.97	
	Poor	0.62		2.122	
	Very Poor	0.62		3.202	
	Not sure	0.26		1.414	
	999-missing item	0		1.492	
	Missing questionnaire	-		1.082	
Has he misheard words when not looking at you? (Mishear)	**No**	0	0.897	0	0.868
	Rarely	0.11		0.57	
	Often	0.56		1.829	
	Always	0.56		3.057	
	Not sure	0		0.933	
	999-missing item	0		1.217	
	Missing questionnaire	-		1.033	
Has he had difficulty hearing when with a group of people? (Group)	**No**	0	0.915	0	0.861
	Rarely	0.06		0.573	
	Often	0.47		1.835	
	Always	0.56		2.989	
	Not sure	0.12		1.018	
	999-missing item	0		1.498	
	Missing questionnaire	-		0.996	
Has he asked for things to be repeated? (Ask repeat)	**No**	0	0.923	0	0.859
	Rarely	0.05		0.389	
	Often	0.46		1.794	
	Always	0.56		2.659	
	Not sure	0		0.661	
	999-missing item	0		1.922	
	Missing questionnaire	-		1.249	

***Notes**:*

The PC weights are the first component from the component matrix for the 4 items considered only, giving a total slightly weighted towards 'best' (i.e. most consistent) items, but highly correlated with the raw total.

The scale values are only for OM8-30. The final OMQ14 values are not exactly the same although the difference is small. For OMQ14, each was re-optimised for predicting the factor score on which it loads most highly.

Missing questionnaires can be given a scaled value in the final scaled values, being a mean for the sample to which they are applied. This is on the grounds that they are missing at random and not an individual reaction to the particular question where the missing item value contributes some information, avoiding a bias; discussions of handling missing data make this distinction.

The PC weights appear to differ between original and new sets of scaled values but this is misleading; the whole is always made of four nearly equal parts. Both between and within versions, more extreme coefficients from the scaling regression seen with a strong item such as the first are then slightly scaled down again by their PC weight. Without the distribution of responses on the item the implications for the effective relative importance of items cannot be deduced, but the fact of selection of items in part for their inter-correlation and hence high internal consistency for the measure leads to importance mostly being homogeneous.

3.0.5. Data from clinical data sheet

Supplementary questioning by the clinician or transcription from hospital case notes also produced general information that characterises the entire sample, and provides what, in the context of occurrence and incidence, would be called background risk factors. Such parent responses also provide independent variables as conditioning determinants of severity within the clinical caseload, which is how they are interpreted here. Table 2.3. summarises these variables and the overall distributions of the response category levels on 2,886 cases. A standard clinical data sheet was produced and translated to facilitate acquisition of most this information or transcription of it from notes. The length of history information was not acquired in this way but from the questionnaire, although it is not strictly defined as part of OM8-30 or of OMQ-14, as uniformity of wording was desired. In one translation and reformatting history was unfortunately omitted. Likewise sex was not acquired in one small centre but is generally unimportant so receives an imputed (ambiguous) value in analyses to enable case inclusion. These, being administrative issues, as with late introduction of diagnosis, are considered missing at random in respect of the covariance structure examined in analyses, so the by far largest component of total % missing is not expected to be biased.

Table 2.3. Outline of background clinical data present and of response distributions

	Mean	SD	Male N	Female N	Non-Manual N	Man-ual N	Comb-ined N	OME N	RAOM N	Miss-ing N	Total N
Sex			1582	1256						48	2886
SES					1304	1448				134	2886
Diagn-osis							395	772	282	1437	2886
Age	62.334	15.469								0	2886
Length of history#	4.660	1.090								482	2886

Notes*:*

2,886 cases are the total sample. Analyses are mostly done on those in age range 36-108 (N = 2,865) Usual exclusions apply.

 # The mean and SD for length of history are for the 2,404 cases having valid length of history.

For the 2,886 cases with imputed length of history the mean and SD are 4.613 and 1.050.

The two large values of missing are due respectively to late adoption (diagnosis) and omission in one large centre (history) so assumed to be largely at random

3.0.6. Sampling of the European population by centres collaborating

Table 2.4. gives a summary of the centres contributing data to the present analyses. It is based on the grand denominator of 2,886, the cases within age range 36-108 months including both terminal months and having a questionnaire, although particular analyses always have smaller N due to availability of multivariable data. It gives the percentage of each centre's cases having tympanometry or audiogram information present, which can reflect both the general resourcing level and whether local referral policy encourages referral of (R)AOM as well as OME to ENT. It also gives beginning and end dates for recruitment, illustrating the Phase 1/Phase 2 difference and the percentage of the centre's contribution seen in the winter-spring (December-May) versus summer-autumn (June-November). This last can throw light on possible centre contributions to data appearing to reflect seasonality or *vice versa.* The differences in this deserve comment. Only one centre had actually reversed seasonal dominance, New Zealand, for special reasons discussed in detail in the Table 2.4 footnote. Taking the extremes outside the range of 60 % to 90% of cases accruing in winter-spring, we notice that two of the centres with smallest samples (Leskovac in Serbia, and

Athens in Greece) totalling only 96 (3.3% of total sample) provide heavily winter-dominated data. On the other hand, New Belgrade, Milan, UK TARGET sample and Bialystok, Poland have cases spread through the year with below 60% in winter. This amounts to 1,168 cases, 40.5% of the total. This property was not designed into the study, for example to illuminate year-round seasonality of severity; but it may be seen as fortunate from the point of view of power to illuminate seasonal differences by having enough summer cases that two of the largest centres had a material number of summer cases. It has to be remembered that the numbers, not necessarily the pattern of severity that they show, are thereby confounded with centre, an issue to be addressed further in due course. In the present fully centre-adjusted analyses, artefact concerning seasonality is ruled out overall although it as to be acknowledged that demonstration of summer and autumn patterns rests rather heavily on centres with more summer data and in particular again on the TARGET sample with its year round spread of more severe and persistent cases running over into the summer. Clinically, centres seeing mostly a highly seasonal winter pattern of referral with RAOM predominating might not see enough cases of the latter type to arrive at an informal impression of seasonal phenotype corresponding to the present findings on seasonality of severity for OM facets when adequate summer cases are to hand.

Table 2.4. Centres contributing data to Eurotitis-2 and to 3 of the studies in this thesis (Study II is on a subset)

	N	HL N	ACET N	First date	Last date	Winter - Spring N (%)
1.00 Belgium, Brussels	116	116	15	25-OCT-2004	19-FEB-2007	93 (80.2)
2.00 France, Paris	53	53	0	22-OCT-2003	12-APR-2006	40 (75.5)
3.00 Finland, Helsinki	85	59	20	01-OCT-2004	20-MAR-2007	57 (67.1)
4.00 Netherlands, Maastricht	197	119	70	29-APR-2004	13-JUL-2005	156 (79.2)
5.00 UK peripheral: Cheshire, Kingston, Epsom	70	68	0	01-MAR-2004	05-APR-2006	53 (75.7)
7.00 New Zealand, S Auckland & Hamilton	80	70	79	***06-Dec-2005***	***21-Apr-2009***	***23 (28.8)***
8.00 Italy, Milan (Paediatrics)	105	15	105	1 –MAY- 2006	08-JUL-2008	57 (54.3)
9.00 Italy, Trieste (ENT)	129	129	125	21-JAN-2006	01-DEC-2008	84 (65.1)
10.00 Hungary, Budapest	140	119	130	1-DEC--2006	02-APR-2008	120 (85.7)
11.00 Serbia, CCS Belgrade 1	325	221	314	26-FEB-2007	30-DEC-2011	213 (65.5)
12.00 Serbia, New Belgrade	102	43	93	31-OCT-2012	18-MAR-2014	57 (55.9)
13.00 Portugal , Oporto	50	33	47	9-JAN-2006	07-MAR-2008	41 (82.0)
14.00 Greece, Athens	47	0	47	19-MAR-2007	18-JUN-2007	43 (91.5)
15.00 Poland, Bialystok	273	256	268	28-JAN-2010	24-NOV-2011	178 (65.2)
16.00 UK TARGET multi-centre RCT	634	631	619	06-SEP-1994	17-AUG-2000	341 (53.8)
17.00 Serbia, CCS Belgrade 2	327	294	313	05-JAN-2012	12-MAR-2014	172 (52.6)

18.00 Serbia, Leskovac	49	32	49	29-JAN-2014	04-APR-2014	49 (100.0)
19.00 Montenegro, Podgorica	104	28	104	05-NOV-2013	31-MAR-2014	73 (70.2)
Total (Time range)	**2886**	**2286**	**2398**	**(06-SEP-1994**	**04-APR-2014)**	**1850 (64.1)**

***Notes**:*

The three number (N) columns give totals with questionnaire data (the general denominator for the database) and then with HL values and for those having valid ACET (i.e. both ears' tympanograms). Thus there are over 2,280 cases with an objective hearing measure; the total with the hybrid HL/ACET does not increase much beyond that for ACET, due to correlated missingness. A form of that is seen here in the general larger number with tympanometry (ACET) than with HL. For the study of hearing measures a smaller total of 1,400 having both is used and a subset of 10 centres.

The last column is the N and percentage of cases in Winter-Spring, after and May having added 6 months to NZ dates in the working data file, to reflect solar angle in the Southern Hemisphere. The fact that New Zealand apparently has fewer winter-spring cases than other centres is correct and is an arbitrary consequence of two short discontinuous recruiting periods in two sub-centres. It is not an error in data checking related to the Hemisphere. However the dates shown have been're-corrected' here so that the columns give actual historical dates as their names imply, and these entries have been italicised to attract attention to this note.

Serbia CCS, Belgrade 1 includes OM8-30 cases and OMQ14 cases, 2011 and earlier; CCS Belgrade2 includes 1st visit cases from an attempted 2-visit audit, plus the OMQ14 cases dated 2012 and later. There are 102 cases in CCS1 and 61 cases in CCS2 who did not have the 6 URTI questions before the decision to add this element from OM8-30 back into the standardisation study as an adjunct to OMQ-14 (which in standard form does not include URTI). This explains why analyses of URTI may have numbers between the general number for OM8-30 and the expected larger number for OMQ-14.

3.0.7. Statistical analysis strategy

The Eurotitis-2 statistical strategy is broad and contains many elements not all required for particular analyses reported here, so it is not appended formally in full.

General principles adopted. These are mentioned elsewhere in the thesis where examples of their application are met, but are summarised here for easy reference:

a. Maximum reasonable use of data available; systematic reasoned and explicit handling of issues of missing data;
b. General mention of all significant effects but emphasis on the smaller number of effects that are non-trivial in magnitude; so in general there is no Bonferroni correction (except for some issues in Appendix III on Balkan cases), on the grounds that we are not seeking

isolated differences as factoids to publish from low-powered data, but rather to distinguish important from unimportant magnitudes in a comprehensive approach;

c. Use of partial eta-squared to express effect sizes for both categorical and continuous variables;
d. Systematic acknowledgement of instances where error (particularly Type 2 error) may result from few and or heterogeneous items (this being as potent a limitation to statistical power as sample size is) followed by systematic comparison of alternative analyses to reach un-confounded interpretation.

Modelling conventions. Most analyses in this thesis are General Linear Models (multivariable regressions), usually linear, that is with continuous dependent variables having good measurement properties. I also used logistic regression where it is the appropriate model for simulating a screen with a decision rule giving a dichotomous outcome. These methods permit statements about the general importance of independent variables, statistical control for confounders or additive effects and exploration of magnitude and relationships other than linear where necessary, so they are the most powerful statistical techniques. We generally apply the complete list of available variables as adjusters and state study-specific reasons where this is not done and also address there any topic-specific issues of under- or over-adjustment that the interpretation may raise, e.g. one powerful effect pre-empting another through multi collinearity, entailing that a null result may not mean no effect.

To simplify reporting of multivariable models reported analyses delete unimportant effects by the process known as backwards deletion or elimination. With very large sample size, the risk of Type 2 error through premature deletion of a non-significant term is small, but the adopted standard procedure here is (a) to use 0.10 as *p*-value criterion for retaining a term as a useful adjuster in the model (our criterion for interpreting a significant effect towards a conclusion is much more stringent than this) and (b) exercised the 'intelligent' form of backwards deletion by finally bringing back terms failing retention by a narrow margin. Because of the power and parsimony problems over interaction terms in highly multivariable models, we have adopted a more stringent criterion for proceeding with them in models for component factors ($p = 0.02$) and only tested interactions where there was prior hypothesis or else found suggestion of an interaction in the aggregate data. Examples are explained further and show in detail in Study II (RHD and other hearing measures). In particular, for categorical variables there are two stages even before switching emphasis to

effect sizes that qualify: (i) the overall p-value for the term determining whether it is retained and analysed further, and (ii) the *p*-value for the component of interest and its direction, this generally not involving a comparison with the 'missing' category.

Centre adjustment, degrees of freedom and limited inspection of centre differences. The manner of applying centre adjustments with the means (or intercepts in MLM terminology) of centres fitted to allow for differences was set out in Section 3.0.1. In summary, the disadvantages of fitting centre in this way, with degrees of freedom one less than the number of centres supplying data to the particular analysis are that it increases model degrees of freedom (so, where effects are weak, threatening model stability), and that nothing very certain can be said about what the bases of centre differences are, but the control for them offered when examining other effects of interest is good. There may be some interplay of season with centre such as to give a few centres a major contribution, in terms of numbers hence power, to documenting winter-summer differences. But the power-hungry interaction test for interdependence between an overall effect and centre, with Centre *df* multiplied by the number of *df* in the other effect, might not be viable; the 'other-term' effect occasionally has more than 1 *df* for the other, categorical, variables The advantage of the very large sample size is that it generally enables a safe and stable ratio of model degrees of freedom to residual degrees of freedom. For this, a factor of 10 is sometimes taken as 'safe' but it is wiser to view <10 as positively unsafe. With sample sizes of generally over 1,000, the analyses always have a safety factor of about 50 or more for the models requited, and this is the basis for saying that Eurotitis-2 analyses can afford the centre adjustment. This is a very secure basis for multivariable adjusted analyses.

Transformations and normality of model residuals. Generally we have transformed where necessary to minimise skew of model residuals, have footnoted the particular transforms, and have mentioned where an inversion (e.g. of a square root transformation to make negative acceleration into positive acceleration) that is required for negative skew might lead to problems interpreting direction sign. The reason for transforming is not connected with normality of residuals for taking p-values literally, but with statistical power being highest when there are roughly equal numbers above and below the mean – shifting mean to be near median. With a large sample size it is inappropriate to use significance of departure from normality as the criterion for transformation because most departures are significant even if very small; we used one SE in skew index. For kurtosis the transformation

options are much more restricted, but in general there were very few kurtosis problems outside ACET, and kurtosis does not bear on the scaling issues in the way that skew does. From the point of view of *p*-values, the GLM is in any event highly robust with large samples. The clearly abnormal distribution of ACET and the spiked distributions of a small number of items with a few category levels (but scaled), we have used bootstrapping (Efron & Tibshirani, 1986), an option now available in SPSS 21, for conservative confidence intervals (CIs) from computationally intensive re-sampling. We used 1,000 re-samplings for empirically estimating the CIs. In one other (non-ACET) instance of extreme kurtosis we also bootstrapped.

3.1. Methods for Study I

3.1.1. Approach for studying periodic variations in the disease

The general information on questionnaires, their structure and psychometric characteristics are explained in the General Method approach and need not be repeated. Likewise the way of collecting data, the procedure before filling questionnaires, the age range of children. The number of centres and countries does not in general differ between chapters using the OM8-30 data and any minor differences are specified in Table footnotes. The specific statistical approach to fitting time-functions is explained further under this subheading.

Annual variation in presenting seasonality of incidence, point-prevalence or severity can be presented using a monthly distribution (histogram) with 12 bars. It is the simplest way to present the data and seems natural, but holds dangers for generalisation, parameter extraction and statistical inference. Some of the monthly variation can be due to error and under the general principle of frequency sampling, any such histogram should only be interpreted to an accuracy of 2 months or more coarsely. In the light of the theoretical discussion above of expected phase delays between facets, this could offer a poor picture of time-relations, hence of causal cascade status, between facets and hearing measures. The precise way of representing the basic timing of a periodic function is the use of trigonometric functions, fitting single sin, or sine and cosine functions, with some arbitrary starting time 0, such as 1st January used here although it gave clearer results previously when working with sine and cosine on monthly data. For the sine function, maximum is reached in a quarter of the period (a year) and the minimum after three quarters. As the second half of the cycle is a negative mirror image of the first, in practice with attention to direction sign, the fitting does not need to be done a number of times equal to the year length divided by the time bin (e.g. 52 weeks /1 week = 52 times) but only half this (e.g. 26) times. In practice, as 365 is not an exact multiple of 7 and to assist with tracking direction and wrap-around we have fitted 27. This method reflects the central tendency in the exact monthly distribution in the data, as smoothed by the sine function acting as a sampling window. It does not extract the timing of the exact position (e.g. of the week having the highest single value). Although that seems natural when using visual communication, it is strongly influenced by random fluctuations in the data (‘noise’), and by higher frequency variation in the scores than fundamental.

Periodicity or annual variation of disease is not exactly sinusoidal and different mathematical approaches to extracting it could be used, e.g. Fourier series with harmonic multiples of the fundamental. However to improve precision these would need evidence on what the appropriate model was (i.e. which harmonics should be included), and at the present stage of knowledge, such assumptions for an empirical approach to overall delay, such knowledge is not available. Fitting variation using the week as the time-quantum bin captures more accurately the time of broad maximum that will be general over many years in a way less susceptible to sampling noise. If the period of fitting the sinusoid function is 26 (in practice 27 weeks as ½ of the year, then the maximum is 13 weeks apart (quarter of the year) of the starting week; this gives the delay estimate, even if random fluctuation makes the mean value in that bin for the 13th week not particularly high. The estimation is done by fitting a continuous variable in regression, a set of look-up values for the sine function, running 27 regressions shifting the starting time along one week at a time. The statistical output used is the t-statistic for the fit of this stored function to the data. If the t-values around weeks of maximum fit are negative, then the broad adding 13 weeks (as always, whether positive or negative) but then if negative and subtracting 6 months as a 6-month phase delay equates to polarity reversal for a 1-year function. Clearly this fitting procedure could be done with or without adjustment for other influences on severity: here we did not adjust. The obtained *t*-values varied smoothly with week of start-time because of the smoothing in the sine function. We used a peak-picking procedure to provide the estimates in Table 1.3.

3.1.2. Approach to modelling two possible cascade pathways for disease impact

It is desirable to model causal relations between facets in a way consistent with our understanding of the canonical pathway of OM forms. This involves expressing a logical (causal cascade between OM disease aspects) in the form of a structural equation model (SEM; Haggard et al., 2015) is done as the relation between them. Prioritising parsimony above excellence of fit which may demand complexity, the complete set of facets can be divided economically into two sets of contrasting dichotomies: (1) upstream versus downstream, attempting to summarise causal origination or independence versus dependence, and (2) a mainly health pathway contrasting with an impairment pathways, expressed in the model as the path A and B. The A and B pathways are not strictly separated over very many stages and both data and known pathogenetic influences suggest some cross-linkage between the two pathways. The Eurotitis-2 SEM and its pathways are presented in Figure 1.1. It shows

some re-uniting of cascades at various points. If the pathways were to haves different delay characteristics (an there is no overriding reason for them to be similar) then the re-joining will diffuse the differing seasonalities of the previous stages in each cascades in their effect on the downstream measure(s).This will lead to absence of measurable seasonality in the variable after the re-uniting. Thus the theory expressed in the conceptual diagram (Figure 1.1.) can, using some seasonality delay data, make predictions for variables where no seasonality at all may be measurable. The present delay estimation results are completely new but may be integrated into a rich integrative theory as expressed by the structural equation model.

3.1.3. Statistical analyses

We use 2,786 cases from the Eurotitis-2 survey for the regression analyses fitting 27 sinusoids with delays at multiples of 1 week and measurement interval (bin) of one week. To allow for twin adjacent identical maxima of sine fit to the data in the sine delay function for two adjacent weeks, the delays are quoted to the nearest half week. All the main facet variables supported by the OM8-30 were modelled: URTI, ESS, Sleep disturbance, RHD, and general impact, also the performance measure HL. To examine possible trade-offs between specific validity and general reliability, three of these were subdivided: RHD into the 3 communication items and the single-item hearing rating, and the URTI symptoms scores are separated into two components: infection and obstruction. General impact is heterogeneous, being composed of items originating in, behaviour, speech/language and parent's quality of life and these also were separated. For maximising numbers of cases, the HL models here did include the sometimes omitted 4 centres Milan, Leskovac, Podgorica, and New Belgrade after a check on robustness of findings (these centres have a high percentage of missing data on HL, hence possible biases). The following list of determinants was used as adjusters: centre, age, sex and history of disease. Significant independent variables were kept in the model at $p < 0.05$. In Table 1.3. the maximum severity weeks and strength of seasonality are located via their p-values and partial $\acute{\eta}^2$ is used as the effect size index. Monthly and weekly distributions of the cases are illustrated for visual comparison. Upstream facets are those starting off the canonical pathway and the prediction is that their maximum severities occur near the beginning of the season. Downstream facets such as generic developmental and wider impact should have later maximum severities. We shall see that the results are highly consistent with the former and not inconsistent with the latter prediction.

3.2. Methods for Study II

The General Methods Chapter covers seven main issues in measures, sample and statistical analysis all of which arise in this chapter. However as this is the first appearance in the thesis of multivariable analysis, where effect sizes can be compared between terms within models, the discussion of effect size is amplified below.

3.2.1. Conventions adopted for reporting models of determinants: *p*-values versus effect sizes

The chief analyses used the general linear model (GLM -- also known as analysis of covariance and multiple regression; Rutherford, 2012) for a continuously distributed variate (dependent variable), for the statistical control required by the research questions. Variates were transformed where necessary to eliminate skew, as specified. To avoid premature deletions during the backwards elimination of variables in arriving at an economical model of the various determinants of extent of hearing problem we used a *p*-value to exit of 0.10 so as to document marginal findings of interest but in fact very few effects failed to reach more conventional $p < 0.05$. We adopted a conservative approach to the many possible interaction terms, considering only the 1st-order interactions between the 7 overall (main) effect terms (21-1 = 20 of them) accepting them only at $p = 0.02$ with consequent limited reporting and discussion here.

3.2.2. Centres and centre differences in the analyses of hearing measures

The latter number of Eurotitis-2 centres is 16 for OMQ-14 questionnaire data, but this reduces with any special data requirements such as for HL and tympanometry. Intersecting these with the requirement for all four RHD-4 items, 10 centres contributed earlier-phase but fuller OM8-30 data to the 'complete' 1,400 cases chiefly used here. The centre term thus absorbs 9 degrees of freedom here, about the same as the number of substantive variables in most reported models; this is a quite high 'price' to pay in parsimony relative to how it can be handled in multi-level modelling (MLM). However the large sample size makes the price affordable, and the provision buys power gain through reduction of error and confounding, thus permitting deferment to the MLM context, of the detailed consideration of centre differences and effects carried by centre differences.

3.2.3. Case inclusion and missing data

We used very little imputation of data for these analyses, as given the examination of possible biases in the cases with complete data on the hearing measures it was generally not necessary. With the emphasis on properties of RHD here, the relevant total sample is usefully defined as those with all four RHD items present (2,170); these are 98.5% of the super-sample (2,202) having an OM8-30 questionnaire at all and in the age-range 36-108 months, giving a very high rate of data completion for the questionnaire data. The General Method Chapter gives the data availability by variable and explains guarding for unacceptable numbers of missing items per case in questionnaire scores as part of item scaling and the method of imputing for a small proportion of missing. One level of imputation issue specific to HL, analogous to guarded imputation for questionnaires, is also covered in the General Method chapter. For the objective dependent variables, we did not impute for missing ears if only one ear's data were present, and with ACET did not impute at all but we let these cases be absent.

3.2.4. Specification of standard measures, reported hearing difficulties (RHD) and item scaling

Audiometers in all sites were declared to be calibrated, with use of test procedures to the national (i.e. internationally prevailing) clinical standards in hospitals in the developed world. The situation for tympanometers is less clear in practice, and I have been unable to locate any recent studies of actual tympanometer calibration. Possible centre differences in tympanometry will be addressed within the centre differences report. It is likely that calibration issues widen the differences due to centre, so some caution is required when interpreting tympanometry data pooled over centres. The full centre adjustment described 'purchases' the ability to ignore this for the present purpose of generalisation about other variables, which are as a result documented, in effect, on a within-centres basis.

We have based the present documentation of measure properties on a fuller number of items (4) on a sample size approximately three quarters of the now augmented Eurotitis-2 total sample; the power considerations trade equivocally (accepting ¾ sample versus accepting ¾ items) but the emphasis here is on measure more than on population. The later enlargement of the Eurotitis-2 database has enabled further refinement of the precision of the RHD-4 measure and a more comprehensive 9-item version (Haggard el al., in preparation)

but the principle of scaling is illustrated and present values used are given in the Table 2.2. under General Methods.

3.3. Methods for Study III

3.3.1. Questionnaires

The basis of the OMQ14 questionnaire, created by selection of the best items from the QoL perspective, out of the pool in the longer, OM8-30 form was described in the preceding introduction in some detail. OMQ14 items can be aligned with three domains (i.e. they have their highest loadings on 3 distinct factors): i) physical health [compose from what was originally two facets: general health (1 item) and ear symptom scores (3 items); ii) RHD (3 items) and iii) general Impact [from originally three facets: behaviour (3 items), speech/language (2 items) and parent's quality of life (2 items)]. These three main domains, summarised by varimax rotation of the factors together, make up a broad reflection of symptoms and quality of life which in a summed form as the PC total can be used as a QoL measure. The choice of criterion variable from the antecedent longer form, OM8-30 is based on choosing all possible items originating from the same facets in OM8-30 as are selected for OMQ-14, so the criterion measure has greater reliability and generality. Appendix II addresses the only one of the three (impact), where this was not an obvious alignment with previously used scores, and shows a very satisfactory solution. The other issue for criterion validation is that in OM8-30 all facet scores inter-correlate at least moderately, cases more severe in one facet tend to be more severe in another. This stands in contrast to the situation with measures from OMQ14, where the factor scores are maximally separated (by the orthogonal rotation) and have zero correlation. It is worth facing the possible difficulty for criterion validation that inter-instrument correlations will be lowered, because the number of aspects that could be reliably supported for a profile is reduced by the reduced number of items, necessitating that OMQ14 be scored on the basis of rotated orthogonal factors for maximum efficiency in using all information in the items.

To examine criterion validity, it is necessary to have one or more appropriate criterion measures. The issue is essentially whether the discarding of items in OMQ14 leaves a highly sufficient (but desirably shorter) version of what was available in OM8-30. The different principles for scoring the two instruments make this question not entirely straightforward. For PCtotal and for RHD, corresponding scores exist in the two instruments. For ear infections (ESS) in OMQ14, the extra item of global health judgement is retained, although it is a cross-loading item, it predicts QoL well and loads highly on the 1st PC. The alternative actions were

(a) to include it with this factor, or (b) to leave it out and only contribute to the formula for PC. As it loaded 0.436 (above the often-used criterion of 0.4) it was included (alternative a). For general impact, the issue of appropriate criterion was more complex because such a construct was not used in the scoring of OM8-30 which was structured on a more *a priori* basis. The heterogeneous construct of impact [made up of parent Quality of life (2 items) plus behaviour (3) plus speech/language (2)] was only defined later with OMQ14, when surveying the items selected as highly predictive of QoL and finding *a posteriori* that this set made an acceptable factor. Thus to obtain an appropriate criterion measure from OM8-30 we had retrospectively to construct one with the same facet balance, from the 14 items originating = in the same 3 OM8-30 facets (for details and good distributional properties, see Appendix II). This gave us the full 4 measures for assessing criterion validity.

3.3.2. Samples, design of analyses and details of statistical methods

We used 2,865 cases for the preliminary principal component analyses (PC) and then definition of the factors for scoring OMQ-14. Correlations with OM8-30 were done with maximal number of cases available for the OM8-30 facets. The main-effect models for PCtotal and the three factors are run on the same number of cases, with backwards deletion of variables non-significant at $p = 0.1$ according to the strategy in General Method. Variables in the region $p > 0.05$ make very little difference to the rest of the model. We used 6 independent variables: age, gender, SES, history, season and diagnosis to capture the determinants. Significant variables were kept in the model. Parameter estimates (i.e. coefficients) expressing a difference or regression slope are available, but effect size was estimated using partial eta squared ($ή^2$) to embrace continuous and categorical variables and enable comparisons, and as the scores do not have familiar natural values (See General Method). Effect sizes are presented along with direction of the effect for interpretation in the discussion. This main-effects model (i.e. with significant variables after back deletion of non-significant terms) was then fitted with interaction terms, and interactions significant at $p < 0.02$ were kept in the model (see General Method). This final model thus benefits from a maximum number of cases, but the 1,866 cases with all measured hearing data in (permitting. the same model with added HL and ACET) was of particular interest (see Appendix IV). This leaves the issue, as in Study II of the comparability of the two samples, for example whether the model presented without hearing variables should be the one on this smaller set of identical cases or the one having maximum cases with questionnaire data. Preliminary

comparison showed that no direction of effect differed and the differences were not large but that effects were nearly all slightly stronger on the 1,866 identical cases also having the hearing measures, the explanation presumably therefore being as in Study II, the data quality or severity distribution accompanying data presence.

SES has missing data and in the *p*-value table (Table 4.3.4. – shown later) two values are presented, the first being for missing versus more favourable SES , the second for less favoured. Only the interpretable component, for less versus more favoured SES is presented in the effect size (Table 4.3.5.). All four levels of the diagnosis categorical variable are presented to show how this may be relevant to OMQ14 scores. 'Missing' here includes the first phase of the study where diagnosis was not requested. The levels are; missing (D0), combined [OME and super added RAOM (D1)], pure OME (D2) and RAOM (D3). The *p*-values for D0, D1 and D2 and the partial eta-squared are presented in contrast to D4 (RAOM) as reference. Residuals from the models are nearly normally, distributed, with only slight skew and kurtosis.

In Appendix III, I briefly summarise the data available from Balkan centres (not counting Greece) versus others. The Balkan sub-database has 904 cases from five centres Old Belgrade1 (OB1), New Belgrade (NB), Leskovac (LE), Montenegro (MNE) and Old Belgrade2 (OB2). The data from old Belgrade are divided into two subsamples: OB1 cases with OM8-30 questionnaires and later, OB2 cases with OMQ14 questionnaires. The Balkan sample size is large, about one third of all cases. The comparison between samples is potentially informative about differences and similarities between the rest of Europe and the Balkans, including a picture of healthcare accessibility and OM severity. The overall results did not show big differences, and some variables feature are centre-dependent, even after Bonferroni correction for avoiding type I error testing multiple hypothesis. One feature of this sub-data base is the small number of missing data (3.65%). Some differences between centres present may be explained by reorganisation of health policies on efficacy, rationality, prevention and introduction of incentives in reimbursement. Some further research potential of this subset of the data is suggested by Appendix II but not the topic of this thesis.

3.3.3. Statistical issues and methods for Study III

All statistical analyses were conducted using SPSS version 21. To specify the properties of OMQ14 we report first the final stages in its derivation, using principal components and factor analysis (Varimax rotation). Extraction of the factors followed the Kaiser criterion, with minimum eigenvalue > 1.0 but deciding on the basis of interpretability from prior knowledge of item content, not necessarily extracting exactly the full number of factors having eigenvalue > 1.0. The loading patterns for principal component and for factors express internal consistency in the most comprehensive way. In other work that is typically summarised as the average inter-item correlation, Cronbach's alpha.

Criterion validities of OMQ14 scores are computed as Pearson correlation coefficients on the same cases that also have the criterion measure from OM8-30.The numbers of these identical cases vary from 2,149 to 2,183 for particular measures. Distributions were continuous and normal or near-normal for both PCs, for both impact measures and for the three OMQ14 factors, but with slight positive skew for impact, with platykurtosis for RHD OMQ14 factor and both platykurtosis and positive skew for the OMQ14 ear infection factor. For the RHD and ear infection OM8-30 facets, the small number of items makes the distribution striated, even after item scaling, illustrating one advantage of the factor method of scoring. The deviations from normality are not a concern for two reasons. Where they are to be used as dependent variables, most of these variables can be readily transformed to near-normality or bootstrapping can be used; this was also done here, but it made little difference. For use as an index of agreement and particularly linearity of relationship between measures with assumed interval-level measurement, the Pearson r does not require normality. Significance of some correlation is not in doubt as similar formulations with many overlapping items are being tested with a general expectation of relationships stronger than 0.90. With the very large sample size, the p-values for relationships around r = 0.90 will generally have about six zeroes after the decimal point so significance is not discussed for Objective 2.3.1., only for Objective 2.3.2. and Objective 2.3.3.

The 1st principle Component and the three rotated factors summarising the profile information are the scores in terms of which OMQ14 can be compared to the OM8-30 equivalents (by simple correlation for criterion validity) and to probable determinants or antecedents (four types of General Linear Model with each of these as dependent variable). This latter type of analysis underpins recommendations for control for confounders while

controlling for known significant influences, and so provides information on construct validity (Objective 2.3.2 and Objective 2.3.3.).

3.4. Methods for Study IV

3.4.1. Design of analyses and predictive variables used

We used the 1,400 complete cases from the Eurotitis 2 study database having the RHD-4 questions, plus ACET and HL. Here the RHD questions are separated into two variables (as previously described in the Study I); the first question about overall hearing and other three, communication questions: 1. *Has he/she mis-heard words when not looking at you*; 2. *Has he/she had difficulty hearing when with a group of people*; 3. *Has he/she asked you to repeat things?*. The reason for separating the RHD-4 items into two sets (one single and set of 3) is that the differing seasonalities in relation to other hearing measures or even non-hearing measures as variables in the same regression model could contribute to predictive power of the model. It could do so because of the different timings of maximum severities of different hearing questions during annual cycle, as described in the Study I; hearing rating at late spring, just after HL peak severity and overall three communication questions more towards early summer. The single overall hearing rating is fitted separately despite possible reliability problems, because it is the best item, in the sense of correlation with HL (bilateral average of 4 frequencies per ear from clinical pure-tone audiometry) is independent variable in the underlying GLM. To achieve better distribution of residuals (see General Method), this HL is mildly transformed (using ln(HL+9) for this sample and set of terms). The overall strategy has two logical stages: (1) to use the GLM to describe fundamental relationships for the whole severity range; and then (2) in logistic regressions to simulate a screening application with a categorical outcome (e.g. 'probably > 25dB, so refer') to examine which of the terms in the GLM remain predictive for a particular suggested cut-off in the HL, expressing the cases whom it is thought desirable to refer. There has to be general agreement between the two types of analyses, but details can differ and specific relationships between predictors and criterion HL may not be the same around each cut-off that is tried. The form of any such discrepancy will mostly be that a term significant in the GLM is not significant in the logistic because logistic is less powerful.

3.4 2. The underlying HL prediction model

The underlying GLM had HL as dependent variable (DV) and the following variables in the regression; ACET, overall hearing rating question, RHD-3 communication items and season as sine/cosine pair specified at 1-month precision from the look-up table as described

previously. This underlying model was examined in two forms: i) without centre adjusted and ii) with centre adjusted. The later logistic regressions to simulate screens were also run in both ways, the purpose for both types being to examine how particular effects of interest like season might be subject to instabilities related to particular centres, however the centre-adjusted version was declared in advance to be more appropriate for the logistics regressions as capturing the conditions expected for any one population and assessment centre a real screen environment. The strength of the model is explained using Rsq (R^2) as a measure of the fit achieved by the structural relationships embodied in the model. The model with centre adjusted was the better of the two as expected, and so is used to report these analyses. Effect sizes of variables in the regression model are expressed using partial eta squared ($\acute{\eta}^2$) and interpretation of the value is done according to our adaptation (see General Method) of the principles of Cohen (Cohen, 1988): small (0.003-0.01), medium (0.01-0.03) and large (> 0.03).

3.4.3. The model with interactions

Interactions between the two types of hearing questions (parts of RHD-4) and ACET in influencing HL were considered, because interaction had formerly been shown for the whole RHD-4. For degrees-of-freedom reasons (i.e. stability) they were fitted separately, one at a time. The reason is that for not very strong effects it is entirely possible that for example seasonality adjustment for the rating and for RHD-3 would be in direct competition, and we need to know whether it should in general be present, even if one such adjustment is knocked out by the other due to multi-collinearity. Significant interactions were at this stage kept in the model for $p > 0.1$ to avoid possible Type 2 error in later consideration of the logistic, but in the knowledge that they could well then be eliminated as not contributing more than trivially. The direction of these interaction influences (captured by sign of the regression coefficient: ACET*PC RHD3 items) is important to give a rational explanation for retaining an interaction term, not just to better predict HL in one set of data.

3.4.4. Selection of cut-offs and decision strategy for logistic models

We used four cut-off values for predicting hearing loss in logistics with centre adjusted. After a preliminary survey all interaction terms were dropped because the interaction terms did not account for much variance in HL. This is understandable because when separate cut-off values are defined for separate logistic regressions, the interaction term seen in the GLM

will be represented not as greater or less contribution between those particular analyses, but in their separate corresponding overall ('main') effects. In the logistic for predicted HL we used four variables: HR, ACET, PC-RHD3 and the rating item. I describe later a supplementary model adding an extra variable (ear infection symptom score-ESS).

The appropriate screen performance parameters for the logistic regressions, given below, were obtained for all four cut-offs but I focus here on 20 and 25dB for reasons examined later in discussion. The performance parameters used are given below. It is also possible to discuss positive and negative predictive values to project the implications of having a screen of the type suggested, by using estimated reference prevalence from age 2 - 3 years from two studies. Positive and negative predictive value (PPV and NPV) are the respective probabilities that a child with a positive test truly has at least the defined level of hearing loss, and that a child with negative test does not have such hearing loss. The parameters depend on the prevalence of the disease, so are not only a property of the predictive power of the test. In a population with high prevalence, the PPV will rise and NPV will go down.

1) Sensitivity of the prediction model say what the probability is that the child identified with the screening test positive using RHD/ACET instrument have HL > cut off (20, 25, 30 and 35 dB).
2) Specificity is the probability that the screening test will accurately identify those with HL < cut off value (in our case percent of children rightly identified with normal hearing using RHD/ACET score). However it is not customary to quote these when continuous scores from the screening test or set of variables are available, because there is a trade between them and relatively free choice in where the cut-off in the independent variable (screening test) is applied to match the decided desirable cut-off in the criterion variable (diagnostic test). They are therefore replaced by:
3) Specificity at a given level of sensitivity thought acceptable, usually 90%, and
4) AUC-ROC, which is the area under the ROC curve, a reflection of how accurate the test is in general. AUC-ROC is stable measure of test performance and is estimated using a range of hypothetical cut-offs of differing sensitivity and 1-specificity (false positive).

SPSS logistic regression produces the option of ROCs with AUC summary value and a table of values for reading off specificity at a given sensitivity. After checking for deletion of interactions (and potentially of weak overall effects) in the logistic regression we proposed the best model, with AUC-ROC as the chief summary index. At this stage after the

underlying GLM and with the stated strategy, the entry of the terms into the logistic regressions is not of prime importance. However to allow some comparison of strengths of predictive contribution for different cut-offs, it is desirable to show odds-ratios, and their confidence intervals. Furthermore, to allow comparison between predictor variables at any one, or at all cut-offs, it is desirable to do these on pre-standardised variables. This is because the odds-ratio for a continuous independent variable is given per unit in the independent variable and in general these units will not otherwise be on the same scale permitting comparison. Normality of distribution is not essential for independent variables, but if standardising, having the best approximation to normality is good practice. Transformation to minimise skew in ACET before standardising brings the highest AUC-ROC down from above to just below benchmark value of 0.9 but the difference is small and the intention is unbiased analysis not seeking ways to present impressive figures. The odds-ratio (OR) measure is the exponentiated B-value from the logistic regression equation considering the ratio of probabilities of being above or below the cut-off as a continuous variable. In medicine it is thought of as a 2 X 2 association index and sometimes this is appropriate. But the real reason for its widespread use, as here, is the ability to apply multivariable statistical control and estimation of size of effects as in a GLM, for either a continuous or categorical predictor (independent variable). OR is one type of measure of effect size, on a different scale from, but conceptually equivalent to partial eta-squared measure.

4.1. Results of Study I

4.1.1. Descriptive month versus week scores for facets severities

Despite the reservations about documenting seasonality with an 11 *df* fitting of the monthly pattern, the distribution of the facets/factors scores and hearing measure over the 12 months serves for initial descriptive purpose. Number of cases, mean score and SD of the ear infection score (ESS) are presented in Table 1.1. and the monthly score severity in the Figure 1.2. A similar descriptive view of HL with table of the monthly mean scores, SD and number of cases is given. From the tables and figures it is not easy to see any dominant pattern of periodicity in these upstream markers, only the crude monthly variation. Some months' scores are very high and the absence of nearby high scores confirms the point made that these cannot be taken as reliable, due to fluctuations and possible error when we allow and estimation process with 11 *df*. Likewise the larger of differences between adjacent months does not necessarily show up expected switches between a high and low season, due to middle and high frequency oscillation over months. The best that can be said from this type of inspection is that amid much variability, the general pattern of scores is higher in the first four months but there is also an unexpected rise in July, not matched in the month before or after. The number of cases giving the ESS data, was the highest in the February and after in March while severity incidence in the July, than January, March and May.

Table 1.1. Monthly N of cases & mean ESS score and SDs in 2,881 cases in Eurotitis-2 Study

	N	Mean	Std. Deviation
Jan	152	1.5237	.32464
Feb	208	1.4754	.34177
Mar	197	1.5082	.34325
Apr	188	1.4972	.32863
May	129	1.5079	.36117
June	101	1.4536	.36223
July	65	1.5579	.29989
Aug	52	1.4334	.33875
Sep	67	1.3931	.32067
Oct	128	1.4903	.34652
Nov	109	1.4743	.32515
Dec	124	1.4264	.30834
Total	1520	1.4841	.33650

Table 1.2. Monthly N of cases & HL severity, (transformed for models), means and SDs in Eurotitis-2 study

	N	Mean	Std. Deviation
Jan	152	6.0037	.75747
Feb	208	5.9408	.87630
Mar	197	5.9767	.90058
Apr	188	6.0818	.86028
May	129	5.9991	.85991
June	101	6.0268	.88537
July	65	6.1236	.82741
Aug	52	5.6960	.92248
Sep	67	5.9402	.88732
Oct	128	5.9403	.78491
Nov	109	6.0988	.79830
Dec	124	6.0503	.78106
Total	1520	5.9995	.84606

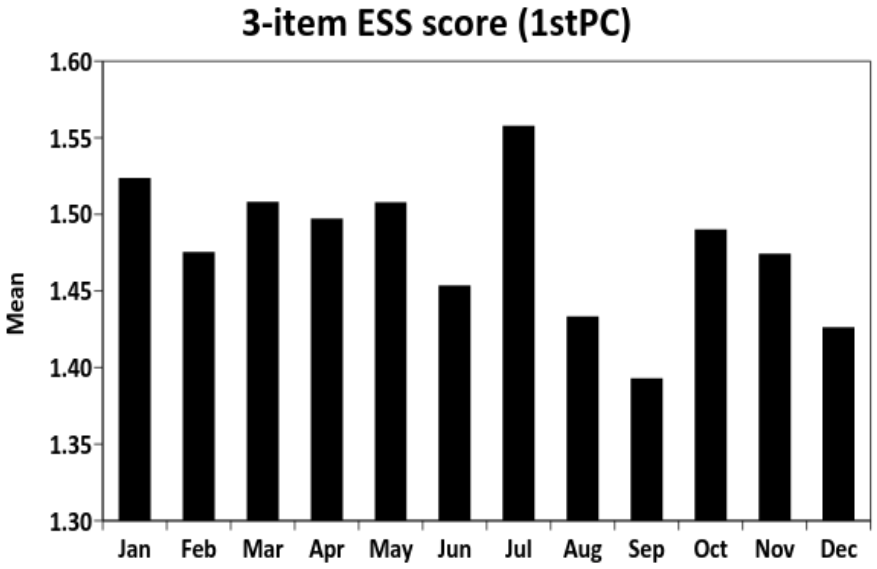

Figure 1.2. ESS scores: distribution over months of the year in 2,881 cases with data

For HL the highest individual month values are for April and July (Figure 1.3.) but generally late winter and early summer have more severe cases than other parts of the year, although this is not convenient to documented starting from this type of data presentation. Corresponding results are presented for other facet scores in Appendix I.

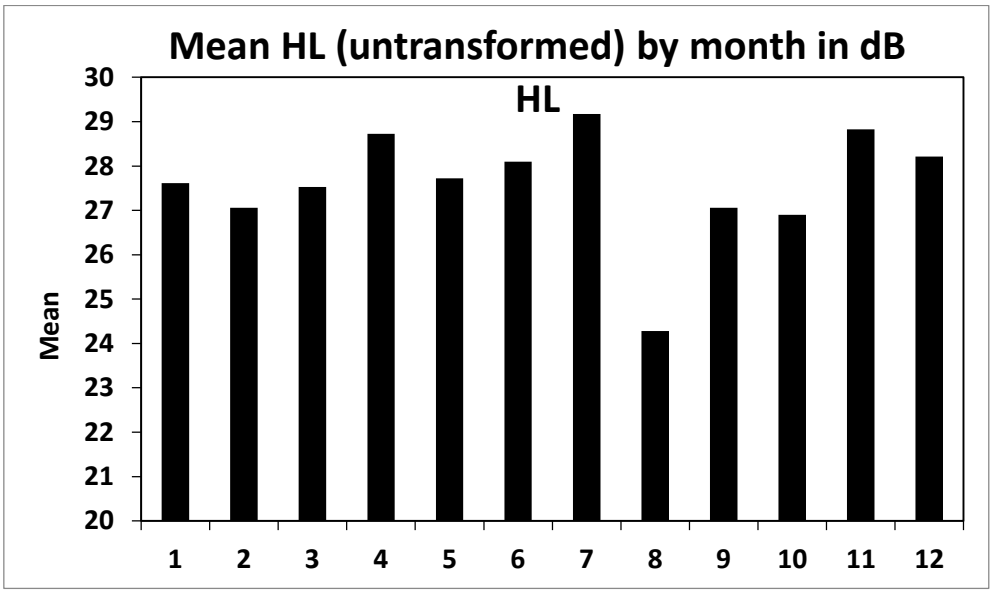

Figure 1.3. Illustrative monthly seasonal pattern for mean HL of 1,520 cases with data

***Notes**: The untransformed scale shows the modest differnces in the mean, but the analysis models used transformed HL: square root of (HL+9).*

Using the sine function fitting technique with week as a time period for descriptive analyses gives more precise information about the overall timing of the broad annual peak of maximal severity; just how far the week as unit can improve on the month, cannot be predicted in advance as it depends on intrinsic variability and event-rate – here the number of cases with a single consultation. This way of minimising error within the limits of the data potentially offers a better view of inter-facet comparisons in characteristic delay. Figure 1.4. shows the weekly pattern of the ESS severity scores. Weekly distribution of the other facet score severities and hearing measures is given in the Appendix I. Some of these descriptive diagrams show a strong fundamental (sinusoid) but also show some time structure with shorter periods or oscillations of higher frequency than the fundamental a structure possibly worthy of replication, weeks, but tending to noise at periods shorter than 6 times the fundamental (2 months).

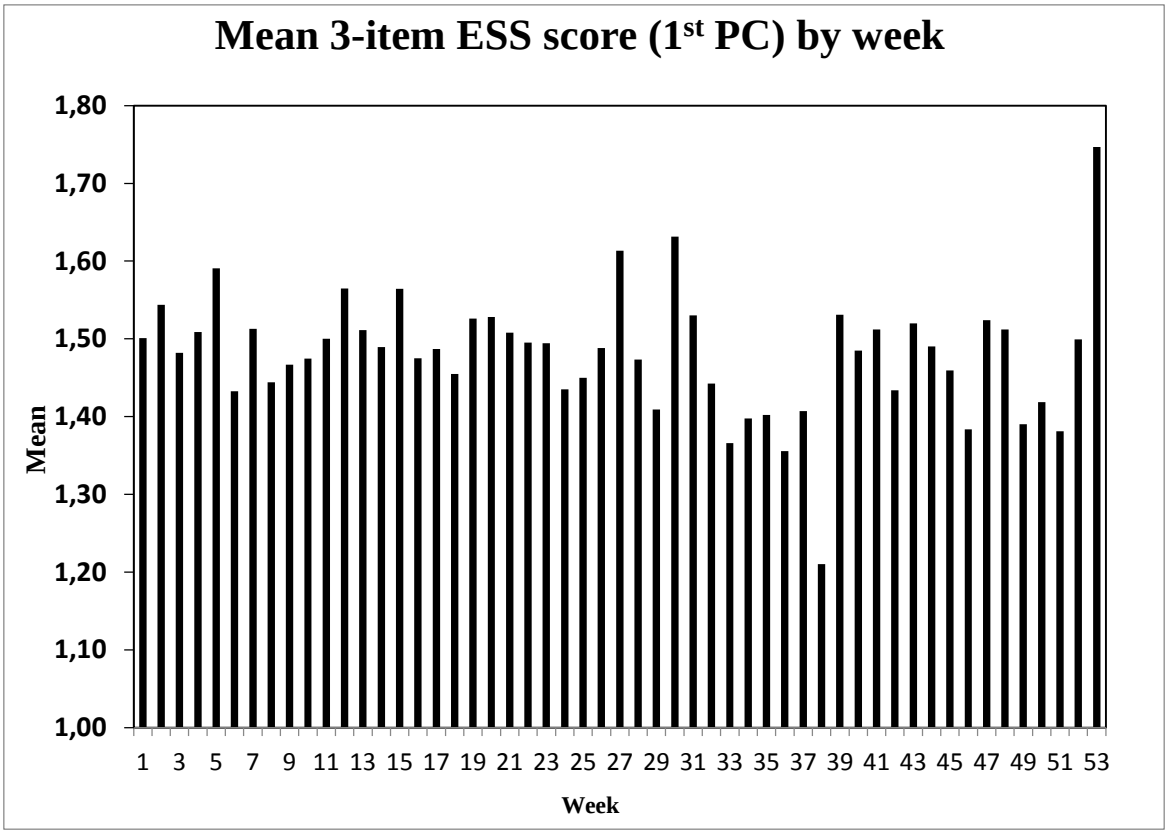

Figure 1.4. Weekly distribution of the ESS score from the starting week, the first in January

4.1.2. Seasonal severity for main facet scores

Upstream facets: Adopting weeks as time unit for the regression models of these dependent variables gave in all instances a sufficiently distinct annual picture for facet score severities and for hearing level to postulate a characteristic number of weeks delay from the beginning of the year. In Table 1.3. are presented *p*-values for the fit of the sine function for those weeks having maximum fit with the centred maximum activity occurring 13 weeks (one quarter period) after this as explained in preceding pages; also given is the effect size (partial $\acute{\eta}^2$). All variables used in regression are kept as significant at $p < 0.05$ except SES score ($p < 0.06$) which was marginal for ESS but is kept in the data for consistency of documentation as it is known to be important for certain dependent variables. For URTI infection, ESS and HL severity scores, sinusoidal seasonality is significant at $p < 0.005$. Partial $\acute{\eta}^2$ is here considered worth discussion if it is between .003 and .009 moderate from 0.01 to 0.10 and strong if above this, a set of anchors slightly more lenient than for univariable ANOVA (Cohen, 1988; Miles & Shevlin, 2001).

For the upstream facets, the delay estimates cluster together in the early part of the year as expected. Of these the earliest estimated delays, are around the beginning of March for URTI obstruction and URTI infection; the small estimated timing difference between them is within the margin of error. These also have the strongest seasonalities. For sleep pattern, the severity has no significant seasonal peak, i.e. its effect is very weak so the apparent timing of the maximum should be viewed with scepticism; for what it may be worth, that timing occurs a week before the timing for URTI obstruction (7.5 and 8.5 calendar week respectively). These three upstream measure severities (URTI, ESS and HL) peak early at the year, 8.5, 10.5 and 11 weeks. They are very closely related and their effect is very strong. It is perhaps surprising that timing for HL is so close to the other two, but the shape of the HL data may help explain this. An early peak for high HL only half a week after AOM peak severity may convey that the MEE in AOM and recurring (so presumably previously known) OME can return rapidly. However the high HL season is long, consistent with later emerging awareness of continuing hearing loss and related problems in cases traditionally described as OME, that is a continuing flow of cases due to differing views on what is a serious and continuing problem worth consultation, this occurring through late spring into early summer.

Table 1.3. Maximum severity weeks of main variables, starting week, p-values and ή²

Dependent variable (Facet, k- items & cascade of impact§)	N	Model's adjusted Rsq	*p*-values	Other determinants (of those tested*) significant at $p < 0.05$	**Calendar (week) of maximum severity#**		Partial ή² sine (week)
URT infection (3) A, B	2649	0.087	0.000	History, age	*22/ 23*	**9.5**	0.006
URT obstruction (3) A	2689	0.062	0.112	SES, history, age	*21/ 22*	**8.5**	0.001
ESS (3) B	2786	0.026	0.005	SES, history	*23/24*	**10.5**	0.003
Sleep (3) A, B	2128	0.063	0.478	SES (@) History, age	*20/21/22*	**7.5**	0.0002
HL A	2168	0.124	0.000	SES (@), History, age	*24*	**11**	0.009
RHD (4) A	2170	0.214	0.016	SES, history, age	*2/3*	**15.5**	0.003
Hearing rating	2170	0.107	0.018	SES, history, age	*1/26/27*	**13.5/14**	0.003
PCRHD (3)	2170	0.228	0.028	SES, history	*3/4*	**16.5**	0.002
Behaviour (5) A,B	2104	0.114	0.163	SES, history, age	*10/11*	**23.5/24**	0.001
Speech/language (3) A	2133	0.089	0.456	Sex, SES, history, age	*24/25/26*	**11.5/12**	0.0003
Parent QoL (5) A,B	2137	0.094	0.621	SES, history, age	*10/11/12*	**24.5/25**	0.0001

Notes:

 # Differing seasonalities are expressed in weeks as delays reflecting how up- or down-stream the facet is in the causal cascade; Italic letters represent starting week and bold the maximum severity week. Transformations used in the week-based sinusoid analyses were as follows: URTI-6: None; URT infection: SQRT(6.3-inf), URTI obstruction: SQRT(obstr+3.7)), ESS: SQRT(ears+2.3); Sleep disturbance: SQRT(2.6-sleep); Hearing Level: SQRT(HL+9); RHD-4: SQRT(RHD4+6.2); Hearing rating: SQRT(hearing item+0.9);. PCRHD3: Untransformed, but later transformed for some analyses; OM8-30 Impact: SQRT(impact14+2); Behaviour-5: SQRT(beh5+3.4); Speech & language: LN(speech+1.05) and parent Quality of life (PQoL): LN(PQOL+1.9).

@ *SES determinant marginal (p = 0.06).* ******* *Independent determinants potentially adjusted for were: centre, age, gender, SES and history.* ***§*** *A, B = main location in the two postulated parallel cascade for disease impact in conceptual diagram; A –hearing/language and B-physical health impact Figure 1Obstruction seasonality was not significant. We therefore took the more reliable re-fused 6-item form of URTI ('resp', both obstruction and infection items included) and tested its interaction with sine week 22, the delay at which maximal seasonal fluctuation is detected, with both age and SES. Both interactions were significant but not very strong; Age*sine week 22 p = 0.037, partial eta squared = 0.002, below what we would normally report; SES overall, so including effects of missing p = 0.021, with partial eta- squared 0.003 on the lower margin of small effects we would report. The directions are that older-age children show more seasonality at this delay; and the lower SES group shows more seasonality (p = 0.070 for this component, but this is highly marginal).*

Downstream facets: The downstream facets are more diverse and their effects have lower magnitudes and only RHD4 ($ή^2$ = 0.003) emerges with a qualifying degree of seasonality. The most sensitive single item is the overall hearing rating by parents. Again the difference is only of the order of the margin of error, but its seasonal peak occurs two weeks before (13.5 weeks) the value seen when the score from overall RHD as parent's concern about hearing (all-4 items used in OMQ14, giving peak severity centred on 15.5 week). Consistently, the peak severity for the 3 RHD questions is one week later than for RHD 4, at 16.5 week. The discarded item in the OMQ14 short form is 'asking to repeat things said' present in the OM8-30 form. The small difference between RHD measures needs to be replicated, but the longer delay of 2-5 weeks after HL severity peak for the communication questions shows that the time-accumulated effect of the hearing loss is central to RHD. The maximum week severity, located 13 weeks after starting time of the sinusoid with maximum coincidence, places the peak severity into summer, at the beginning of the July.

Other downstream variables do not show strong seasonalities and among them the behaviour facet severity shows a clearer cycle permitting a delay estimate than speech/language and PQoL. The behaviour problem peaks in June, with maximum at 23.5 and 24 weeks, but with negative cosine coefficient so the peak seasonality is related to summer. All these facets are kept in the model because it is meaningful to record the delay between groups, and then assess whether these delays are consistent with some kind of path model that enhances causal inference like SEM. Speech and language severity (very weak peak) appear to have a similar timing to the hearing problem, but it is hard to draw any firm conclusion in the absence of a strong seasonality effect. We need further studies, most preferably longitudinal, to measure severity of speech/language in relation to hearing and other severities scores thought of as OM sequelae. The behaviour facet has maximum severity at 23.5 and 24 weeks from the reference week at the beginning of January. The effect size for behaviour at the severity maximum is weak ($ή^2$ = 0.001) even though behaviour appears to give the most reliable seasonality effect of all downstream facets. The other impact measure, PQoL, has severity maximum a week after behaviour and its p-value is not significant and also the seasonality effect size is very weak (Table 1.3.).

4.1.3. Consistency of causality information from delays and from a structural equation model (SEM)

The SEM developed for the Eurotitis data is not the topic of this chapter and is not offered as new work in the context of this thesis. However it is mentioned briefly to show the possibility of achieving consistency between that form of description for understanding causal relationships, and the present one of causal explanation from seasonal delay between maximum severities of facets. SEMs usually involve the influences of latent variables (factors) in explaining severity of disease presentation in more than one facet area. In the SEMs, traditional risk factors such as: SES, age and history of disease take the role of driving initiators of the causal cascade.

The current stage of SEM modelling of the Eurotitis-2 database (Haggard et al., 2015) will be briefly summarised here to show the promise of SEMs and the present delay estimation techniques informing each other and producing estimates of causal sequence that agree. The current SEM offers two main pathways, each cascade sequence, which in turn offers a satisfactory way to explain the relationship between OM facets in the form that they are currently understood. A graphical representation (Figure 1.1.) is necessary to hold the elements all together in a single framework. A useful simplification is to think of the model as largely a hearing and speech/language pathway through development to QoL, and a largely separate pathway from URTI and RAOM through sleep disturbance to at least behaviour and other aspects of development to PQoL. Standard regression weights (SRW) between URTI and ESS, as well as between HL and RHD and RHD and PQoL are all > 0.3. Such strong relationships strengthen the inference that there is a causal cascade between the facets linked; URTI and HL and RHD and PQoL. Between Sleep and Behaviour, the SRW is less but substantial > 0.150. This second causal pathway appears weaker in the data but it takes in physical symptom scores and its influence on PQoL. Latent variable's loadings (SEM and Age) are high > 0.394 and 0.638 respectively. This suggested model has good fit with Root mean-square error of approximation (RMSEA) < 0.070 (Figure 1.1.).

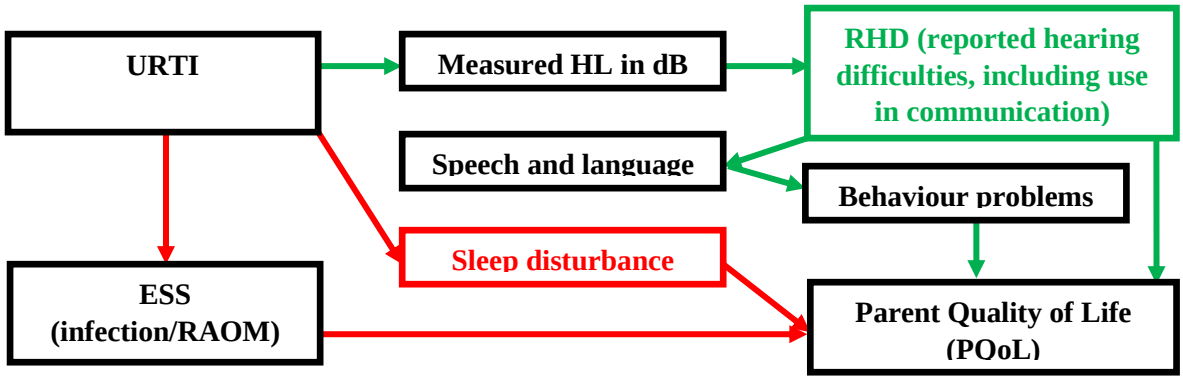

Figure 1.1. Conceptual diagram corresponding to a simplified form of the Eurotitis-2 SEM (Haggard et al., 2015)

Note:

'Upstream' in causal terms corresponds to part(s) of the model in the upper left, and downstream to lower right. Originating variables; age history and length of history have no plausible inputs in the Eurotitis-2 data. They might influence other variables at any cascade stage, but such inputs are strongest for RHD. Latent variables (i.e. for factor totals under-lying two or more observed variables) have been tried for various combinations of the variables shown, but those do not produce any better fit than having separate variables for each facet as here, so none are shown (conventionally, ellipses). The midstream mediator for Communication Pathway A is RHD (Green); for Physical Health pathway B it is sleep disturbance (Red).

4.2. Results of Study II

4.2.1. Sequence of presenting data

The means and SDs of the chief variables are presented for both complete and maximum cases, and comparison between SDESs are taken as the measure difference between them (Table 4.2.1.). The significance of contribution of independent variables for explaining variance in hearing measures and their PC total are presented in Table 4.2.2., while the effect sizes and their directions in Table 4.2.3. present magnitudes and directions of these influences on hearing measures of differences which are significant (i.e. almost all). The influence of the two objective hearing measures, alone and together on RHD with other independent variables is set out in Table 4.2.4 explaining RHD variation using only variables other than hearing measures, i.e. repeating values from Table 4.2.3. Next in Table 4.2.5. follows an analysis giving a more literal interpretation of 'discrepancy' using the standardised difference between HL and RHD. Finally we use mediation analysis (Figure 4.2.2.) to infer the underlying causal basis of then correlations between the two hearing measures and between them and RHD. Using substitution of HL by ACET I suggest a new method for imputing missing HLs by using ACET (Table 4.2.6.). For Table 4.2.4. and Table 4.2.6. especially, to avoid the text becoming very long and detailed, the *Nature* style of Methods and some details of Method as extensive footnotes is adopted here.

4.2.2. Completeness of data and properties of corresponding sub-samples

Balancing the generality of drawing maximally upon the data available against completeness of desired control requires analysis with two degrees of inclusiveness in analyses. The two main fields of Table 4.2.1. permit the comparison of samples with (a) broader inclusion, i.e. maximum cases on each variable separately, and (b) included within the former, the 1,400 complete identical cases having all three hearing variables. The difference in N of about one third is not trivial, so comparability is important for generalising certain analyses that are feasible only on the smaller group (1,400) of complete cases. The two panels of Table 4.2.1. (a & b for continuous and categorical variables respectively) compare the samples' descriptives. Comparing data rows 2 & 3 also shows the relationship between raw HL and transformed HL, for eliminating skew, in terms of their respective means and SDs. As often, raw is as good and more familiar for descriptives, but models were run with a square root transform to improve skew of residuals. General comparability is seen

in the general similarity of corresponding column and row entries between the first and second and fields. Given the two thirds overlap of samples, a sharper contrast is seen when comparing the 1,400 identical cases with the approximately 700 (data not shown) in the complementary set making up the maximum cases for each variable. In this sharper form of contrast, the two variables showing non-trivial effect of case selection on values (expressed as SDES) are length of history (large SDES, complete identical cases longer) and RHD itself (mild to moderate SDES, surprisingly with complete cases milder). The explanation for the effect of history is simple: in the complete identical-cased data the sample from the TARGET RCT is disproportionally represented because of the attention to data quality and completeness that is feasible within a publicly funded RCT. (Of 1,448 non-complete cases that any variable might draw on, only 38 came from this source (2.6%), whilst of 1,400 complete, 596 did, making 42.6%.) This leads to the expectation that with the more serious cases in the complete-data, any sub-sample effects involving RHD and history would be stronger within complete data, because of the higher mean levels of both. The basis for this expectation would be explicit selection for the trial, built on the back of a policy in the UK of treating AOM in primary care plus watchful waiting to keep cases out of hospital care unless suspect for continuing hearing problems.

In addition we examined missingness of categorical variables sex (1.67%) and SES (4.64%). Such variables have the compensating advantage that the missing category can be estimated in the analysis, allowing case retention, and its value does not need to be pre-imputed. For sex, the % that were girls differed by only 2.5% and the number of lower SES (maternal education) differed by only 4% between the two sample compositions. There is thus an adequate basis of similarity of inclusion for examining consistency of results on determination between the two sets of cases (Table 4.2.1.b.).

Table 4.2.1a. Descriptives on main variables used for overlapping maximum and complete cases, and the SD effect sizes (SDES) for complete-data cases versus cases with missing data (difference set)

	Maximum cases					**Identical cases**					**SDES between identical & partly missing cases**
	N	**Mean**	**SD**	**Skew (S.E)**	**Kurtosis (S.E)**	**N**	**Mean**	**SD**	**Skew (S.E)**	**Kurtosis (S.E)**	
RHD	2170	2.828	0.184	-0.038 (0.053)	-1.46 (0.105)	1400	2.806	0.183	0.117 (0.065)	-1.448 (0.131)	0.341
HL untran	2168	28.237	10.048	0.094 (0.053)	-0.424 (0.105)	1400	28.613	9.892	0.012 (0.065)	-0.520 (0.131)	0.106
HL transf	2168	6.044	0.842	-0.25 (0.053)	-0.277 (0.105)	1400	6.077	0.828	-0.301 (0.065)	-0.381 (0.131)	0.111
ACET	2398	1.332	1.149	0.099 (0.050)	-1.183 (0.100)	1400	1.373	1.094	0.171 (0.065)	-1.198 (0.131)	-0.086
Age	2886	62.334	15.469	0.414 (0.046)	-0.503 (0.091)	1400	63.475	14.348	0.412 (0.065)	-0.296 (0.131)	0.144
History	2886	1.986	0.260	0.049 (0.046)	4.194 (0.091)	1400	2.093	0.318	-0.959 (0.065)	4.080 (0.131)	0.873

***Note**: The two underlined ES values at the far right show that only Length of history and potentially RHD need consideration as differing enough between the two chief samples of differing completeness as to possibly make the complete-case analyses unrepresentative; for detailed implications see text.*

Table 4.2.1b. Outline descriptives for sex, SES and seasonality of maximum and complete cases

Panel b	Max cases		Identical cases	
	N	%, date of annual max	N	%, date of annual max
GENDER	2886		1400	
Female	1256	43.5	644	46.0
Missing	48	1.7	0	0.0
SES	2886		1400	
Manual	1448	50.2	759	54.2
Missing	134	4.6	90	6.4
Month max	2170	Late March	1400	Late March
Month max	2168	Late March		
Month max	2398	Late March		

***Note**: For compactness of table, and because Study I addressed seasonality in a more powerful way, the quantitative information underlying season of maximum severity is omitted. The parameter estimates obtained from sine and cosine fits to monthly numbers of cases differed little and in no instance by more than two SEs between the samples distinguished.*

4.2.3. Significance of determinants

In complex epidemiology, multivariable analysis is essential for control purposes, so *p*-values are appropriate in deciding what variables should be retained in models to produce probably more precise and less confounded analyses; within these justified models it is possible to see in a controlled way and to discuss effect sizes for the effects of chief interest, those for which the study will generally have been powered. Here, each variable can act both effect-of-interest and control for other effects, so p-values are unavoidable but the logical status, and the criteria to use, differ between these two roles. Comprehension is favoured by concentrating on a few comparisons of predicted direction and magnitude of effect. However, in the present systematic combination of 3 dependent variables plus their aggregate the principal component (PC) with 6 independent variables (assuming fitting sine and cosine as a seasonality pair), we needed some form of 'screening' of influences for reliability (Table 4.2.2.) before proceeding to address magnitude and direction (Table 4.2.3.) for interpretation. The majority of the 24 (or of the non-redundant 18 if disregarding the PCtotal) effects are very highly significant ($p < 0.0005$) as might be expected from the large sample size. We therefore imposed an initial retention threshold of $p = 0.10$ for all overall (main) effects in the model but only discuss effects of interest at the conventional $p = 0.05$ threshold before progressing to the more important magnitude criterion expressed in partial eta-squared.

The table of overall effects (Table 4.2.2.) can be summarised in four simple generalisations: (i) considerably more variance is explicable by the available determinants, both in tympanometry (coded as ACET) and in RHD than is explained in HL, the division not being an opposition between subjective report and objective measure; (ii) barring exceptions listed below (a-c), the trends for identical and maximum cases are very similar, (iii) sex is not significant, not relevant to within-diagnosis severity; (iv) effects of length of history are strong and similar enough to not make the slight difference in the mean of these between the two samples a source of artefact or instability.

Table 4.2.2 Reliability of effects: Centre, 5 determinants, & season as (sine, cosine) pair, for all three measures & PC (expressed as p-values for overall effects)

	N	Adj	Centre	Age	Sex	SES#	Diag*	History	Sine	Cos
		R-sq	*p*-val	*p*-val	*p*-val	*p*-val	*p*-val	*p*-val	*p*-val	*p*-val
Complete cases										
RHD	**1400**	**0.241**	0.000	*0.272*	0.119	0.001	0.000	0.000	*0.130*	0.008
HL	**1400**	**0.140**	0.000	0.000	0.653	0.001	0.000	0.000	0.000	0.798
ACET	**1400**	**0.235**	0.000	0.002	na, 0.278	0.016,0.346	0.002,0.002	0.002	0.005	0.207
PC (RHD HL ACET)	**1400**	**0.209**	0.000	0.000	0.861	0.000	0.000	0.000	0.000	0.138
Max cases										
RHD	**2170**	**0.211**	0.000	*0.009*	0.388	0.014	0.000	0.000	*0.008*	0.105
HL	**2168**	**0.157**	0.000	0.000	0.071	0.051	0.000	0.000	0.000	0.977
ACET	**2398, 2351**	**0.285**	0.000	0.001	0.009,0.075	0.067,0.507	0.001,0.001	0.001	0.001	0.568

Notes:

**The collapsed form of Diagnosis (3 levels) was used in these models to conserve df: Missing, RAOM and OME with superadded RAOM (combined), OME. In later tables the number of categories is re-expanded.*

#SES: The component of interest is Manual versus non-Manual (reference non-manual) but in this table the overall term including missing values is fitted.

Italic font: pairs of results between which the difference is expected on general power grounds.

The following transforms were used to improve the residual distributions: LN(22-RHD), SQRT(HL+9), LN(34-ACET), LN(PC+4.6). All ACET models were additionally bootstrapped for both maximum and identical cases, with 1,000 re-samplings and these are the quoted values. Uniquely for ACET, two values are provided to reflect possibly special effects of missingness on variability under bootstrapping, the first being for the overall effect and the second being for the component of interest.

4.2.4. Consideration of interaction terms

For this study we adopted a special approach to interaction terms, considering them chiefly as possible limitations to generalisation of the interpretations of overall effects. As an initial constraint we used the maximum cases samples but only proceed with first-order interaction on individual measures if a term had met a p = 0.02 criterion on the PC as dependent variable. We then used a p to retain of 0.02 for interactions with multi-level categorical variables, and applied the same 0.02 again to the individual component terms

before accepting a term as needing interpretation. Because some of the interaction related to missing values, the resulting picture was of very little interaction at all. For RHD as dependent variable, although three, two interactions met the joint criterion of $p = 0.02$ overall, only two met it in the component due to values other than missing and also had partial eta squared 0.003 or greater. These were the length of history*age ($p = 0.0017$; partial eta squared 0.003) and SES*age (lower versus higher SES component $p = 0.013$; partial eta squared 0.003). The magnitudes even of the significant interactions emerging from this restrictive screening are only on the margin of worthwhile consideration, and that summary conveys a satisfactory generality for the simpler overall (main-effect) models in Table 4.2.2. For HL the two interactions were stronger: diagnoses*age (overall $p = 0.00026$ and partial $\acute{\eta}^2 = 0.010$) and SES* length of history (overall $p = 0.000197$, partial $\acute{\eta}^2 = 0.008$). The directions of these are complicated to decode given the inverted log transform used for RHD and of the four above only two had partial eta squared for the interpretable component above the marginal 0.003, those for HL. For avoid repetition it is left to the Discussion. For both RHD and HL, length of history is more important in older children, where there would be more time for longer (and more easily reported) lengths of history to have developed. This ability to explain what is found assists the interactions in serving as some internal validation of data, rather than as informative substantive findings. The SES interaction for RHD expresses a lowered importance of length of history in lower SES (actually maternal education) is of possible relevance to policy or to the interpretation of the overall (main) SES effect, so is revisited in the Discussion. Possible explanations might be lesser awareness in respect of perceiving or interpreting signs and symptoms in younger children, or less precision in recalling and reporting them. As there is precedent, this finding could be used to justify modulating clinician questioning in respect of length of history according to maternal education, and possibly in adjusting formal scoring of history items in questionnaires.

4.2.5. Stability of determinants across inclusiveness of samples

The generally extreme significance levels restrict the amount of comment necessary on incompleteness of data (Table 4.2.1.). The three chief apparent differences with sample composition are: (a) for RHD, age is moderately significant in the sample with maximum number of cases but not in identical cases, though it is significant for other hearing measures; (b) similarly sine (of month), conveying late winter maximum severity, is significant in the larger maximum cases sample, but not in the smaller one of the identical complete cases; and

(c) cosine of month is significant only in identical cases. It should be recalled here that reference 0 month for sine and of monthly data is December. Importantly, such modest apparent contrasts in significance pattern will not themselves generally be significant. Points (a) and (b) would be consistent with the difference in power following directly from the respective Ns, so demand no special explanation. Point (c) requires some explanation; the apparent delay of maximum severity in RHD will be probed later more directly in other ways. In general, sine-cosine pairs should be examined as pairs without statistical thresholding or backwards deletion of terms. Seen thus, the two results on (b) and (c) agree in locating the maximum between March (the peak for positive sine) and June (peak for negative cosine). It is entirely plausible that the cases (or cases at particular centres) giving fuller data should be those with more emphasis on hearing problems and so more accumulation (with delay) of maximum reported hearing difficulties. Indeed, the small sample discrepancy in the means (Table 4.2.1.) showed identical complete cases as having worse HL and ACET. These facts are congruent with the identical cases sample, despite being smaller, being specifically the more powerful for the issue of delay in RHD maximum (negative cosine term), making the apparent difference in significance level consistent. This has an aspect related to centre differences too detailed to examine here, but these comparisons reveal no impediment to interpretation of magnitudes and directions or generalisation of findings, provided that the qualification is registered that they may be strongest in samples containing a high proportion of serious cases.

4.2.6. Magnitudes and directions of effect on complete identical cases

Table 4.2.3. documents the effects requiring substantive interpretation, and effect size comparison as cases are identical (only the complete 1,400 cases shown). They are all in the expected direction and largely consistent across the three hearing measures, although the magnitudes vary. On this index, values less than 0.003 are deemed too small to discuss, values above 0.03 large. In the following five comparisons of effect size, some require further discussion, although it is generally not feasible to put a *p*-value on the contrast being drawn, an issue revisited later in Table 4.2.5 (i) The SES effect (i.e. lower SES are worse affected) is large for RHD, small for the objective measures, despite some expectation that more educated parents would be more aware of hearing and communication problems with an unselected population not yet cases. (ii) Within the clinic population, severity declines with age, except for RHD. (iii) There are two possible dichotomous contrasts within 3-level

diagnosis, of which the one used here brings out conservatively the expected higher value of the hearing measures for pure OME than for collapsed category RAOM plus combined (OME +RAOM), despite broader impact being high for combined diagnosis. (iv) Length of history is moderately influential for the two objective measures, but is very strong for RHD and for the PC total of all three measures. (v) The seasonality columns show clearly the alignment of the objective measures nearer to a late winter maximum (sine of month positive), but nearer to a late spring maximum for RHD, and very weak for tympanometry quantified as ACET.

Table 4.2.3. Direction & magnitude of determinant effects (extension from table 2)

Variable & direction of effect on DV	**Sex**	**SES**	**Age**	**Diagnosis**	**History**	**Season (sine of month)**	**Season (cosine of month)**
RHD	NS	Manual higher*	NS	OME higher#	Long higher	NS	-ve: Late spring max
partial ϩ2		**0.006**		0.0019	**0.046**		**0.005**
HL	NS	Manual higher*	-ve	OME higher#	Long higher	+ve: Late winter max	NS
partial ϩ2		0.0017	**0.012**	**0.011**	**0.012**	**0.010**	
ACET	NS	Manual higher*	-ve	OME higher#	Long higher	+ve: Late winter max	NS
partial ϩ2		0.0006	***0.147***	**0.007**	**0.009**	0.0044	
Total (PC)	NS	Manual higher*	-ve	OME higher#	Long higher	+ve: Late winter max	NS
partial ϩ2		0.0032	**0.057**	**0.010**	**0.027**	**0.009**	

***Note**: Significant effects are in bold. *Low-SES families had 'worse' values on hearing measures; #Cases with OME diagnoses were worse. Only the 1,400 complete cases are used. For space reasons, centre is omitted as only a background adjuster, although highly significant. As the sine-cosine pair captures one effect (season) with a totally complementary pattern, there is only one truly non-significant effect in the usual sense apart from sex: age upon RHD.*

4.2.7. Discrepancy between objective and subjective hearing measures

Table 4.2.4. permits closer examination of the apparent distinct patterns of results for RHD in complete cases through one way of accessing the discrepancy of pattern between the measures. Sine (month) was fitted as an adjustment but is omitted from the table here and from the final model, as it was not significant and the concentration is on the relative delay

represented by cosine specifically for RHD. Diagnosis is also omitted, as the effects in the model are seen irrespective of diagnosis but it has been fitted as adjuster. Here in the four vertically separated fields, objective measures are additionally fitted as predictors of RHD alongside determinants (in sequence, HL, ACET, both, and neither; the last is a control analysis closely approximating the RHD entries in Table 4.2.3.). The first two columns simply show that the additionally fitted objective measures are contributing materially to explaining variance in RHD. The table gives the role of the objective measures in explaining RHD, but also by implicit comparison between the fields for two objective measures and neither of them, gives an impression of how this may through multi-collinearity, alter the pattern of influence form the other RHD determinants.

Table 4.2.4. Joint determination of RHD by objective hearing measures and other determinants: regression coefficients & SES

Term → **↓Extra** **Predictor**	**HL**	**ACET**	**Sex** **(Male worse)**	**SES** **(lower mat education worse)**	**Age**	**History**	**Cosine month**	**Adj RSq** **(Partial eta-sq, center)**
HL	**0.006**	X	0.015	0.022	0.001	**0.035**	-0.015	0.347
SE	0.0004	X	0.008	0.008	0.0003	0.005	0.006	X
Partial $\dot{\eta}^2$	0.135	X	0.002	0.005	0.005	0.034	0.005	*0.049*
ACET	X	**0.008**	0.017	0.023	**0.001**	**0.037**	-0.014	0.310
SE	X	0.001	0.008	0.009	0.0003	0.005	0.006	X
Partial $\dot{\eta}^2$	X	0.086	0.003	0.005	0.009	0.036	0.004	*0.032*
HL&ACET	**0.005**	**0.003**	0.016	0.022	**0.001**	**0.034**	-0.015	0.353
SE	0.001	0.001	0.008	0.008	0.0003	0.005	0.006	X
Partial $\dot{\eta}^2$	0.063	0.009	0.003	0.005	0.008	0.033	0.004	*0.042*
NEITHER*	X	X	0.013	**0.027**	0.0003	**0.043**	-0.016	0.246
SE	X	X	0.009	0.009	0.0003	0.005	0.007	X
Partial $\dot{\eta}^2$	X	X	0.002	0.006	0.0005	0.045	0.005	*0.046*

***Notes**:*

Regression coefficients are unstandardized so it is necessary to inspect partial eta squared to compare magnitudes (importance) along rows. However, as HL and ACET are on the same dBHL (or equivalent) scale, the magnitudes of these two coefficients can be compared, in the context of other determinants; this is comparison between the first two data columns, both between the first two fields and within the third. Comparing estimate/SE for each separately, ACET is the less reliable, (only 8 times its SE compared to 13 times for HL). However its best estimate shows a larger contribution (0.008 in Field 2) in terms of coefficient

magnitude per dB unit of ACET to RHD than HL does per unit in dB (only 0.006 in Field 1). This ACET contribution gets reduced by multi-collinearity in the GLM (which takes reliability into account) in Field 3. Thus relative coefficients or partial eta squared when both are fitted in GLM do not give a full picture of relative weight. This comparison triggered the mediation analysis.

For space reasons, diagnosis, though fitted, is not shown, as not of chief interest here. The objective hearing variables (despite raising model df) explained much more variance, making it justifiable and supportable without instability to add one further model df (distinguishing all 3 diagnosis categories, by separating 'combined')

Comparisons within the last column show the proportion of total variance (Rsq in roman font) contributed by Centre (partial eta-squared, italic font); it ranges between about one tenth and one fifth of the part of the total that is explicable. These are large centre differences but sufficiently small to not overshadow effects of interest. Contrasts between entries in the last two fields reflects collinearity of objective hearing measure with other effects; the smallness of such differences conditional on presence of the objective measures supports the principle that these are two largely independent (additive) classes of effect.

Table 4.2.4. gives several important messages: a) objective measures contribute strongly to RHD, HL doing so more than ACET; b) when neither of the objective variables are fitted, the determinants still explain nearly one quarter of the variance in RHD (Rsq = 0.246); c) HL and ACET are highly collinear so their combination in PC whilst adding reliability does not add much predictive power from that present in the stronger single one alone (i.e. HL alone); d) contribution of centre is large but not overriding, still below 5% for the sample of complete-data cases. The variance explained by centre dropped when tympanometry is fitted, suggested centre differences in tympanometry calibration of different meeting criteria for OME/RAOM or referral policy. Boys appear to have worse RHD for their measured HL, but as this is the only significant sex effect obtained, it requires caution and replication. Interestingly, in all four models (Table 4.2.4.) the estimate for cosine exceeds twice its SE, meaning that the delay to RHD severity maximum in the annual cycle survives control for the objective measures which express a late winter seasonality. The relativity issue raised by this expression, i.e. delay relative to the delay seen for HL, is addressed in Discussion, Most of the effects are modest, but only two (sex) are very small, 0.002 or less in partial eta-squared. The effect of history is reduced by fitting the objective measures, but remains strong. The significant effects of determinants on RHD are thus not undermined by collinearity of hearing measures but the magnitude of these effects could be.

The form of analysis in Table 4.2.5. (adjustment for objective hearing measures) is equivalent to fitting the objective term(s) in a pre-regression then modelling with all the other determinants together. It does not give a direct index of discrepancy between measures. We therefore also used a different approach, Z-diff, the difference between standard forms of the two scores as a direct and explicit measure of discrepancy. The standardised difference scores were made the dependent variable in a model of the same form as in Table 4.2.5. The two approaches gave overall rather similar results except that centre effects are more important with Z-diff, as is age; history, on the other hand, remains more important with the two equivalent methods, which fit HL as an independent variable (see Table 4.2.5 footnotes).

Table 4.2.5. Two approaches to discrepancy testing HL and RHD using a) covariance apportionment (covariate or residuals and b) direct differencing (Z-diff)

Term → **↓Dependent Variable**	**HL**	**SEG** **(lower education level)**	**Age**	**History**	**Cosine month**	**Overall** **Adj RSq** ***(Partial eta-sq, for center)***
RHD with HL in model	0.006	0.016	0.001	0.035	-0.011	0.312
SE	0.0004	0.008	0.0003	0.005	0.006	X
Partial etasq	0.139	0.003	0.008	0.027	0.002	*(0.096)*
Residuals from fitting HL*	X	0.016	0.001	0.035	-0.011	0.209
SE	X	0.008	0.0003	0.005	0.006	X
Partial etasq	X	0.003	0.008	**0.027**	0.002	*(0.099)*
Zdiff #	X	0.065	0.010	0.112	-0.052	0.176
SE	X	0.052	0.002	0.034	0.039	X
Partial etasq	X	0.001	**0.021**	0.006	0.001	*(0.133)*

***Notes**:*

SES, age and diagnosis are omitted from analyses as non-significant in the final model.

The two fields in the upper half of the table simply demonstrate the formal equivalence of fitting the HL term in the model to taking the residual from a preliminary regression model with only HL as independent variable, and then modelling that residual. The RSq difference (0.312 to 0.219) is considerable but simply reflects the known large contribution of HL that has been removed by the preliminary step. In contrast, taking the standardised difference as a direct index of discrepancy produces a somewhat different pattern of results, with Rsq lowered to 0.176 because in general relative error is increased in a difference but reduced in a sum.

For SES and strength of summer seasonal maximum severity, the differences are too small in absolute terms to be usefully compared. However for age and history, there is a marked crossover (bold) between the methods. Z diff reflects age more strongly but the equivalent covariance-based methods reflect history more strongly. Centre differences are also a little stronger with the direct Z-diff method. For interpretation see Discussion.

4.2.8. Mediation analyses

Even with HL in the model, the addition of ACET improved Rsq in RHD in Table 4.2.4., but the effect size of HL is seven times stronger than that for ACET. In order to estimate what part of the contribution of ACET is not mediated by HL, we used mediation analysis with the Process software (Hayes, 2013). The strength of each of the three paths is estimated by regression coefficients, in effect partialling for the other paths. Thus a trivariate context is used for comparing the effect of ACET on RHD mediated by HL, and also the direct relationship between ACET and RHD. The total effect of ACET on RHD has a standardised coefficient of 0.360 ($p < .001$, $t = 14.436$). More interestingly the separation of direct from indirect effect by this software gave a direct effect of 0.203 ($p < .001$, $t = 6.400$). Using re-sampling methods (1,000 bootstrapped samples), we estimated variability hence significance for the indirect effect at 0.157 ($p < .001$; no t-value given for this procedure, but 95% CI [0.117, 0.200]). The indirect effect is essentially the product of multiplying the regression coefficients (0.247*0.637) and analogous to serial resistances in an electrical network; it represents how much of the variance in the first variable 'gets through' to control the third in the indirect path (Figure 4.2.2.). Thus, ACET is an independent predictor of RHD, and this explains why the model with both hearing measures explains more of the RHD variance than the one with only HL. Significance of the indirect path makes it a better model than the one with direct paths only, and in this better model, the ratio of indirect to direct contributions for ACET (value 1.29) is more informative and fairer to ACET than the sevenfold ratio estimate obtained from the relative weights in the GLM with no mediation path.

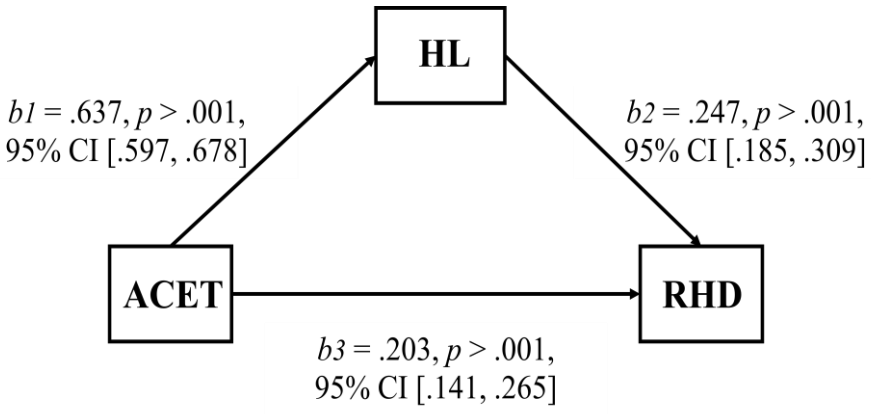

Figure 4.2.2. Standardised regression coefficients as the strength of path between three measures

4.2.9. ACET as replacement for missing HL data

The raw correlation of 0.598 and the similarity of determinants suggest that substituting ACET-based values for missing HL-values could be widely use in clinical practice. In order to present how much of HL data could be safely imputed in this way thus we randomly replaced HL with ACET in 5%, 10%, 15% etc. up to 35% of cases in the complete data sample of 1,400. This work (Milovanovic et al., accepted) showed that the correlation of the imputed hybrid with the true original HL on the same cases only fell below 0.92 (not shown) when the randomly missing rate went above 25%, and this was generally encouraging. This correlation coefficient with RHD (bottom row) is the only available validation paradigm; the first and 2nd column show the single values for HL and ACET are reasonably close. The issues raised by the use of the hybrid are more fully addressed in the footnotes to Table 4.2.6. Its main points are that even with a regaining of case numbers approaching +50% as given by the Eurotitis-2 missingness structure for HL, the substitution of ACET based on tympanometry leaves the shape of the distribution little changed and places the correlation value exactly where it is expected -intermediate between the value seen on complete-data cases for HL and ACET alone.

Table 4.2.6. Adequacy of a substitution for missing HLs with the ACET value: distributions and correlation of the two objective hearing measures (with SEs)

On Complete-case data			Maximising case-gain in Eurotitis-2 for the pattern of missingness in the data		
Measurement(s)	HL	ACET	0%HL+100%ACET	0%ACET +100%HL	Hybrid: HL or ACET where HL is missing
Number of cases	1400	1400	1716 ^	1793^	2094^ ^
# Skew of distribution (SE)	0.012 (0.065)	-0.171* (0.065)	-0.180 (0.059)	0.006 (0.058)	-0.046 (0.053)
# Kurtosis of distribution (SE)	-0.520 (0.131)	-1.198 (0.131)	-1.173 (0.118)	-0.568 (0.116)	-0.447 (0.107)
r with 'true' (100%) HL	1.000	0.598	----	----	----
@ Correlation coefficient with RHD-4 (SE)	**0.378 (0.023)**	**0.300 (0.024)**	0.223 (0.023)	0.360 (0.021)&	0.330 (0.019)&

Notes*:*

 # Skew is the left-to-right asymmetry or leaning of the obtained distribution, versus its having a strong central tendency. Positive skew (long tail to the right, at higher, more positive values) is naturally the more common direction of skew for pathological states -- whether in a population or a clinical sample – there being few extremely severe cases, but many milder ones having passed some threshold of concern. Positive skew is usually highly remediable with some simple transform like logarithm or square root, and transforming has been done in the past chiefly with the aim of being able to take resulting p-values literally. Transforming can also favouring power by contrasting two more equally populated ranges of the variable when it makes the mean and the median more nearly coincide. Kurtosis is the peaked-ness versus flatness of the central part of the distribution. This is not easily remediable by transforming without challenging the basis of the measurement and if extreme, kurtosis potentially requires results to be bootstrapped. The kurtosis, not skew is the problem here, mostly in respect of the bimodality of ACET which is a special instance of negative kurtosis (extreme more than flat top).

^ Note that columns 1 & 4 are to be contrasted here with columns 2 & 3. The two columns before the last one are on maximum cases, so cannot exactly replicate the corresponding complete-cases data in the first two columns (e.g. 100% HL should = HL, column 1 with column 4). However, for both kurtosis and skew the given pairings of columns (data columns 1 with 4, 2 with 3) show that the distributions are closely similar. Such statements of similarity depend on the standard errors (SE), where a difference of 2 SE becomes marginally reliable. So for example the difference for skew between the last two columns is 0.121. It is gratifying that it appears actually to favour the proposed hybrid in the last column, but no strong claim can be made on this because the difference is only just over one SE (0.116).

^^ *The hybrid in the last column is the chief interest, showing the consequences of applying to the Eurotitis-2 data the rule of using ACET in actual cases where HL is missing. For the last 3 columns, correlation with true HL on the same cases is not available, as the case gain is based on using cases which do not have all variables. This case gain over the 1400 complete-data cases is + 28.3% or +28.1%, in the two columns previous-to-last, when we consider use of either variable but only one of them and this would correspond to a loose specification of the a priori research strategy, and reporting as separate analyses probably with correction for multiple testing, depending on the exact hypothesis. The case gain further increases to + 49.6% on accepting the hybrid HL/ACET in the last column. This has the conceptual advantage of a single variable defined by the substitution rule, allowing for example avoidance of the further power penalty in adjustment for multiple testing in those circumstances where that would be considered obligatory.*

@ It is clear that the hybrid preserves the distribution of HL well, and that in both types of correlation shown, with true HL and with RHD, it shows the modest drop (by 0.078 and by 0.030) expected from the absolute correlation obtained with true HL, because HL and ACET are imperfectly correlated. It is likely that the rather low correlation (0.223) of ACET with RHD in the 1796 cases, of whom 396 have ACET but not HL, is due to these cases with no HL measure available being milder and so not deemed clinically to require HLs; hence the correlation would be lowered by having a more limited range of RHD.

& The standard error around the estimate from the hybrid HL/ACET as produced by the missingness pattern in Eurotitis-2 is actually slightly narrower than on the maximum cases sample having complete data for HL (last two table entries). This means that the admixture of ACET data having a cruder quality is more than offset, in error terms, by the greatly increased sample size from accepting ACET where HL is missing. The correlation value may be slightly lower but we can have more confidence in its precision.

4.3. Results of Study III

4.3.1. Criterion validity

Principal component and subsequent factor analyses of OMQ14 items defined three factors: ESS (ear symptom score), RHD and general developmental impact, called 'impact' here for short. The different basis of using items to define facet scores (OM8-30) and factor scores (OMQ14) means that the inter-score correlations must differ between the two instruments. Table 4.3.1. shows that on the set of cases with data available, the inter-factor correlations for the two instruments do indeed differ in the way expected from their derivation. Only for the OM8-30 correlation between impact and RHD (bold in table) is the correlation high enough to suggest that the criterion validity correlation between corresponding scores in OMQ-14 and OM8-30 could be threatened by the rotated factor derivation with OMQ14 forcing it to be zero.

Table 4.3.1. Inter-factor correlations (and Ns) for OM8-30 and OMQ-14

Correlation pair→ Instrument ↓	ESS, RHD	ESS, IMPACT	RHD, IMPACT
OM8-30 (Discrete-item-set PCs)	0.167 (2190)	0.165 (2154)	**0.323** (2149)
OMQ14 (Varimax factor scores)	0.019 (2183)	0.013 (2183)	-0.003 (2183)

***Note**: the inter-factor correlations for OMQ14 are not exactly zero as the cases available with the full 'other' data are not exactly the same as the cases for the original derivation.*

Criterion validities for the OMQ14 scores are given by Pearson linear correlation coefficients between the OMQ14 factor scores and the OMQ8-30 equivalents (Table 4.3.2.).

Table 4.3.2. Criterion validity correlations between corresponding OMQ14 factor scores and OM8-30 facets scores and also for total PC

Variable	ESS	RHD	General impact	totalPC
Pearson r (N)	.966 (2183)	.951 (2191)	.903 (2149)	.907 (2180)

The high correlation coefficients and large sample size avoid any issues about absolute significance of these correlation coefficients. The small difference in item content between

instruments OM8-30 and OMQ14, on ESS and RHD (only two ESS items discarded and only one RHD item) is sufficient to explain the especially high correlations between the corresponding forms of these measures. The third factor, general impact, is more heterogeneous, as shown by its lower item loadings. However the factors explained similar amounts of the total variance. Given the difference between the factors, the PC as their weighted total is by definition a heterogeneous construct and half the items in it come from the more heterogeneous impact domain. The criterion validity correlations still exceed 0.90 as required, but they are reliably lower. The way in which the measures are defined for heterogeneous constructs is probably responsible. The difference between the higher two and the lower two has to be explained because it is not just a matter of number of items involved in the facet (or high-loading in OMQ-14), which would give a difference in magnitude in the opposite direction. Using Fisher's Z technique for testing the difference between either of the first two correlations and either of the last two gives all $p < 0.001$ (highly significant.) The underlying explanation for the generally good correlations is probably that discarding some items (7 from 14 in OM8-30 and 18 from 32 in the PC) reduces the heterogeneity present in OM8-30 and encourages slightly more homogeneity in the factor. The difference is not significant between the last two (impactQ14 and totalPCQ14) consistent with the above suggestion explaining that the lowered correlation for the PC score results from it being driven half by the inhomogeneous impact items forming half the number of items present in the questionnaire.

4.3.2. Internal consistency

Results on internal consistency are presented in the Table 4.3.3. using Cronbach alpha. These are satisfactory (> 0.70) for all scores on the sample of 2,865 cases. Only for general impact did the alpha-value fall to below the usual standard of 0.70. The reason is the known heterogeneity of this general impact factor, with the items having been drawn from three separate original facets (speech/language, behaviour and parent quality of life) thus touching more than one impact problem. Cronbach's alpha does not strictly apply for factor scores, so to create an equivalent for comparison we created a special facet score, not normally used, from the small number of items that load highly on the factor score. Such internal consistency values serve for comparison with other measures and instruments.

Table 4.3.3. Internal consistency for the OMQ14 facets scores structures

Measure → (k @) Index ↓	totalPCQ14 (14)	ESSQ14 (4)	impactQ14 (7)	RHDQ14 (3)
Cronbach's alpha $	0.762	# 0.802	0.668	0.847
% Variance explained*	52.148	18.395	16.914	16.838
Mean item factor loading	0.487	0.756	0.560	0.852
Lowest factor loading $	0.304	# 0.435	0.454	0.832

Note:

k @: k in brackets is the number of high loading items for each factor; # :the global health item is a good QoL predictor and is assigned to ESSQ14; $:lowest among only the high-loading items that would enter a discrete item facet score, ignoring low-loading items that load highly on the other factors. Some low loadings are expected in a PC.

4.3.3. Determinant models for OMQ-14 scores

General linear models for PC total and the three factor scores were run on all 2,865 cases using 6 variables; age, gender, history, diagnoses, SES, season (sine/cosine pair by month, for economy of effort). Partial eta squared ($ή^2$) of determinants at 0.1 significance in the overall (main-) effects model without interactions are shown in Table 4.3.5. Variables not significant at $p < 0.1$ are back deleted and p-values in the main-effects model are shown in Table 4.3.4.

Table 4.3.4. P-values for totalPCQ14 and the three constituent factors in the main-effects model without interactions

	totalPCQ14	**ESSQ14**	**RHDQ14**	**impactQ14**
Adj Rsq	***0.182***	***0.168***	***0.276***	***0.109***
SES	0.000,0.000	0.024,0.007	0.072,0.506	0.000,0.000
Age	0.000	0.244	0.000	0.000
Sex	0.578	0.137	0.994	0.006,0.002
History	0.000	0.107	0.000	0.000
Sine	0.024	0.001	0.140	0.702
Cosine	0.074	0.291	0.706	0.122
Diagnosis	0.000,0.000,0.000,0.025	0.000,0.000,0.007,0.000	0.000,0.445,0.000,0.000	0.067,0.037,0.089,0.014

Note:

Model with 2,865 cases (main-effects model); terms significant at p > 0.1 are back deleted and the p-values presented are from the back-deleted model. In this and the following models form this study, the residuals from the models were all sufficiently normal to not require transforming. All SES p-values are given for missing and non-manual versus manual (reference non-manual). For SEX male versus female (female reference). For Diagnoses [4 levels: missing, combined (RAOM+OME), OME and RAOM]: missing versus RAOM, Combined versus RAOM and OME versus RAOM (RAOM reference).

Table 4.3.5. Effect sizes from main-effect models on 2,865 cases (i.e. without interaction terms)

	totalPCQ14	**partial $\dot{\eta}^2$**	**ESSQ14**	**partial $\dot{\eta}^2$**	**RHDQ14**	**partial $\dot{\eta}^2$**	**ImpactQ14**	**partial $\dot{\eta}^2$**
SES	Manual worse	**0.008**	Manual worse	0.0026	Manual worse	0.0002	Manual worse	**0.007**
Age	-ve: Older better	**0.012**	NS		+ve: Older worse	**0.011**	-ve: Older better	**0.059**
Sex	NS		NS#		NS		Males worse	0.0035
History	+ve: Long worse	**0.023**	NS		+ve: Long worse	**0.019**	+ve: Long worse	**0.009**
Sine	+ve: Late winter max	0.0018	+ve: Late winter max	0.0042	NS		NS	
Cosine	-ve: Late spring max	0.0011	NS		NS		NS	
Diagnosis Missing	Better	**0.006**	Better	**0.029**	Better	0.0002	Worse	0.0015
Diagnosis Combined	Worse	**0.010**	Worse	0.0025	Worse	**0.010**	Worse	0.0010
Diagnosis OME	Better	0.0018	Better	**0.068**	Worse	**0.015**	Worse	0.0021

***Note**:*

Values are after back deleted non-significant variables (p > 0.1) from the model in the Table 4.3.4. SEX is not significant overall and is back deleted from the finale model (marginal at p = 0.055 & partial $\dot{\eta}^2$=0.0013). Entries are partial $\dot{\eta}$2 and direction sign for effects of main variables are given.

Age is significant in all models except for ESSQ14. We do not in general find that factors score differ in respect of gender, except that boys have higher impactQ14 scores than girls, as would be expected from the general developmental delay literature on male susceptibility , but at partial $\acute{\eta}^2 = .0035$, this effect is very small. Children with D1 diagnoses (combined) are 'worse' (i.e. compared to pure RAOM) than children with other diagnoses on all the factor scores. They are particularly so for RHD, which seems to carry most of the similar effect on PCtotal in terms of the amount of variance explained in RHDQ14 versus what is explained in other factors. The RHD results are directly comparable with those in Study II. Children with a longer disease history have more severe disease profiles except for the ESSQ14 score, consistent with the distinction acute/chronic and children with longer history generally have more severe symptoms and impacts. Children with less educated mothers (marker of lower SES) are worse affected in all dimensions of the disease profile: RHDQ14, ESSQ14, impactQ14 and totalPCQ14. All of the investigated independent variables have some demonstrable and explicable influence on disease profile (pattern across factors) and also on overall disease consequences, although for gender this is minimal. Age, gender and history do not influence ESSQ14 score, but the severity of this factor is the one which varies most strongly through the year, being worse during late winter (as summarised by fitting sine of month). The effect of season is relatively weak but present and not significant for other factors, as discussed in more detail in Study I.

For the model with interactions (Table 4.3.6. and Table 4.3.7.) the interactions to be tested were first constrained by seeking positive findings on the more powerful OM8-30 data. All of the three shown as significant in the preliminary analyses on OM8-30 measures, were strong enough on OMQ-14 measures to mention and to include in an alternative best model. For PC total these are: Age*History, SES*History and Age*SES (Table 4.3.6. and Table 4.3.7.).

Table 4.3.6. Interaction terms from models with interactions, p-values and adjusted Rsq for totalPCQ14, ESSQ14, RHDQ14 and impactQ14

	totalPCQ14	**ESSQ14**	**RHDQ14**	**impactQ14**
Adj Rsq	***0.193***	***0.17***	***0.28***	***0.112***
SES*Age	0.012,0.052	0.996,0.966	0.012,0.008	0.126,0.598
SES*History	0.000,0.001	0.008,0.008	0.001,0.052	0.197,0.189
Age*History	0.000	0.367	0.071	0.001

***Note**:*

Interaction terms from models with Interactions (2,865 cases). When main effects: age, SEX and SES were not significant they are necessarily retained for interaction testing. Interactions significant at $p \leq 0.02$ are kept in the model, and adjusted Rsq values are taken from the final model containing all interactions significant at $p \leq 0.02$. Where interactions were not significant at $p \leq -.02$, the p-values presented are taken from the model before variables are back deleted.

Table 4.3.7. Interactions, from main-effects model with interactions added, $\acute{\eta}^2$ and direction of effects

	totalPCQ 14	**partial $\acute{\eta}^2$**	**ESSQ14**	**partial $\acute{\eta}^2$**	**RHDQ14**	**partial $\acute{\eta}^2$**	**impactQ 14**	**partial $\acute{\eta}^2$**
SES*Age	synergism	0.0013	NS		synergism	0.0025	NS	
SES*History	synergism	0.0041	synergism	0.0024	synergism	0.0013	NS	
Age*History	-ve	0.0044	NS		NS		-ve	0.0036

***Note**:*

*Values are obtained from adding interactions to the final back deleted model (Table 4.3.6.). Effect sizes with partial $\acute{\eta}^2 > 0.003$ are: SES*History for totalPCQ14 and AGE*History for totalPCQ14 and impactQ14. The direction of the interaction SES*History is positive (synergism) but AGE*History is negative (antagonism) for both factors (totalPCQ14 and impactQ14).*

The caution about these interactions is that the model with interactions should be required to explain more variance in the dependent variables after adjustment for the *df* if the interactions are to demand an attempt at explanation. However the *df*-adjustment is fairly minimal on a large sample, and not heavily weighted to theoretical parsimony, i.e. not towards explaining a lot of variance with few variables. Giving most importance to PC as total, two of these three interactions: Age*History and SES*History (for each, partial $\acute{\eta}^2$ =

0.004) are stronger than the third SES*age, which can be disregarded as trivially small. The SES*History interaction is in the expected sensible direction; children *both* from *low SES* families (low maternal education) *and* with *longer history* of disease are disproportionately worse in totalPCQ14 than expected from just low SES and long history absolute in ESSQ14 and RHDQ14 scores. This is one of relatively few convincing demonstrations on a large sample of the often suggested synergistic interaction of risk factors for consequences of OM. Younger children with relatively long histories are worse in impactQ14, and in the totalPCQ14 score to which it contributes, than would be expected considering the summated effect of the two determinants alone.

In the 1,866 cases (complete hearing measures) out of the total 2,865 with questionnaire data where both objective measures were available, we next added HL and ACET to the main-effects model. The results for each added independent variable are always in the expected direction, worse hearing giving worse outcome measure, but with HL generally being much stronger because of its more continuous distribution and wider effective range; the exception to this is seen with ESSQ14 where the presence of fluid ACET as represented in ACET is expected and relatively stronger. Some of the variables become more significant after this fitting of the hearing measures due to the higher percentage of variance explained. The effect of the hearing measures on Rsq and the individual variable p-values are presented in Tables 4.3.8. and 4.3.9. For all factors in the disease profile, hearing status thus affects severities and overall effects of disease in the expected direction but in general developmental impact this is marginal for HL and not significant for ACET. For totalPCQ14, current HL is a stronger influence even than history, but not by a large margin.

Table 4.3.8. P-values of hearing measures and adjusted Rsq in 1,866 complete cases with HL and ACET

	totalPCQ14	**ESSQ14**	**RHDQ14**	**impactQ14**
Adj Rsq	***0.275***	***0.188***	***0.409***	***0.106***
HL	0.000	0.365	0.000	0.052
ACET	0.003	0.002	0.002	0.706

Note*:*

Model with identical cases (1,866). Rsq values are taken from the model with both hearing measures in: HL and ACET as well as the determinants tabulated previously. p = 0.000 to 3 decimal places means p < 0.0005

Table 4.3.9. Main effect model in 1,866 cases with hearing measures in, $\dot{\eta}^2$ and direction of effect

	totalPCQ14	**partial $\dot{\eta}^2$**	**ESSQ14**	**partial $\dot{\eta}^2$**	**RHDQ14**	**partial $\dot{\eta}^2$**	**impactQ14**	**partial $\dot{\eta}^2$**
HL	+ve: high worse	**0.029**	NS		+ve: high worse	**0.075**	+ve: high worse	0.0020
ACET	+ve: high worse	0.0049	+ve: high worse	**0.005**	+ve: high worse	**0.005**	NS	

Note*:*

The effect size of HL is strong for totalPCQ14 and RHDQ14 and weak for impactQ14; ACET effect is moderate for ESSQ14 and RHDQ14 but weak for totalPCQ14 and non-significant for impactQ14. The final model before adding ACET and HL is from 1,866 cases, but is only slightly different from the final model with maximum cases.

4.3.4. Low SES

The main-effects model in the sample with identical cases 1,866 did not differ materially from the corresponding model with questionnaire data in the sample with maximum cases (i.e. from the model on which it is based and shares 1866/2183 = 85.5% of cases. Overall the partial $\acute{\eta}^2$ values (Table 4.3.10.) are slightly stronger in the 1,866, suggesting higher data quality in the more complete data or possibly a more symmetrical distribution of severity (there being more clinical incentive to acquire hearing measures in more severe cases). Because the measurement is normalised, with variability providing the measurement unit, it is in practice hard to separate these two classes of explanation. As one exception to representativeness, the effect of SES in the 1,866 is considerably stronger, about double, what it is in the maximum cases. This finding does not alter the meaning of the influence but it is unlikely to result only from more power in the larger sample, as the other increases in partial $\acute{\eta}^2$ are much smaller increases as well as the smaller sample being more powerful for some effects. It is possible that more of the cases with complete data come from centres that happen also to have steeper socio-economic gradient. This issue is touched upon in the appendix on the breakdown of data from centres in the Balkans (Appendix III).

As the ACET score is used here only as an independent variable, there has been no need for bootstrapping and the measurement assumptions of the GLM are met. The residuals for the main effect model have close to a normal distribution (Table 4.3.11.) with only two slight exceptions (in bold) needing special consideration such as transformation and bootstrapping respectively: the positive skew for impactQ14 and the negative kurtosis (flat top) for ESSQ14.

Table 4.3.10. Effect sizes of independent variables in 1,866 complete cases with hearing measures

	totalPCQ 14	**partial $\dot{\eta}^2$**	**ESSQ14**	**partial $\dot{\eta}^2$**	**RHDQ14**	**partial $\dot{\eta}^2$**	**impactQ 14**	**partial $\dot{\eta}^2$**
SES	Manual worse	**0.016**	Manual worse	0.0047	Manual worse	0.0020	Manual worse	**0.011**
Age	-ve: Older better	**0.010**	NS		+ve: Older worse	**0.005**	-ve: Older better	**0.042**
Sex	NS		Males better	0.0022	NS		Males worse	0.0030
History	+ve: Long worse	**0.031**	NS		+ve: Long worse	**0.030**	+ve: Long worse	**0.009**
Sine	+ve: Late winter max	0.0019	+ve: Late winter max	0.0022	NS		NS	
Cosine	-ve: Late spring max	**0.005**	-ve: Late spring max	0.0030	-ve: Late spring max	**0.0015**	-ve: Late spring max	0.0019
Diagnosis Missing	Better	**0.009**	Better	**0.039**	Better	0.0012	NS#	
Diagnosis Combined	Worse	**0.012**	Worse	0.0040	Worse	**0.014**	NS	
Diagnosis OME	Better	0.0032	Better	**0.077**	Worse	**0.015**	NS	

***Notes**:*

Direction of effect and $\dot{\eta}^2$ in model with all variables and hearing measure. Diagnoses NS for impactQ14. Diagnoses (4-level); missing, combined, OME, RAOM/versus RAOM. Sex: males versus females (reference female). SES: non-manual versus manual (non-manual reference). Season: cosine effect for RHD significant but effect below the magnitude to justify interpretation.

NS#: For impactQ14 diagnosis is not significant overall and so dropped in the final model, and so non-significant in this table. However the missing category is marginal at p = 0.073 with etasq 0.0017 in the model just before diagnosis drops out.

Table 4.3.11. Residual distributions with the final backwards-deleted model on 1,866 cases

OMQ14 Variable (& standard error)	totalPCQ14	ESSQ14	RHDQ14	ImpactQ14
Skew (SE 0.057)	0.204	.171	0.178	**0.449**
Kurtosis (SE 0.113)	-0.164	**-0. 573**	-0.243	-0.222
KSP#	.0063	0.014	0.304	< 0.0005

***Note**:*

Standard errors of the distribution parameters for this N are in brackets. # KSP is the p-value for the Kolmogorov-Smirnov non-specific test for overall departure from normality. On this sample size, ≥ 0.01 is satisfactory, which only impact fails so generally requiring transformation.

The natural positive skew of severity for ImpactQ14 (fewer extreme cases) is the underlying cause of the first exception. Greater KSP (p value for the Kolmogorov-Smirnov non-specific test) means that the difference from normality is less significant. When HL (or ACET or both) were added to the predictors in the model the distributions of RHDQ14 and totalPCQ14 became extremely satisfactory showing that the skew is natural, not a property of RHD as a measure. For the others, adding HL or ACET did not provide sufficiently strong an entry to the regression to achieve this re-distribution (e.g. for impactQ14, the *p*-value for HL is only 0.059). Generally, attention to distribution shape will be required with the Impact scores and potentially to the shape for ESS also. The distributions of residuals for main effects models are discussed in more detail in Appendix II.

4.4. Results of Study IV

4.4.1. Underlying GLM without centre adjustment

In the logistic regression model for HL prediction without centre adjustment, the hearing rating question and ACET had strong effects (partial $\acute{\eta}^2$ = .358 and .069 respectively), while effect of season, early spring, was modest (partial $\acute{\eta}^2$ = .004). The measure of the goodness of the model was good (Rsq = 0.469, adjusted value) (Table 4.4.1.). The seasonality entered the model significantly when hearing rating item is in the model, reflecting maximum severity in early spring, late winter (Study I). This means that the rating item, because like HL itself it has an annual cycle of severity but one offset by more than two months, has to be given a different scale value for predicting HL at differing times of year. In the model with both types of hearing questions (score of 3 communication items also in) the interaction between ACET and communication and also seasonality were significant, reflecting the fact that the best way to combine them additively takes account not only of their complementary roles at an average low and average high severity but the early summer maximum in report of communication items. The full model with both type of items and interaction and cosine was best (adjusted Rsq = 0.471); however although interaction and seasonality terms are significant, the difference that they make to Rsq is largely under 0.005. As with partial $\acute{\eta}^2$, this order of difference is marginal.

4.4.2. Underlying GLM with centre adjusted

The models with centre adjusted are all better than the unadjusted versions for the GLM (and indeed later for the logistics). This assists comparison with the version of logistics decided to be appropriate. The predictive effect of the hearing rating item is still high (though slightly less than in the unadjusted model) while effect of interaction ACET * PC RHD3 communication items is modest. Communication items still have significant interaction with ACET in the model but the effect size is weaker than in the centre-unadjusted model (partial $\acute{\eta}^2$ = 0.004 versus 0.005 in model without centre adjusted). The interaction terms between communication items and ACET are in the same direction for centre-unadjusted and centre-adjusted models: positive, i.e. higher RHD communication score complements ACET to better predict HL, particularly within the higher range, which ACET because of its limitation at double B tympanograms cannot do well. That additive effect of communication items, and also this interaction effect particularly in predicting HL within the higher HL range, can only

happen because they are imperfectly correlated, a necessary but not sufficient condition for their complementarity. Because the logistic models sample particular parts of the continuum and introduce noise (from which cases happen to end up just on each side of a cut-off) we expect only overall general correspondence of GLM results with logistic results, so another complementary function of the score form the communication RHD items may occur in models with particular cut-offs.

Table 4.4.1. Underlying model of HL prediction with ACET and Hearing questions, season and interactions, expressed as p values and partial ή²

Model without centre adjusted					With centre adjusted	
Variables	Without interaction Rsq adjusted 0.469		With interaction; Rsq adjusted 0.471		With interaction Rsq adjusted 0.514	
	Sig	ή²	Sig	ή²	Sig	ή²
ACET	0.000	0.358	0.000	0.351	0.000	0.358
Hearing rating item	0.000	0.069	0.000	*0.038**	0.000	0.030
PC RHD3 communication items	/	/	0.027	*0.004*	0.368	0.001
ACET*PC RHD3	/	/	0.007	*0.005*	0.025	0.004
Season/sine	0.020	0.004	/	/	/	/
Season/cosine	/	/	0.038	0.003	0.303	0.001

***Notes**:*

*Transformed version of variables: ACET and HL; * hearing rating item effect stronger (0.038) than PC RHD3 items and ACET*PC RHD3 interaction (0.004 and 0.005).*

Model with interaction is better (Rsq 0.471) as is the model with centre adjusted (0.514); most entrances are significant and large enough in effect size to be worh interpreting (underling).

Non-significant terms back-deleted at p < 0.1.

4.4.3. A range of cut-offs in screening target condition (logistics without interaction terms)

20 dB HL prediction: Although of theoretical interest in the GLM as being explicable on grounds of information theory, the interaction term ACET * PC RHD3 communication items is left out of the following exploration of variables useful in the prediction of desirable cut offs. This was decided for two reasons: 1) on general screening policy grounds (adequate sensitivity) the two lower cut-offs, at 20 and 25 dB HL were the more important possibilities to examine, and according to the GLM this is not now *a priori* the area for which the interaction is allocating a role to the RHD communication items (i.e. there was no need for precise prediction in the high-threshold range). (2) when we fitted interaction terms they always remained marginal in the logistic regressions. Analyses were also run for 30 and 35 dB cut-offs but are not presented in this application-oriented chapter. In the present logistic regressions, four independent variables entered the prediction significantly for the two relevant cut offs presented in Table 4.4.2. The following exposition starts from the lowest threshold, 20 dB, widely used as the basis of the clinical decision for further assessment, and is sensitive enough to capture the effects of fluctuation and long-lasting minimal hearing loss in OM children. The independent variables' *p*-values, OR(s) and CIs for four predictors for 20 dB HL cut off are shown in Table 4.4.2. The first step model had five terms: ACET, HR, PC-RHD3, PC-ESS and ACET*PC RHD3 items interaction. Interaction term and also ESS items were deleted as marginally significant and the 2nd step backwards-deleted model without interaction term was adopted for interpretation. The OR(s) for hearing rating item were 1.366 (95% CI, 1.080-1.728) but for communication items 1.591 (95% CI, 1.228-2.061). Surprisingly (from the point of view of the previous finding that the interaction term involving the three RHD communication items resolving in the high severity range) this term has a larger odds per unit increase from PC-RHD3 items in predicting 20 dB HL than this variable receives at other cut-offs, a paradox we have yet to resolve completely. At any rate the result emphasizes the importance of three communication items overall, including for screen accuracy at a lower cut-off in the target condition. For 20 dB cut off, given the requirement for 90% sensitivity, the specificity was 73.2% as is shown in the 2x2 classification table (Table 4.4.3.). The AUC-ROC in Fig. 4.4.1. for 20 dB cut off is .899, an acceptably high figure for the general ability of the proposed test combination to distinguish cases below and above 20dB HL Figure 4.4.2. For application of screens, positive predictive value is a useful health-economically relevant performance parameter. As predictive values

depend on the prevalence of the condition, we have calculated, and tabulated for various prevalence values logarithmically spaced from 0.1 to 30%, the PPV for a screen with 90 % sensitivity and 75% specificity, the latter being an indicative round figure rather than the preceding exact finding. Assuming 10% prevalence, the PPV had a yield of 28.5% (Table 4.4.5. and Figure 4.4.2.). This means that a child who has the combination of ACET and hearing questions defined as a positive screen test for higher HL has a ~28% chance of having > 20 dB HL when prevalence of OME is 10%.

The ACET OR is less than 1.0 because of an inverting transform, but .202 for example simply means slightly less than 5 times (i.e. reciprocal) the odds per additional unit in the independent variable. The unit is now the SD, the range is from about -2.0 to + 2.0, so about four SDs. For an upward shift of one SD in ACET which equates to about 7 dB equivalent, the odds of exceeding the 20 dB HL criterion increase by a factor of almost 5.0. So for a span of -2 to +2 SD we see about 20-fold odds increase for having the screened condition (HL > 20dB).

25dB HL prediction: For 25 dB cut-off, the estimates for the three independent variables; ACET, Hearing Rating and PC-RHD3 were slightly different from those for 20 dB with OR: 1.240, 1.466 and 1.337 respectively. The OR for the hearing rating question was higher than for PC-RHD3 at this cut-off. At reference sensitivity for the model of 90%, the specificity was 69.3% (Table 4.4.2.). The AUC-ROC was .866. The 25 cut-off model is thus also good but slightly worse than 20 dB prediction, giving lower specificity for the reference sensitivity of 90%.

The 20 dB HL cut-off model has better performance of parameters: sensitivity, specificity, AUC-ROC. The role of RHD items is not identical over different cut-off criteria, shifting between stronger rating and stronger communication items played significant role: a) they are more sensitive for lower HL screening criterion among the two models examined in detail, b) they showed positive interaction with ACET in the GLM better explaining HL in the higher range. These two observations are not entirely compatible but it is not worth detailed effort to resolve the paradox because for an extreme cut-off (low, 20 dB) the number of cases below this in the sample is not large, and the logistic model is generally less powerful and specifically unstable as cell totals become low, and the multi-collinearity between the two variables derived from RHD is high, making all models containing both terms slightly unstable.

Table 4.4.2. Summary of screen performance data for two definitions of the condition (in dBHL) to be detected

cut-off dB			Direct role in prediction Standardised Odds-ratios (Est, 95% CI)			Context of adjuster variables *p*-values as main effects		
	AUC-ROC	Specificity at 90% sensitivity	ACET	RHD rating item	RHD PC3 (comm. Items)	Center	Season (sine)	Season (cosine)
20	.899	0.732	.202 [.162, .250]	1.366 [1.080, 1.728]	1.591 [1.228, 2.061]	<.0005	.694	.083
25	.866	0.693	.240 [.203, .284]	1.466 [1.205, 1.785]	1.337 [1.084, 1.648]	<.0005	.081	.663

***Note**:*

The overall performances for the two cut-offs are generally similar but show slight differences: (i) for 20 dB we see dominance within RHD of the 3 communication items and cosine seasonal adjustment but for 25 dB we see dominance of the hearing rating, with sine seasonal adjustment. Strong conclusions should not be drawn as these patterns would not differ significantly. More reliably (as the 25 dB CI excludes the 20 dB mean, ACET is stronger within the 25 dB model (*presumably as the cut-off is nearer the median of the sample) but yet the overall screening performance is slightly poorer.*

Table 4.4.3. Raw classification table for ≥ 20 dB cut-off

		Predicted HL		Percentage Correct
		.00 < 20	1.00 ≥ 20	
Observed HL	.00 < 20	197	109	64.4
	1.00 ≥ 20	58	1036	94.7
Overall Percentage				88.1

Table 4.4.4. Raw classification table for ≥ 25 dB cut-off

		Predicted HL		Percentage Correct
		.00 < 25	1.00 ≥ 25	
Observed HL	.00 < 25	375	146	72.0
	1.00 ≥ 25	97	782	89.0
Overall Percentage				82.6

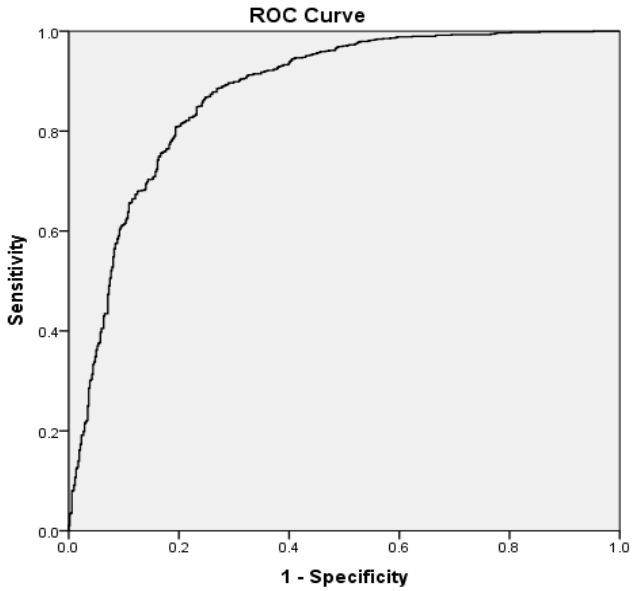

Figure 4.4.1. ROC curve for HL ≥ 20 dB, giving 73.2% specificity at 90% sensitivity and AUC-ROC 0.899

Table 4.4.5. Calculated table to project the PPV values in any screening programme, for a reference screening test with sensitivity 90% sensitivity and 75% specificity

p (in terms of percentage)	Sensitivity	Specificity	p	(Sensitivity)(p)	(Sensitivity)(p) + (1-Specificity)(1-p)	PPV
0.1%	0.9	0.75	0.001	0.0009	0.25065	0.003591
0.3%	0.9	0.75	0.003	0.0027	0.25195	0.010716
1%	0.9	0.75	0.01	0.009	0.2565	0.035088
3%	0.9	0.75	0.03	0.027	0.2695	0.100186
10%	0.9	0.75	0.1	0.09	0.315	0.285714
30%	0.9	0.75	0.3	0.27	0.445	0.606742

Note: *OME prevalence in children up to 2 years of age is around 4-5.5% (Bhutta, 2014; Rovers et al., 2000).*

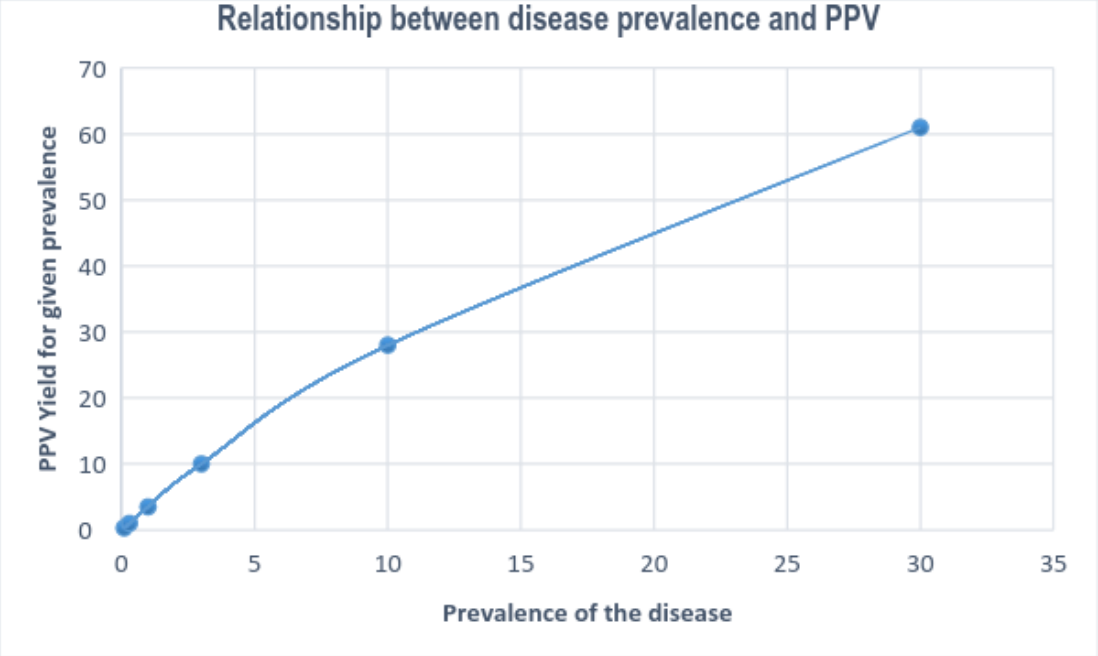

Figure 4.4.2. Graphical relationship between disease prevalence (X axis in %) and PPV (Y axis in %) in a test with 90% sensitivity and 75% specificity

Notes: *These sensitivity and specificity values are indicative rounded values close to what was found in the logistic model of 20 dB HL prediction.*

5.1. Discussion of Study I

5.1.1. Seasonality of the symptom scores

***Upstream variables**:* We found strong seasonality of symptoms severities for upstream facets: URTI, ESS, and HL. Some patterns are near-sinusoidal but others (e.g. ESS, HL) are not; nevertheless, the sine function provides smoothed estimation of the timing of annual maximum severity. The timing of maximum URTI severity is strongly present in the data, around the beginning of the March (the precise delays estimated have to be viewed as a centre for what is a broader peak). This peak coincides broadly with the peak in case incidence, and the difference between low and high severity at timing of peaks most marked in the age range most prevalent to disease. Nasal obstruction seems surprisingly to precede the infection symptoms in the annual cycle, just after sleeping score severities which appears as first in the annual sequence but as these are weak without significant seasonality they should not be forced into interpretation. The sinusoid fit for sleep disturbance severity is not significant but appears to be a week before URTI obstruction. As the application of this technique to quasi-time-series is new, wherever seasonality is not strong (and for these two it is not) interpretation of timing of maximum must be cautious. However, there is some reason not to entirely doubt the finding in the instance of sleep disturbance, but to see it as a conflict or cancellation rather than total absence of pattern. Clinically, sleep disturbance is often a symptom of AOM and its severity may also hint at URT obstruction problems – mucosal inflammation is involved in both. The possibility of two types and phases of contribution to sleep disturbance needs further investigation.

The winter peak in URTI is a highly expected finding from a highly predictable event, the known start in late autumn. This event of fairly regular timing results from the seasonal genetic shift in viruses and a shift in host immunity. The immune variation results from seasonal variation in gene receptor expression (Dopico, 2015) phenotypically responsible for more pro-inflammatory action during winter. The promoter of the seasonal changes is the daily temperature and shift in daily hours of sunshine challenging human environmental adaptation in the temperate zone. Winter virus infections promote pathological changes in upper respiratory tract mucosa and ET dysfunction. Denudation of mucosal layer and polymorphonuclear dysfunction promote planktonic bacterial binding and cytokine production leading to fluid accumulation and developing AOM. In experiments on

chinchillas, the highest incidence of pneumococcal OM occurs when the bacterium is inoculated 4 days after viral infection (Giebink et al., 1980) showing the mechanism for viral triggering of bacterial AOM. In another experiment study in chinchilla RSV promoted Moraxella catarrhalis middle ear infection 7 days after inoculation bacteria and 4 days after RSV infection (Brockson et al., 2012). In our study time delay between URTI week severity and ESS was one week. Also the month of the highest URTI incidence (Heikkinen & Chonmaitree, 2003; Biles, Buffler, & O'Donell, 1980) overlaps with the present estimate of the month of the highest URTI symptom severity.

Given that the animal studies are true time-series for the individual and we are inferring in quasi-time-series to the dominant presence of such sequences (i.e. from the patterning in large sample data across individuals), the congruence on the approximate one week delay is very encouraging. The distinction between the two different type of studies could be called cross-sectional (this epidemiological study) versus longitudinal (animal and clinical). In the canonical pathway for pathological sequence of OM expression, URTI is the antecedent of AOM and AOM the antecedent of OME. We here find a similar sequence of disease phenotypes in the annual cycle, but expressed by the phasing of maximum severity of symptom scores.

The seasonalities thus differ across measures, but for a simple and comprehensible account of the difference, the facets can be grouped into two major groups with a general causal link between them causally inter-related; upstream and downstream. The first variable, URTI, has its severity peak in the late winter, March. The sequence in these peak timings is essentially the same as with the sequence of categorical incidence of disease entities: URTI, AOM, MEE, and OME (Heikkinen & Chonmaitree, 2003) in human clinical studies. Timing for HL maximum severity is very close to that for AOM also in the late winter time. ESS peaks in the second week of the March, just a few days before the maximum in severity of hearing problems (HL measure). The reasons for similarity seem to lie in: i) justification of the level of symptom difficulties; ii) heavier cumulative impact overall severities from a general shift down the canonical pathway; iii) combination of new incidence with persistence of MEE from earlier post-acute effusions, responsible for raised air- conduction thresholds. The estimated time delay in the quasi time-series (in weeks) between URTI and ESS symptom severity is 1 week, very close to what is suggested by studies where sequence information is available in individuals, namely 3-5 days (Chonmaitree, 2008).

5.1.2. Relation of seasonality estimates to control for non-OM variables

Possible background variables showed modest effect on upstream facets and were taken into account as additive effects (confounders) with seasonal severity and not as determinants of the seasonality pattern itself. Interaction with other variables was not estimated except for URTI, the most promising example, due to strength of its seasonality. In this supplementary analysis we found when fitting also a marginal interaction for SES a positive interaction for age with the sinusoid term for the starting week that had shown strongest seasonality in the model for overall effects (week 11). In this test, the seasonality was slightly stronger in older children (p = 0.037; partial eta-squared = 0.002); a possible reason is that older children may be susceptible to 'new' autumn viruses but that their immunity and contact patterns make summer episodes rare. As these results were marginal and reasons for seasonal patterns to be conditioned by other determinants are hard to envisage, such interactions were not pursued and will not be further discussed. Significant overall effects were: SES, age and history in manifestation of OM and hearing problems. This multi-determinant pattern as in Study II is seen for most of the measures subjected to the present seasonality analysis and the reduction of error by fitting those adjusters similarly assists the seasonality 'signal' to emerge. Multiple determinants clearly need to be considered as the default approach in future studies of all OM facets. The evidence is mounting that OM is really a multifactorial disease and the causal relations never depend on one single factor but more on a combination of multiple factors, so only emerge clearly when this is fully considered by multivariable adjustment.

Like SES, length of history entered every regression run for these fine-grained seasonality analyses and in the direct longer → worse-affected. This parallels an earlier finding that a stronger effect of history on severity of AOM and hearing loss is present in otitis prone children; those with longer history are more severely affected (Biles, Buffler, & O'Donell, 1980). The effect of history on RHD is reported and explained in Study II and the related article studies (Milovanovic et al., accepted). Here the length of history has an additive influence on the RHD score, this emerging again when symptom severity is modelled on a weekly basis. From the present pattern of history effects throughout these analyses, it would seem that the basis for this is the fact of the report and to some extent the need for a specified recent reporting period on which the parent is requested to focus (although that period should not be over-precisely interpreted). In other words we have no

evidence or reason to believe that the length-of-history effect is specific to RHD as a facet. Rather, within the hearing measures, it is specific to parental report.

The findings generally support a conventional picture of the continuation of acute forms to chronic forms, but with these new estimates of specific temporal parameters we have added a relatively precise bridge to the sort of data obtainable on large samples. Some obvious consequences follow for evolution of disease and impact: i) there is a latency or silent period, without physical symptoms of disease but high severity of hearing measures; ii) the aggregation of hearing severity in communication difficulties, expressed also in high parent concern, reflects the fact that this latency period is not free of harm; iii) the eventual reflection of this on the downstream variables, usually seen as outcome measures, speech/language, behaviour, parent's quality of life happens later and variably in the more severe and persistent cases. The counterpart of this evolution is a dominant profile of clinical presentation that varies through the annual cycle, making the date of presentation useful information to guide case assessment.

Downstream facets*:* A history of hearing difficulties in preschool and school aged children has effects on behaviour, language, cognition and social activity; this is now accepted because observed and published in many different types of studies (Milovanovic et al., accepted; Bennet & Haggard, 1998). The positive history of OME is reflected in poor peripheral sensitivity and perhaps further forms of degradation of the afferent input while disease is present. However central auditory function, and especially bilateral aspects of hearing (localisation, and discrimination of sounds arriving from different directions in physical space) have been subjected to intense experimentation (Hogan & Moore, 2003). These sensory and cognitive manifestations of OM are diverse, and result from multiple upstream processes; so when they are measured and expressed as severities in the annual cycle the effects will have accumulated over differing time-scales, even though driven by a small common set of hearing difficulties and physical symptoms. The existence of two main pathways of influence on downstream variables in the conceptual diagram (Figure 4.2.1.) allows for dual influence on several of the downstream measures. This makes it more difficult to provide a simple overall account of time delays to peak severity than would a single cascade. Detailed time delays between pairs of variables considered must then depend on the pair under consideration, and predictions based on the particular upstream variables may be for multiple or very broad peaks of maximum severity. A fuller exploration of this

spreading is given in next section below. We may lose the ability to see, via match to a single sinusoid, a fundamentally longitudinal sequence reflected in the correlation structure of the cross-sectional data. We need a simple overall summary account on which further evidence can be allowed to accumulate in future, discussing sets of variables together. For the present we may have to be content with being able to explain generally why seasonality is weak for downstream variables. Although seasonalities for individual downstream variables are not statistically significant, collectively they seem to point to a delay by up to a quarter of a year – a whole season or 12-13 weeks.

It is not clear why speech/language should show only little delay, but there are relatively few items in all and their heterogeneity (e.g. speech versus language) may make it inappropriate to interpret the finding, due to a measurement problem. An analogy to language development may be helpful here, although that is unidirectional not oscillatory like the seasons. There are sequences for average timing of appearance of various types and levels of language structure and for their appropriate functional deployment. But the causal links are complex because development of one part of the system facilitates development of the other parts (Bowerman, 1985). Therefore a broad test that inevitably reflects several parts of the linguistic system maturing at different stages will not show sharp growth spurts that a narrow test of one function can, and indeed broad tests of language do not show sharp patterns of growth. Carrying this analogy back to facet seasonality with questionnaire items, we cannot be sure that different item timing as with RHD items is to be invoked and the items may just be relatively poor. But this discussion sets up a clear contrast between simply absent seasonality for the function, with the likelihood that this is due to non-discriminating items and contradictory seasonalities within the set of items, versus directly contradictory seasonalities. Future work should have the design and power to force this distinction. Thus with a short, catch-all speech/language questionnaire measure of only 3 items and know from the psychometric development stage not to be the best items in the questionnaire, we could not expect to do better and this may explain the lack of clear annual peak severity. The latent period between cause in upstream symptoms and effect in downstream ones is consistent with the idea of the accumulation of communication deficit being the main underlying process and some of the estimated delays for downstream variables show similar delays as expected. This explanation of paradoxically short delay to peak in speech and language severity score in terms of inconsistent item seasonality is not satisfactory but providing a strong explanation would take serious further studies.

In the timing of typical cases of OM driving the patterns in the mass data, the severity of the behaviour problems score is highest in June, a week before PQoL peaks, but three months, or 13 weeks, after HL severity maximum, and 9 weeks after parents have observed hearing difficulties. Although weak, this plausible timing of peak severity for behaviour makes the one for speech/language problems, early in the first part of the year just after HL severity, seem strange but suggest an additional element of explanation other than just measurement problems. It could be that at the level of data there is a whole year's accumulation involved in some cases, and that consultation is only triggered late in the year by the arrival of a new URTI season prompting a realisation on the part of the parents, perhaps by observed speech and language patterns then suddenly deteriorating (again).

The seasonality effects on all downstream variables are weak in effect size but stronger for child behaviour problems than for the other two. The weakness of the seasonality in downstream facets severity is logical when we consider the two pathways of influence from upstream variables; 1) physical symptoms of URTI and AOM and 2) impaired hearing functioning expressed by low HL and RHD. ESS makes a large contribution to total impact of OM (seen for example in the loadings on the PCQ14 in Study III and the fact that it emerges first in the 3-factor solution). In general, the fact of two pathways of influence makes difficult the separation of influences from upstream to downstream variables. The influence of originating variables, SES, age and history, raises the necessity for their control in calculating seasonality. Considered as risk factors in the literature, all these variables seem to have additive contributions here, and promote the pathological sequence from an acute to a chronic form of disease. The originating variables can influence both upstream and downstream facets groups, but they seem to be most influential downstream, in what is here called general impact. Their precise effect is not within the scope of this chapter.

5.1.3. Summary of correspondence of finding with conceptual schema for causal cascades

The complete set of facets discussed in this work and available from OM8-30 can be divided economically into two sets by contrasts in two dimensions, two ways, in effect two independent contrastive dichotomies: (1) upstream versus downstream, attempting to summarise causal origination or independence versus dependence, and (2) two pathways, with a mainly health pathway contrasting with a hearing-driven impairment pathway. These are expressed in the model as the path A and B. The A and B pathways are not strictly

separated through their several stages, and both data and known pathogenetic influences suggest some cross-linkage. A simplified form of the SEM for the OM8-30 data from Eurotitis-2 was presented in Figure 1.1. It shows some re-uniting of separate cascades at various points. If this major occurrence of parallel pathways (and some other minor parallel instances also) were to have different delay characteristics in the parallel path stages (and there is no over-riding reason for the delays to be closely similar) then the re-joining will spread the differing seasonalities of the previous stages into a broader even bimodal or flat distribution each cascades. This will lead to absence of measurable seasonality in the variable at the stage after the re-uniting. Thus the theory expressed in the SEM can, using part of the seasonality delay data, make predictions for other variables at least at the relative, ordinal level (X1 later than X2, etc) for delays. For example, downstream variables as given by the arrow directions in the figure should have later maxima than variables upstream in relation to them, and I have shown confirmatory evidence for several of these ordinal relationships. Not in every case will such predictions be confirmed exactly, due to the mentioned problem of parallel paths, and there may even be influences of unknown pathways not appearing in the figure at all and unknown influential variables missing from our data. However, that downstream delay is the general expectation and there are enough confirmations of it in the present results to continue with it and test it further. For some variables where no seasonality at all may be measurable we have heterogeneity as a rational, if preliminary, explanation. The present delay estimation results represent a completely new approach, using quasi-time-series data that are not true time-series; but the results potentially can and certainly ought to be reconciled with and brought into a rich integrative theory. They agree with many types of data already, as expressed by the structural equation model and are worthy of comparison with yet further types.

5.2. Discussion of Study II

5.2.1. Homogeneity versus 'discrepancy' among the three measures, especially between RHD & HL

Clinical research is insufficiently aware of the relevance of measure quality, and reliability in particular, as a basis of statistical power to answer questions, one of equal importance to sample size. Where a scientific or clinical aim in OM prioritises the reliability of a measure (for power and, precision and for generality), the inter-correlations seen here would justify aggregating the three present hearing measures. The lower correlation of RHD with each of the objective measures would not usually justify taking a total in the form of 1st principal component of variation. However, the patterns of influence (Table 4.2.3.) are sufficiently similar to not undermine validity of such pooling. The principal component is a standard method for pooling across differing metrics. However, where the aim prioritises specific validity, it is necessary to be aware that the three measures do not behave identically. The diagram (Figure 4.2.1.) conceptualises the tension between having high inter-correlation, arising here from the close causal sequence of the processes underlying the three measures, and having incomplete identity of the pattern of influences. For example, ACET is less seasonal than HL (as all participants are clinical cases of one or other condition for which this is a basic marker). Greatest relative distinctiveness is seen for RHD, which is more dependent on SES and length of history, and has a delayed seasonality pattern, relative to the objective measures. However the absolute similarity of the list of influences allows that where measurement effort has to embrace the diversity of the three present measures reliability can also be obtained at no extra effort or assessment cost by totalling them in this way. Even the mediation analysis increases the justification for totalling, by linking ACET to RHD in a way not exclusively mediated by HL. Table 4.2.3. showed a consistent, but slightly more powerful, pattern of determinants on the PCtotal compared to those of the measures separately. Given the complementarity in objectivity and range between ACET and RHD, a later study IV probes the potential value of ACET (i.e. tympanometry) and RHD in offering an HL-surrogate in settings (family practice, screening) where HL is not feasible, for cost or acoustical isolation reasons. That and other approaches to balancing quality of measurement with its cost, feasibility and efficiency stem from the present documentation of inter-correlations and determinants.

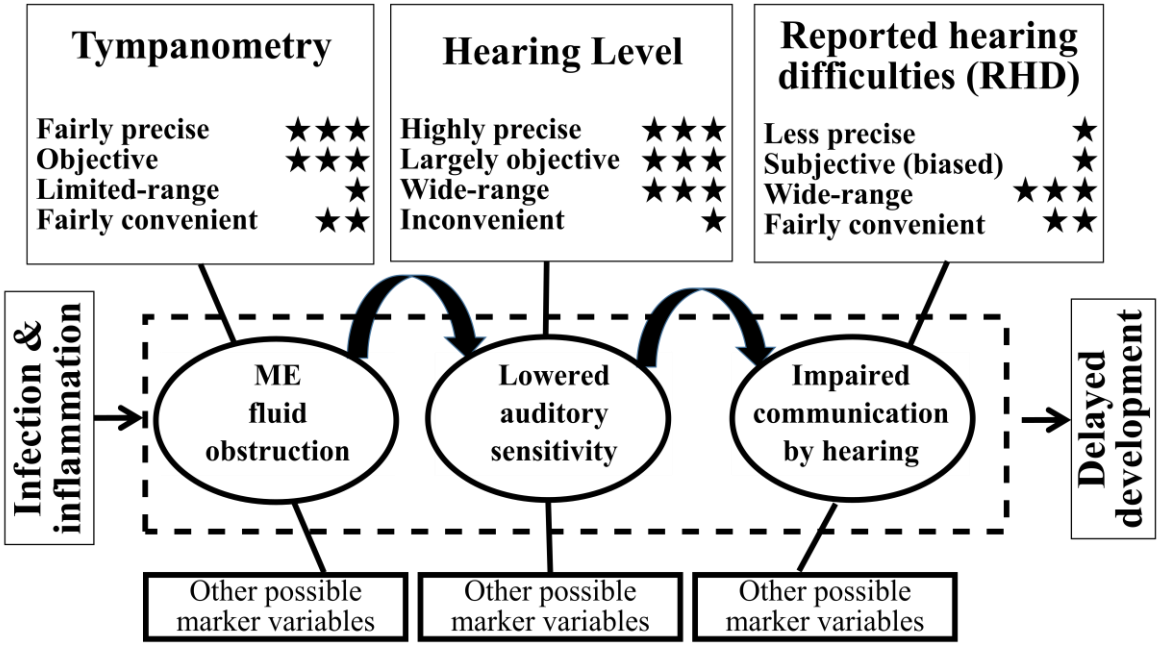

Figure 4.2.1. The three hearing measures relevant in OM, their underlying constructs and properties of the measurement techniques.

***Note**:*

The main point of the analysis and this diagram is that the direct non-mediated path is not trivial in strength.

The counterpart of this acceptability for pooling is the discrepancy between RHD and the two objective hearing measures, but conceived as particularly relevant for HL. The particularly differing effects on RHD of SES, season and length of history are noted. However the most direct attempt to capture discrepancy (the Z-diff model in Table 4.2.5.) does not reveal dramatic discrepancies: only moderate determination of discrepancy by length of history and strong determination by age. For the history finding, the Z-diff form of the result is a totally adequate alternative to an interaction test, and useful in expressing directly that RHD reflects history significantly and materially more strongly than HL does even if HL does somewhat. Explanations for other differences in patterns of results between HL and RHD are therefore entitled to invoke Length of History as a plausible explanation. The direction of the age effect is positive: disproportionate RHD in older children. Scientifically less can be made of this age finding, despite its strength, because there are probably three confounded and somewhat obvious contributions: greater parental awareness of deficiencies in hearing-related behaviours from the expectations for older children, and more, also more long-lasting, hearing problems in the older children who consult at hospitals

with OM. The balanced view of this reality (some discrepancy, much similarity justifying totalling) poses a difficulty for dissemination into clinical practice. People, nor just doctors, like single simple black or white stories, but that preference does not justify excluding one story from this genuinely mixed picture.

5.2.2. Effect of socioeconomic status or maternal education

For present broad-brush purposes these influences are taken as one; in recent decades maternal education has displaced employment category as the preferred SES indicator for a simple single data item in social science research. Many studies (e.g. Paradise et al., 1997; Tharpe & Bess, 1991) suggest that less privileged families consult later or only at higher severity. Leaving aside some over-use and over-intervention on those that can pay in advanced countries, on a global scale families' health behaviour is more appropriate, and healthcare uptake is more complete, in the more educated. In the present data we do not see great evidence of over-consultation, as all three measures suggest that, in less favoured SES, healthcare uptake occurs only at a higher level of severity (because within the sample, the lower SES children are more affected). This socially regressive trend in healthcare uptake is however more marked for RHD. It can be safely supposed from studies like the two cited that for a given objective health care status, higher educational level corresponds to greater awareness. The discrepancy between RHD and HL could be construed as a cognitive bias based on parental concerns or beliefs differing between individuals (and potentially with SES), somewhat like a subjective placebo or nocebo bias in a reported health state. The present findings on temporal contributions to RHD discourage that interpretation. We cannot rule out part of the SES effect being of a cognitive bias nature. However, the SES portion, in the presence of control of both, ACET and HL, provides only about one tenth of the explanation of the variance in RHD (i.e. of the cumulative total partial eta-squared values, and roughly of the Rsq). This is not consistent with interpreting RHD as reflecting mainly response bias; rather the fact that RHD is sensitive to duration (also a report-based measure) allows a direct deduction from the former principle (a) that lower-SES children reaching secondary care would also have objectively longer histories on average. This direction of association is opposed to that from the expectation (b) that lower SES families will have also been less aware of behaviour patterns in very young children suggestive of OM or hearing loss, so objective early onset and length of history would be cancelled behind a shortened (because less sensitive) remembered and reported history.

A supplementary analysis on all data with available SES and history showed that process (a) strongly pre-dominates over any process (b). Making length of history now the dependent variable, parents of lower-SES children produce answers to the question about months of having the condition that are longer by an average of 0.26 months (N = 2,404; t = 5.8; p< 0.0000005; partial eta squared = 0.014). One week (approximately) is an apparently small difference within a wide variance but it is highly trustworthy as a finding. The interaction terms from the analyses threw some further light on these issues but were not tabulated in Results. Of four interaction terms meeting initial significant criteria and having overall partial eta-squared of 0.003 or above (in Results), only one justified interpretation as having the non missing (so interpretable) component also above this criterion. It was SES*length of history for HL as dependent variable (p = 0.005, partial $\acute{\eta}^2$ = 0.004 for component lower versus upper SES). The direction is that lower SES with longer history has disproportionately poor HL. It is not possible to draw specific conclusions about lack of timely access to treatment, but this finding underlines a scenario of where the cases of most concern typically come from, one in which SES-related lower healthcare uptake plays a major part. There may be socially regressive implications of differential access to or uptake of secondary care in the first place, to be addressed by more proactive public health policies, but those are a separate matter. The trace of socially regressive awareness left in this ENT sample is minor, and the incorporation of an RHD measure into treatment criteria applied to the child who has actually reached hospital would therefore be a socially progressive, because of its link with past persistence, not a regressive, step.

5.2.3. Season and delay – implications for practice

The distinctiveness in determination of RHD by both delayed seasonality and length of history is parsimoniously explained by two suggestions about otitis media and RHD, for which there is other independent evidence: (i) in addition to reflecting the measurable short-term HL (which fluctuates in the more numerous milder cases), the RHD questions to parents reflect a medium-term length. Hence RHD is likely to be reflecting in part at least the scope of the parental recall period requested in the questionnaire. This ‘medium length’ is not precisely definable, although a scope of 3 months was requested on the questionnaires. This makes one element within RHD reflect an accumulation of the effect of auditory deprivation. The second suggestion about RHD (ii) invokes the fact that persistence and recurrence are longer-term characteristics of individual pathogenesis, so that until we have separate, perhaps

genetic, biomarkers of these characteristics, the best predictor of medium-term persistence and recurrence may be past persistence/recurrence. In this light, imperfect correlation of RHD with HL is a virtue not a shortcoming; along with history length, it predicts persistence/recurrence, hence the ability to benefit from treatment. Failure to supplement the objective measures of current ear state with information on prognosis and so with the possibility of avoiding side-effects of treatment if it is likely to be non-beneficial, could be seen as clinically negligent. We can hope that more precise bases of prognosis emerge from current biomedical research, but the present work offers a simple, low-cost and evidence-based interim solution.

Perhaps the most striking difference between measures in patterns of influence is the delay of 2-3 months in peaking of RHD. It has to be acknowledged that in the direct and explicit modelling of discrepancies, season (as cosine of month) is not a major determinant of the difference between RHD and HL, giving partial eta-squared below 0.003 and on the margin of significance. However, this does not necessarily mean that the delays in peak severity are similar, a question needing to be approached in a different way, preferably with true time-series data. Also the variability in difference scores always makes phenomena hard to show. Rather it should be concluded that at the individual level current season does not come through as a major determinant of the discrepancy, alongside age, history and other unknown sources of variation at the individual level. In the present summary models, sine/cosine pairs applied to monthly data summation periods offer a simple single-pass estimation method for appropriately adjusting the examination of other influences (Study I). There is finer seasonal detail in those other analyses, in slightly differing delays for the communication behaviours tapped in the various RHD questions which require more precise analysis and weekly data summation, to be reported elsewhere. Tympanometry as ACET is partly related to diagnosis and explaining why the child is being assessed at all. From Table 4.2.3. we saw that ACET contributes very little seasonal variation, which arose mostly in RHD. There are many ways in which clinicians need to look beyond diagnosis to profile in assessment of OM; of these, one is thus the orientation to children consulting and meeting OM diagnostic criteria from late spring through to mid-Autumn. Of course, these are fewer than in the opposite half of the year, but they are more likely to have long-duration auditory deprivation and to have or be heading for wider developmental impacts.

5.2.4. Pattern of missing data in Eurotitis-2

This topic is more fully covered in the General Methods Chapter, in the related publication (Milovanovic et al., accepted) and in Table 4.2.1. The sub-samples with ACET present and even more with HL present define a more affected subpopulation with more hearing problems. In one sense this can be considered biased availability of data but in another it is an appropriate (if variable and under-specified) way of defining a sub-population that it is more necessary to study with hearing measures. Centres were given no explicit request on this beyond asking for HL for study purposes on as many cases as possible. As the sample is large, we may assume that practice in centres' decision to acquire HL is broadly representative of real world practice, responding to resourcing constraints and other practical obstacles. Where dual analyses on complete-data and maximum cases analysis are called the hybrid HL/ACET has good distributional properties and can offer very large cases gains by comparison with deletion of cases lacking HL. It is attractive in conserving degrees of freedom (e.g. a single analysis not two) and in making best use of data available. I have not actually reported analyses on > 2,000 cases with this hybrid as dependent variable; this is because it offers a new method that it would be best to await acceptance before using it as a powerful tool to document phenomena. We made a general recommendation in the article version of the material in this chapter (Milovanovic et al., accepted), based on random simulations of rates of missingness and substitution of ACET for HL. This recommendation was to not exceed 25% case gain by such hybrid imputation, in order to maintain high correlation with the hearing measure value ($r = 0.92$) that would have been given by true HL. In an independent variable, or under special circumstances, the recommendation might be over-cautious. But it would not have been good scientific strategy to make it, then immediately break it.

***Limitations*.** The use of a large sample and the emphasis on similarity of large effects across sample inclusion and across several hearing measures protect this work from the most usual and not always admitted research limitation, low power with unadjusted multiple testing. This driver of publication bias – along with that incentivised exploration for isolated factoids of low replicability - has recently been recognised as a threat to the reputation of biological, social, psychological and medical science (Ioannidis, 2005). The strength of multi-variable adjustments in the GLMs is that they limit the arbitrariness seen in univariate associations, as each is shown in a rather fully controlled context. Two classes of limitation in

the work should however be acknowledged, but each is more likely to lead to under- rather than over-estimation of effect size. The first concerns precision of measurement of some independent variables. For SES, maternal education is widely used as a very good single item, perhaps the best in advanced countries since about 1980; so as a stratifier likely to also be used in further OM studies, it offers feasible control where SES may be important. However as a basis for serious specific study of SES effects, their interactions and mechanisms of influence it would be grossly inadequate in ecological validity (i.e. the sub-domains sampled) and in reliability (number of data items). Fitness for purpose has to be the basis of choice of method.

Except for the funded TARGET sub sample of Eurotitis-2, there are quite high rates of missing data, as expected from the style and administrative basis of the study. We have transparently reported this and presented specific information and comparisons on missingness (a practice not widely followed), showing where it may bear upon main conclusions and no Conclusion is undermined by this. For SES data missing, we know from this and other studies that data are not missing at random (MAR), because the adjusted mean estimate from non-respondents is 'worse' (i.e. more severe disease or impact). This is probably just social desirability bias, the item being about something which the most adversely affected do not wish to admit. Here floating the estimate for 'missing' specifically avoids making the (false) assumption that data are MAR, which a majority of clinical studies make in the interpretation, even if they have reported missingness adequately. The retained cases contribute power for other effects, and the reported SES effect for the cases giving the data is not subject to bias.

5.3. Discussion of Study III

5.3.1. Criterion validity; OMQ14 versus OM8-30

The short form of questionnaire, OMQ14, showed good criterion validity in the form of high correlation of PC total and the three factor scores (giving the disease profile) with the corresponding total and facet scores of OM8-30. Discarding 18 items from the 32-item pool in the larger form of questionnaire to get to the 14 in OMQ14 raises several challenges: of reliability, of supportable richness of profile and heterogeneity versus homogeneity of the items in the facets or factors. In particular, to support prediction of QoL with the items best predicting it and at the same time to make a supportable downstream factor with the few items affordable (in respect of burden, not finance) in a short form required acceptance of some heterogeneity. With the items available, it was not obvious from the factor analysis what pair of homogeneous (high alpha) factors (created by splitting the inhomogeneous impact measures item set) could have been supported (taking 4 factors as the absolute maximum supportable on 14 items). The lowest criterion validity correlations were observed for the two inhomogeneous scores (PCtotal and Impact) and this contrast with the other two, homogeneous, ones was itself significant; however the impact items also suffered the highest proportional reduction in numbers of items when composing OMQ14, and this is the more probable explanation of their lower but still acceptable criterion validity. The lowest correlation is the one between the impact factor of OMQ14 and a corresponding impact score extracted from the developmental domain in OM8-30, probably because of discarding whole facets (sleep) as well as a large number of items (e.g. from schooling, language) not found to predict QoL particularly well. This item reduction left such facets with too small a number of items, and not enough interrelation between these facets, for any form of aggregation for reliability and validity to really boost these properties. Such a shorter length of questionnaire is an appropriate instrument for clinical assessment or fast clinical follow-up, because it includes overall OM impacts and also severity in 3 profile dimensions, with less burden than the longer form OM8-30 research instrument. Internal consistency was satisfactory (all but one alpha > 0.70 for a facet score corresponding to the factor score). This gives a fair summary of the present factors' consistency values for those who do not think in terms of the factor-analytic equivalent.

The dissociation between factors from the orthogonally rotated factor solution was observed for OMQ14 in contrast to the modest inter-correlation for OM8-30 facets. This intended poor interrelation between factors is desirable for economically summarising the diverse presentation in OM. It also makes it easier to discuss and to demonstrate possibly separate patterns for the separate scores, useful whether they are being considered as dependent or independent variables. This principle of orthogonality (statistical independence) was confirmed as being closely realised in practice on the sets of cases available for the correlations. More importantly, the available set of 7 determinants (including what may be considered the classical demographic risk factors) from the questionnaire scores was next examined for its detailed pattern of influences; as well as the similarities expected for a single disease group some very clear differences of pattern did emerge between factors. The fact that these are largely as expected and explicable in terms of existing understanding of OM is an important overall demonstration of construct validity.

5.3.2. Determinants of the separate factor scores

Ear symptom score (ESS): The pattern for the ESSQ14 factor (RAOM severity) is somewhat distinct from the pattern for the two other factors in terms interactions seen between independent variables. ESSQ14 shows a clear late winter peak in severity, reflecting the peak incidence, and the well accepted pathogenesis cascade after winter URTI, recently summarised as the canonical pathway for OM (Bhutta, 2014). The most prominent result for the ESSQ14 factor is that the score is worse for children from low SES families (low maternal education). Thus RAOM severity is under the influence of two major risk factors: season and SES. History alone is not a strong overall determinant of RAOM severity, but is an important modulator, the ESSQ14 factor score being worse in children from low SES families. This is consistent with the highly otitis- prone children due to immune system immaturity being more often from low SES families (Robertson et al., 2012; Steptoe et al., 2011). The fact that the severity of the RAOM is raised in children with both low SES and longer history makes this explanation likely, as it is the immature immune children who have longest histories (Sharma & Pichichero, 2013; Whelan et al., 2006). The history of ear symptoms overall and age do not affect the ear symptoms score, but the length of history enhances the effect of SES in this dataset.

Reported hearing difficulties (RHD): The results with the OMQ14 factor scores from questions on reported hearing difficulties (RHD) closely parallel those reported for the OM8-

30 version in Study II. This RHD score and the one for the general OM impact are more similar in their relation to risk factors and other determinants, despite the fact that the scores are by definition uncorrelated. However there seems to be a particular cross correlation between the three RHD items about effect on communication and three behavioural items within the impact factor. Although continuous RHD is moderately correlated with both the standard HL and tympanometry (scaled as ACET) so can be used as a substitute (Study IV), its essence has to be appreciated as very different. Parental concern is the first sign of notable communication impairment observed by proxy responders and deserves special attention (ASHA, 1997). Our findings do not show any strong influence of season or gender on severity of hearing report with referred children. With the clinic sample, the older children are worse affected, as are those with longer history; even though each variable is adjusted for the other one in the model, it is possible that the general relationship between age and longer OM history underlies both findings. However we do not find any interaction between age and history in the direction that longer history would make older children particularly badly affected. Both risk factors, age and history of OM disease, are important catalysts of a third risk factor, low maternal education used as a measure of SES. SES is not a strong predictor of RHD severity but the SES effect of influence is stronger in the more positive history of disease and older age. Interactions of age and history of OM emphasized importance and effect of SES. Older children from low SES families are worse affected and it may be reasonably assumed that the most frequent reason is unrecognised, untreated OM (E). Even if the interactions do not add greatly to the variance explained, these interactions of determinants or risk factors have an important role in interpreting the complex expression and presentation of the form of the disease.

General Impact: Diagnosis is a substantial predictor of general impact via its influence on RHD, which includes 3 aural communication questions. Combined diagnosis and OME children have worse RHD score. This diagnosis effect is reduced on fitting HL and ACET, which are the strongest and most direct predictors of RHD severity. Poor hearing expressed as flat tympanograms over time (which is addressed in more detail in Study I) and as average higher hearing threshold, or as both, exerts a cumulative effect in communication difficulties as the result of persistent poor auditory input. The present RHD measure can be said to be ¾ a measure of communication that shares similarities with impact factor items. The clinical lore recognises that RHD can often be bad in those with unrecognised disease, and the other measured effects here of low educational level of parents, long history, older OM children,

and diagnoses of OME or Combined OME (with super-added RAOM) can presumably be summated, so invoked to explain severe cases. It can be useful that the feasible RHD questions about hearing have this dual reference – sensory and socio-linguistic, but it also has to be considered as a reason for disaggregation in an analytic context (Study I) or even in an approach of optimised application (Study IV).

The OMQ14 general impact items are a very broad group of 7 items. Despite the heterogeneity resulting from the having three facet sources (speech/language, behaviour and parent quality of life) the facet inter-correlation is evidently high enough to lead to justified justify extraction of this factor in OMQ14; the factor score is sufficiently correlated with the total of all 14 items on the corresponding OM8-30 facets to claim criterion validity. The fact that the correlation is nearer 0.90 than 0.95 is presumably due to the combination of items not having been selected to maximise criterion validity of a whole measure but on the basis of individual item correlation with a separate measure of Quality of Life (Dakin et al, 2010). There was no guarantee that the balance between the facets would be maintained as half of the items were discarded. The equivalent facet scale Cronbach's alpha is just below the general criterion of 0.70 but in the circumstance of defining the measure as heterogeneous (hence the word 'general' retained the title) the convention of making factorial purity the sole index of measure quality can be suspended. This is instructive for measure design.

The younger children with OM are typically more affected as to general impact than the older and more prone to developmental problems; speech and language, behaviour, parental tiredness. This is not strange, because the OM history and symptoms have the greatest influence in development of younger children especially when other RF are present (Hall, 2014), mainly low SES. High HL significantly influences impactQ14 score but the effect is not large. Children with positive history of disease are worse in impact measures and longer history in younger children increases OM impact severity. Age is significant as main-effect risk factor but fitting the age*history interaction spreads the covariance too thinly, leaving no significant age main effect. This means that history is an important marker predictive for communication and developmental impact in younger children. A longer history in younger children makes the measured general impact more severe as expected. History, age and HL are the most powerful overall determinants for impact score. Interestingly, the different diagnoses do not make a large difference to impact within cases receiving hospital referral. The likely explanation for this absence of any large difference with diagnosis is that the

general impact includes items from both the physical health and the hearing - communication pathways to lowered quality of life. In an unpublished exercise, equivalent to mediation analysis but using structural equation modelling (Filipovic et al., 2013), we showed that the health pathway (physical symptom items) has roughly equal influence on parent QoL as the more function-related measure (hearing, speech/language, development) provided that its mediated action through the later is taken into account. These findings together shows that general impact, despite its extreme brevity at 7 items does capture diverse aspects of quality of life, via a multi-aspect capture of disease influences, and supports the acceptance of heterogeneity for some types of measure. The pattern of determination by risk factors, hearing measure and diagnoses, for impactQ14 factor has something in common with that for the RHDQ14 factor as noted above, but also some distinct features.

5.3.3. Global score

The totalPCQ14 score is evidently a global score of OM influence, summing the three chief aspects of the disease that are predictive of QoL. I do not address here the precise optimal formulation of OMQ-14 to map QoL, but two considerations argue that it must be good when completed. (1) Using OM8-30, Dakin et al showed that a weighting of the facets best predicting QoL also had a very high correlation with the 1st principal component of these facets, in which the items had been weighted only to express their interrelatedness not some external criterion variable. (2) Only items well predicting QoL entered OMQ14. On this basis, the PCtotal can provisionally be accepted as similar to a mapped QoL measure, mapped because although valid and predictive the ESS and RHD factors are specific not generic and do not themselves truly fall in the Quality of Life domain. The general impact could perhaps be accepted as falling in the QoL domain. Some writers (Christina et al., 2014) do refer to specific symptom scores as also falling in the QoL domain but this unfortunate usage results from overweighting the contrast between objective markers and subjective reports (QoL being taken as essentially subjective), and under-weighting the important distinction for healthcare decisions between specific and generic measures. A further reason to accept PCtotal as a surrogate QoL measure is that ImpactQ14 provide about half of the total pool of items and explains 10.6% of the total variance in the factor analysis.

Given this contribution of the impact score to the total, it is not surprising that the influences of age, history and HL on PCtoalQ14 are similar in general pattern to those on the general impact measure. The seasonality seen for PC total is not exactly similar to that seen

for any component factors but a compromise, a blend resulting from the mentioned influence of two pathways. Severity of PCtotalQ14 is the highest in late winter and early summer as expected from the downstream delays established in Studies I and II. The first pathway is the health pathway from AOM (ESS) and the second, the early summer accumulation of hearing deprivation is expressed via the communication and development path to general impact. In the basic model for general impact the sine term (late winter) is strong, but cosine (early summer) surprisingly not significant. Fitting the hearing measures into the model then reveals powerfully how these influences work. The hearing measure themselves capture the sine severity peak in late winter, leaving no room for an extra such term. On the other hand, the early autumn peak becomes more powerful, showing the expected significant negative cosine effect. Impact on global life quality depends on hearing level accumulating over time. This finding parallels the finding (Study II) where RHD items became seasonal after controlling for hearing measures. However a corresponding seasonality switch is not seen here with the RHD score based on the items in OMQ14. The reason is not immediately clear, but it could be the result of several possibilities: i) the underlying results may be close to the significance boundary and not truly in conflict but readily influenced in one direction or the other by small systematic or chance features of the data; ii) different types of cases because different centres (much of the data in OMQ14 but not in OM8-30 comes from the Balkans – See Appendix II); iii) also the sample differ additionally in the dropping of one item in the RHDQ14 item pool, possible obscure effects of the low loading items incorporated, and more missing items in the larger OMQ-14 dataset data pool; iv) there is also a slightly different way of fitting the determinants in the model. In the more complete data sample of 1866 cases the effects of some determinants are stronger; this is especially true for the SES effect on PCtotal. Three interactions, SES*history, SES*Age and Age*history, influence PCtotal score; two emphasized the role of age and history in the SES influence on generic outcomes and the third influence of history on age. These findings strength our conclusion that fitting interactions in the model for QoL measure is crucial in understanding complex interrelation between dependent variables and independent variables. The younger children with longer history are worse affected overall as also are the children with longer history combined with low SES.

The two hearing measures are the strongest predictors of severity in RHD and PCtotal as a surrogate quality of life measure. The effect of HL is stronger than ACET and this is reasonable on grounds for range; we know that ACET explains very little of the variance in

HL above > 30 dB. The fact that it enters at all when HL is in the model is almost certainly due to its discrimination among minimal conductive hearing loss less than 20 dB, where lack of good sound isolation and concentration difficulties make measurement of HL unreliable, leading to the decision in many clinics to not acquire the measure or to log it in a coarsely quantised fashion (Study II, discussion). The 'combined' diagnosis is worse affected on all OMQ-14 factors, all aspects of the disease profile emphasizing this diagnostic category as deserving further study and conceivably special clinical consideration (e.g. lower criteria for urgent treatment). Two striking features of combined diagnoses are high ESSQ14 score due to RAOM and high hearing threshold estimated by HL and ACET, both predictors for poor impact. SES is the most influential risk factor, probably responsible for much of the centre differences in the Eurotitis 2 sample. This issue is taken up further in the discussion of the Balkan sub-sample (Appendix III).

5.4. Discussion of Study IV

5. 4. 1. Introducing screening for OME; yes or no?

OME prevalence is high enough to consider a policy of introducing preschool case identification for OM children in order to treat those who on further assessment are expected to benefit. This does not mean that the case is strong. Publications have not explicitly concluded in favour of long-term benefits after early screening and treatment. The outcome variable most often mentioned when considering screening is speech and language as a direct consequences of hearing deprivation at early age (Rovers et al., 2000; Paradise et al., 2003) and this may have unduly limited the examination of potential benefit. In order to test the benefit of early intervention in OM children we need evidence of substantial effects from having fluid in the middle ear on hearing threshold and auditory signal processing in peripheral and central auditory pathways. Many groups of studies have addressed that issue, one important group being experiments in animals showing pathological changes in brainstem, midbrain and olivo-cochlear pathways after experimentally induced unilateral, or bilateral OM and CHL (Sanes & Kotak, 2011; Webster, 1977; Myers, Ray & Kulesza, 2012). Such experiments cannot mimic the exact conditions of OM including its variability, because the experiments are done in strongly controlled conditions and observed in close time window (Whitton & Polley, 2011). The lack of possibilities for close observation, longitudinal follow up and evaluation, big individual differences between children and tests for language examination give confusion in conclusion of OM outcomes and consequences. Language performance tests in OM children confirm expressive language disorder in age 3-7 as an OM sequela (Zielhuis, Rach & van den Broek, 1990; Shriberg, Friel-Patti, Flipsen & Brown., 2000) and affect sociability and other aspects of behaviour (Vernon-Feagans, Manlove & Volling, 1996; Bennett & Haggard, 1999).

In Study II we have analysed in detail the relation between three hearing measures documented previously. We found that ACET, HL and ACET & HL together predict RHD items and that these three hearing measures are interrelated, complementary and that they are under influence of the same variables: diagnoses, season, history and SES as examined in detail in the full paper (Milovanovic et al., accepted). These interrelations provide the justification for the present fairly successful attempt to predict HL from ACET and RHD in a type of ‘pincer movement’. RHD items as a part of OM8-30 questionnaire provide a marker

of longer-term hearing function as a predictive mediator of quality of life and hearing performance of the child (as observed by parents or caregivers). Starting from these extensive analyses in Eurotitis-2 on more than 2,800 cases, we may assume that the choice of questions and the scoring values for RHD offer a promising and relevant instrument for hearing assessment of OM children in the population of age range relevant for OME. Demonstrations of practical value could justify the search for more or better questions of similar type.

In countries with UNS, the aim of a preschool hearing check-up screen may include identification of rare cases of mild permanent and progressive hearing loss that have escaped earlier screening arrangements. However, at the first point of contact with the health system place recording of fluctuating, mild to moderate, conductive hearing loss in its persistent or transient form. Starting from the fact that CHL due to persistent or transient middle ear fluid affects more aspects of child and parent life, hearing as an upstream variable (in the sense of the conceptual pathway diagram similar to a Structural Equation Model) and other disease markers relevant to impact should be identified in a timely fashion and lead to further selection for treatment. Also we point out that episodes of CHL during recurrent OME exhibit cumulative effect on auditory performance and could be expressed in accumulated parental concern.

5.4.2. Underlying GLM

For the underlying GLM, adjustment for individual centres would be the usual approach, but certain effects need consideration both with and without such adjustment for their interpretation. The following summary of the GLM is robust across these two versions; whilst magnitudes of effects may differ, the general form of the model is the same. This contrasts with the position on moving to logistic regressions with differing cut-offs where distinct patterns are seen. Differences between centres can cause small differences in the model, which is usually strengthened by an adjustment for a strong influence. The reverse can happen where for example centre adjustment might over-correct and absorb into the centre term an effect that could be considered as having a more specific origin. However this was not seen here and weak terms of early summer season and interaction of RHD3 items with ACET were clearer with than without centre adjustment. This is fortunate because there are reasons to consider the adjustment necessary when using the logistic regression to simulate a screen and the approach assumed it was more appropriate.

RHD items showed good prediction of HL and the different forms of items contributed in a slightly different way: (i) the overall hearing rating item is generally the more robust in predicting HL because the question has a wide range; ii) its severity peak in the annual cycle is close to the peak for HL; iii) better predict higher threshold cut offs in logistic (next paragraph) and iv) the interaction with ACET is not significant because it is a strong wide-range item and the overall effect dominates over any need for it to supplement ACET within some particular part of the range. The RHD communication items entered the model significantly as PC-RHD3, and also as interaction term with ACET. These items make a contribution generally similar to that of the hearing rating, but specifically complement ACET in the range <30 dB. However, considering the partial eta-squared values for the overall and interaction effects, the single rating still contributes more to prediction. The clear implication from this is a need to increase reliability by having more good items like the rating. As with visual analogue scales (VAS), the obstacle to doing that is the inability to ask the same question twice (offensive to respondents). However further improvement could result from also having a visual analogue scale VAS as a different form and permitting questions very similar to the rating but referring to differing recent time periods. There are still possibilities for the improvement of questionnaire items and their scoring.

The peak in severity in annual cycle for RHD3 came later than for HR, although both came several weeks after HL's severity peak (March). This justifies the interpretation that the communication items reflect the effect of accumulated hearing deficit on auditory perception. In some samples and with some sets of variables in the model the summer maximum for RHD becomes marginal. Here, by separating out the hearing rating with its earlier maximum and the RHD3 items return to showing a summer maximum (negative cosine term). The underlying model for the screening study points this out by showing that to improve HL prediction, the cosine term for early summer maximum needs to be used when the communication items are in the model.

5.4.3. Logistic models for 20 dB and 25 dB cut-offs, based on HL prediction, differences and goals

Using the logistic model for an entire set of simulated screen target cut-offs (20-35 dB HL), ACET always contributed very strongly via the underlying HL prediction, but the RHD items had a smaller contribution, possibly differing at different HL cut-offs. This is most evident between two lowest HL cut-offs, 20 and 25 dB. The rating item is the strongest single

item and suppresses the contribution of the other 3 combined for discriminating HL > 25 dB, and at the higher HL cut-offs the OR for the rating rises still further. Despite the contribution of HR item at higher HL values, the overall performance of the combined formula (measured by AUC-ROC) deteriorated for cut off > 25 dB and progressively beyond. This is explained by ACET always being the strongest term and the known fact that it resolves better in the marginal to mild region.

The picture was different with the communication items, not being so strong in the underlying GLM but relatively potent for the 20 dB HL cut-off. The findings offer some new insight into apparently mild HL deficit in fluctuating OME when the tympanometric result swings between C2 and B forms. The items capture the cumulative effect of the mild average HLs accompanying these swings through marginal ear status. The formula is reflecting the conjunction of an enduring past history (RHD3) with a problem continuing till now and objectively confirmed even if currently mild. There is some evidence (Whitton & Polley, 2011; Sanes & Kotak, 2011) that auditory deficits even with these small effects on thresholds can affect auditory capacities beyond. Even unilateral hearing deficit can alter inter-aural time (ITT) and level difference (ILD) and masking level difference (MLD) changing perceptual acuity in the child, redirected synapses in the Superior Olivary Complex and inhibit signal clarity in the sensitive years for language development (Webster, 1977). The results for the communication items re-alert us to the fact that these fine differences in hearing of 5 dB are enough to give OM perceptual sequelae. The discrimination offered by the formula for 20 dB HL cut-off was highest with AUC-ROC, 0.899 with specificity of 73.2% at 90% sensitivity; at 90% sensitivity this gave specificity higher than for other higher cut offs. Moving the cut- off value for regarding as a case up to a higher criterion standard value must lower sensitivity, but would improve specificity and PPV, so this is an option in a wide health-economic context. However the data argue that the natural properties of the measures available make a 20 dB cut-off more appropriate for this type of combined-element screen, giving an evidence base for the best approach reconciling case-finding with preventing over-referral and cost per screen per child.

We are left with the issue of why RHD-type items perform usefully here, but have not performed well in (in terms of predicting HL) in some publications both on OM and SNHL. Possible explanations for that poor performance were presented in the introduction to this study and there are enough differences in study conditions to not regard the results as directly

in conflict. In one other study, the sensitivity of hearing questions to predict HL (pure tone audiometry) was unsurprisingly lower than sensitivity of the sweep test as realised in a community context; however their capability to predict eventual ENT decision was better than for the sweep test (Hind et al., 1999). This is consistent with the questions better reflecting a long-term problem, as suggested by the RHD3 results here. The better agreement between sweep test results and audiometry is probably due to the two similar tests occurring within a very short time interval, and the effective cut-off points being quite well aligned. Any URTI in the child at the time of testing could give disagreement between tests and questionnaire if this is not captured by a simultaneous URTI or ESS score. Similar results are found in other publications, with same cut off value > 30 dB of hearing test and sensitivity of questionnaires of 56%. The results remain consistent with the principle that with good selection of questions, with justified quantitative scoring and a sufficient number of questions for reliability, short questionnaires can illuminate auditory function. We have shown this in the application context of their contributing to prediction of hearing level. In OM, hearing prediction is not the prediction of ear state but the prediction of functioning; as such it may have relevance to decisions on treatment and may also have application to the evaluation of benefits from treatment. Even if more comprehensive instruments (i.e. more items so more reliable) are required to answer scientific questions about treatment benefits as in randomised controlled trials, short forms such as the present formulation can become a widely feasible standard for many purposes such as routine monitoring of patient outcomes.

6.0. Conclusions

6.1. Conclusions of Study I

The concept of seasonality in the occurrence of otitis media, most usually measured as case incidence, is widely present in the literature. Here however all research participants are clinical cases, (co)varying in severity, in ways that the analysis can make use of. Thus seasonality here is in the severity of various facets of OM: differing seasonal severity of upstream and downstream aspects of OM as a disease, and so different delay between facets in what has been nosologically construed in the past as acute versus chronic forms of the disease. The novelty of the present approach means that this Study I is more descriptive and exploratory than the others but it provides direct support for the canonical pathway with timings, hence strong causal inference. Consideration of seasonality helps us see the form of OM as only a label for a phase in a more general unfolding pattern. As illustrative support for this general conclusion and the improved understand that it represents, I have answered the specific research questions as follows:

6.1.1.a. Annual peak severity maximum of URTI and ESS: ESS (ear infection severity) in consulting cases shows peaks at 10.5 week after the synchronisation reference of first week in January, at around the middle of the March. This late winter peak is only a week after the URTI infection week severity (and 2 weeks after an unreliable, so less certain, peak in URTI obstruction symptoms). The delay in maximum ESS score severity in relation to URTI justifies a causal interpretation of the cascade, linked to incidence and prevalence OM in other studies (Bhutta, 2014). ESS maximum is 3-4 days before HL maximum: but without postulating a very direct causal link, this small interval could just be regarded as a coincidence within a generally similar set of delays expected for all upstream symptoms.

6.1.1.b. URTI and ESS relation to corresponding time delay to maximum severity for hearing measures: The severity peak for hearing measures HL and ACET is just half a week after ESS severity, in the middle of the March, 11 weeks from the reference week January and just 2-5 weeks before the peak in severity of overall hearing question, RHD4 and RHD 3 items. The seasonality of the objective hearing measures is among the strongest seen and presumably contributes to the delayed seasonality in the RHD and Impact and PQoL measures (Study III).

6.1.2. Item seasonalities for individual items within the RHD facet: RHD score severity peaks later than HL severity, by 2-5 weeks. Different items in RHD facets have different seasonalities. The first item, overall hearing rating, is the most seasonal, manifest only 2 weeks after HL severity, in late March or the beginning of the April. This is the most sensitive item in the pool, closest in correlation and temporal terms to HL, with the strongest effect size. The RHD-3 items in OMQ14 and RHD-4 items in the OM8-30 questionnaire have a later timing for their severity peaks and this scatter would tend to reduce the correlation with an HL measurement taken on the same day. The RHD-4 score peaks a week after the overall hearing rating item severity does, and a week before RHD 3 item severity. This is because it in effect averages the two values of delay of its two components mentioned. The item presents in the RHD but not presents in the OMQ14, 'asking you to repeat' is probably more seasonal than the rest of the items. RHD item's score severity extends from the April till the beginning of the July.

6.1.3. Phase delay in upstream and downstream disease aspects in relation to other knowledge: Weakness in any seasonality for annual severity of downstream variables is evident; the effect sizes are < 0.003 in partial eta-squared. Possible explanations are to hand: a) AOM is a self-limiting disease, resolving in the majority of children, but with accumulating impacts on a smaller subpopulation of children; b) The influence of upstream variables on downstream variables is not summarised by a single pathway, although the useful simplification into two (hearing and physical health) assists the explanation of weaker downstream seasonality; c) Not all upstream variables influence all downstream variables, their interrelations require further approach and research.

6.2. Conclusions of Study II

6.2.1. Similarities and differences in determinants of the three hearing measures: In a comprehensive examination of the determinants of severity in the three hearing measures, variables significant for all three hearing measures were: age, season, SES, history, diagnosis. Sex was not generally significant for any hearing measure. Generally there were no differences in pattern of findings between complete-data and maximum cases samples, except an age effect for RHD (not present in more complete data) and also that summer season (cosine of month) is significant only in the complete subset of data (Table 4.2.1. and Table 4.2.2.). Effect sizes for these influences on hearing measures reveal some differences between RHD and the objective measures; effect of SES and length of history are both stronger for RHD than they are for HL and ACET. The direction of these effects is such that children in the lower-SES families had worse RHD scores, as did those with longer histories. Considerably less variance was explained in HL by the determinants available, than was explained in ACET or RHD and this important finding could be considered a fundamental limitation in the use of HL measures alone.

6.2.2. Intercorrelation of the three hearing measures and their potential for substitution or totalling: Building on the inter-correlations and similarities of determinants, mediation analysis showed two forms of ACET influence on RHD, direct and indirect (mediated by HL), with strengths that are not very far from equal (standardised regression coefficients 0.203 and 0.157); this contrasts with great inequality when the stronger suppresses the weaker in a GLM (collinearity). The direct influence of ACET on RHD quantifies the extent to which the relation between ACET and RHD is not all mediated by HL. a) A total of all three measures together is more reliable as a reflection of what is meant by ‘hearing’ because they have complementary roles, and in any adding of correlated variables, random error reduces relative to common ‘signal’. b) ACET can substitute for a proportion of missing HL in research and clinical practice; as independent variable in the Eurotitis-2 sample with a case gain of +49.6%, the correlation value with RHD was reduced for the hybrid measure using ACET substitutions for missing HL, by the inclusion of the poorer information, but only from r = 0.36 to r = 0.33 (by about 2% of the variance explained).

6.2.3. Designs able to make direct inference about discrepancies between RHD and objective measures: Between scores from hearing questions and hearing level as the better

of the two objective hearing measures was attacked directly via two types of model for the effect of determinants. The first (covariance) approach had two equivalent variants, fitting HL as main covariate of interest and pre-regressing RHD on HL then working with the residuals. The second approach was taking a standardised difference between HL and RHD Z-diff, and then working with that as a more direct discrepancy measure. There was no overwhelming reason to prefer one approach for theoretical interpretability or overall variance explained. A difference in pattern was also seen, with age very strong in the Z-diff approach, but length of history very strong in the covariance approach, although the other variable had significant small to moderate effect size in each approach. Thus underlying the decorrelation between RHD and HL (generally above r = 0.3 but below r = 0.4) systematic discrepancies do exist, as well as random error. The systematic discrepancies are attributable to child age and length of OM history.

6.3. Conclusions of Study III

In summarising what the OMQ14 questionnaire offers, age, SES, history, HL, ACET and diagnosis should be considered as the major predictors of OM impact and overall OM child wellbeing. The mentioned variables are main risk factors or catalysts in causal relations of other risk factors in the complex canonical pathway of disease, in which ONE OR MORE underlying continuum can be expressed through patterns ascribed to separate diagnostic entities. The season effect is small in overall QoL score (totalPCQ14) but present with a broad peak arising from constituents with peaks separated in three month time, from late winter to early summer, probably with accumulation of ear symptoms severity score (ESSQ14) into impact measures, RHD and impactQ14.

6.3.1. Properties of factors scoring in OMQ14 and criterion validity: We can say that criterion validity for OMQ14 is high because high ($r > 0.90$) correlations are obtained with the corresponding scores from the much more completely sampled OM8-30 instrument having more than twice the number of items. This is true both, for factor scores and PC total.

6.3.2. Breakdown of the factor scores and determination by other variables in the database: When interpreting severity of disease profiles and impact of disease, the child's age, SES, history, diagnoses and season should typically be controlled for in analyses. Direction of interactions Age*history and SES* history are useful for understanding disease outcomes and prevention priorities, but the evidence from Eurotitis-2 is that these are not strong enough contributors to variance explained to justify fitting as control terms, except in a very large study able to support many independent variables.

6.3.3. Relation of factor scores to hearing measures after control of main determinants in question 2.3.2.: Hearing level and ACET are the strongest predictors of factor scores from OMQ14, explaining more of the variance (partial eta-squared 0.029) than the typically available background clinic information and bio-data. This is true independent of diagnosis, which has been controlled for in the analysis.

6.4. Conclusions of Study IV

6.4.1. The roles of a) ACET, b) hearing rating (HR) and c) 3 other RHD questions in predicting HL:

6.4.1.a. ACET is a very strong predictor of HL because that is the basis of its formulation. Its OR for predicting HL criterion, per unit increase in ACET, is slightly different between the two lower screen cut-offs considered, but this seems to be more a property of the cases and the position in the range than any discontinuity in the HL similar to the one around 30 dB. ACET predicts HL best in the range < 30 dB, as expected from the ACET distribution and from the ceiling effect around 30 dB, and it performs better in the model with centre adjusted (perhaps partly because of differences between centres in equipment calibration).

RHD questions supplement ACET's prediction of HL but within the 4 items we have two parts, differing in their contribution to HL prediction, but in a context-dependent way.

6.4.1.b. The rating item has an OR which mostly rises across increasing values of thee HL cut-off in the range examined and is the highest for 35 dB, though better in 25 dB prediction than 20 dB. The role of the hearing rating item is about equally strong in the models without and with centre adjusted (Table 4.4.1.).

6.4.1.c. Fortunately in the context of the logistic regressions simulating a screen, the PC score for the RHD3 communication items has a complementary zone of best contribution for 20 dB cut-off, although this was not entirely compatible with the form of the interaction seen previously in the GLM. Its OR per unit increase was higher for > 20 dB cut-off.

6.4.2. The roles of interactions; a) ACET*overall hearing rating question and b) ACET*3RHD items in HL prediction: The two types of hearing questions interplay with ACET in predicting HL in slightly differing ways;

6.4.2.a. Interaction between the hearing rating item (HR) and ACET is not significant in either the underlying GLM model, or in the logistic screen simulations. The HR complements ACET generally, but in a way that gives slightly better prediction for higher HLs. This appears to be due to a wide impairment range of coverage by the response levels offered in

the rating and it having less sensitivity in the lower threshold range better explained by ACET. HR item is complementary to ACET but not in a highly specific way.

6.4.2.b. The interaction between PC-RHD3 and ACET is significant in the underlying GLM, improves that model, and explains more HL variation at higher ACET. The interaction effect is small, and is present nearly equally in the model without and with centre adjusted. Interaction terms were however only marginally significant in these prediction models. Communication items are perhaps more sensitive to threshold fluctuations, raising the importance of good hearing in the milder range of HL, compared to the HR item. However small hearing variations will be better captured by the ACET/HL estimation technique.

6.4.3. The difference between 20 and 25 dB cut-offs: There is a slight difference between the two HL cut offs considered in detail, in their screening parameters: sensitivity, specificity, AUC-ROC and PPV. All of the first three parameters showed more favourable values at 20 dB (sensitivity 90.9%, specificity 73.2%, AUC-ROC .899) than at 25 dB cut-off (sensitivity 90%, specificity 69.3%, AUC-ROC .866). The two types of RHD questions seem to swap roles between these two cut-offs: for example the single rating item adds more predictability at the higher of these two threshold cutoffs ≥ 25 dB, while PC-RHD3 at the lower ≥ 20 dB. Empirically better sensitivity and specificity will give fewer false negatives and false positives, but PPV depends on the prevalence of disease. There is some uncertainty about what nominal OM(E) prevalence should be adopted for PPV calculation (considering age and season risk factors and the frequency of conditions worth treating on the basis of need, ability to benefit and need). However a working value of 10% point-prevalence in the main target age-range for a screening scenario lies near the middle of the range of various offered estimates, and does not over-medicalise a common condition. Given the obtained sensitivity/specificity trades, the PPV for this prevalence is of the order of 30%. This means that if other elements in an argument for introducing OM screening or active case triage for OM-related hearing loss were favourable, the PPV would be in an acceptable range also.

POSTSCRIPT

This thesis has examined, using general linear models as a powerful version of correlational analysis, the determinants of the clinical presentations of otitis media in children 3-8 years of age. It has for the first time generalised the known seasonal changes in this disease, from a matter of a gross winter maximum in incidence, into an evolving profile of relative severities of its facets, within the total severity of presentation (Study I). It has not used an accepted generic quality of life measure; however in the course of generating a criterion measure to document criterion validity of a short form multi-aspect questionnaire, it has generated (Study III) a new 14-item one for use in ORL: OM8-30 impact measure. This has good distributional properties, and already has a corresponding form within the short form OMQ-14 questionnaire. The thesis has probed (Study III) the fundamental issues in the measurement of hearing in OM and shown the distinctive properties of reported hearing difficulties (RHD) that distinguish RHD, as a useful and a serious measure. It has documented useful properties, despite obvious limitations, of the pre-existing ACET score, in which tympanometric ear states are mapped via their average HLs to the HL scale. At the same time it has shown a limitation to HL as the traditional objective measure widely used to define 'hearing' -- the fact that the determinants of the related ACET and RHD do not explain nearly as much variance in HL as in those other measures; this 'missing predictabilty' leads to the interpretation that HL has a rogue ingredient, possibly linked to its known fluctuation in milder cases, that limits is usefulness in practice. The documented properties of the hearing measures have led directly to a proposal for a low-cost screening or referral measure combing tympanometry (as ACET) with RHD measures which when applied to the existing database shows (Study IV) very promising screen performance parameters

Doing medical science incurs an obligation to disseminate useful findings. The four studies can stand individually and conventions in current medical scientific publishing will require that they do. However they also stand together and collectively have two properties useful for dissemination, especially when the two are present together. Firstly, the studies offer a novel and comprehensive concept of otitis media in which diagnosis is present but not dominant as that dominance can discard useful information for both research and practice. Secondly they offer useful tools. It remains to be seen whether the seasonal evolution of the facet profile can be applied clinically, but ways to find this out are fairly obvious and simple. The determinants of the hearing measures (Study II) offer a control framework for the

international standardisation of the measures, with negligible cost for the acquisition of the 6 determining variables, and the hybrid HL/ACET measure offers a now documented way to overcome difficulties arising with inevitably missing data in clinical decision or in formal imputation. There are now many new 'patient reported outcome measures' in ORL for diverse conditions, but none with a comparably large reference sample or psychometric lineage to that of OMQ-14, presented in Study III for the first time. Even if no jurisdiction ever introduces new school or pre-school screening based on the conjunction of RHD and ACET (Study IV), the idea of such a rational and evidence-based criterion for onward referral from healthcare settings not able to measure hearing or do otoscopy, to settings that are able, is extremely attractive. The implications for introducing this material to continuing medical education are clear.

Finally as part of the Eurotitis-2 collaboration, the four studies form a critical mass of internationally publishable work to lead publication of Eurotitis-2 results which include further topics such as diagnosis, to seek funding for a further study, Eurotitis-3 which might address evaluation issues in the clinical utility of the tools offered, and to establish use of the standardised metrics from the questionnaires, and to enable their use in international trials (such as the NIH cleft palate one in progress, which adopted OM8-30 as all the necessary translations existed) with a mapped quality of life measure either from the OM8-30 facet scores or the OMQ-14 factor scores.

APPENDIX I

Appendix to results on Seasonality & further examples of sinewave fit

This appendix gives more detail and comments on the sine (week) fitting procedure, and also seasonality charts for variables not given in the main text.

Due to the labour-intensive nature of the fitting of 27 delays in 27 models and of the data-management involved in the graphics, only variables showing some evidence of a seasonal pattern (see results Table 1.3. in the main text) were passed on to the plotting stage. The grain of the charts enables a judgement of how a monthly bar representation might capture all the gross annual fluctuation that exists or in some case might not, because the finer detail might not be reliable. This issue largely overlaps with the measurement or judgement of the general noise level, of weekly fluctuations around the medium-term average. The statistical test for magnitude coefficient of the sine fit at the particular delay giving maximum quantifies that judgment objectively. Whether the patterning seen is then sufficiently captured by a sinusoid is a separate issue, because we are mainly concerned with delayed correlation for causal inference, for which the sinusoid procedure is sufficient. No cosine term as used in Study II is needed in this procedure, because 27 starting times of the stored pattern for fitting to the data cover all possible delay. The negative of the sine at 90 degrees phase lag between the two functions (3 months or 13 weeks) is equivalent to fitting the cosine The 26 +1 for wrap-around is simply to handle half the year not being an integer number of weeks. The second half of the year is handled by reading negative signs – the shape of the second half of a sinusoid cycle is simply the reverse of the first half.

The first two charts are both for the infection sub-scale of respiratory (URTI) – 3 items, but one of them is just the other re-inverted, so high value is ‘bad’. An informal visual analysis is necessary for insights and for checking no error has been made; however it can be deceptive, because the human visual system and scientific conventions make search for sharp peaks a natural and practised process. They do not seem to contain the same information on overall annual pattern. When a peak becomes a dip or vice versa, errors of impression creep in, showing the need for the formal fitting procedure to avoid misjudgement with a visual origin, illustrated below in comparison of one pattern with its inverse.

Despite the reliability limitations in having only 3 questionnaire items, the infection variable has the strongest sine fit of those examined ($p < 0.00005$), with a suitable optimal delay found by its maximum in the series of 27 sine models. A medium-strong fit is seen for the ESS (related to RAOM) ($p = 0.005$). The weakest of those in this batch of significant effects is RHD-3 ($p = 0.028$). This gradation, with diminishing magnitude away from the initiating process, infection, in terms of the conceptual diagram in the main text is exactly what we would expect from where in the causal cascade each variable sits (upstream stronger). The strong overall annuality gets washed out. However visually, this is not how it appears to the unskilled eye, because there is much distracting visual interference at periodicities higher than the fundamental (period = 1 year). The difficulty is that this is likely to be just low-amplitude variability ('noise' in signal analysis terms) and so be error when estimating terms for the fundamental period and picking the peak delay among these. Hence there is a need for a formal fitting procedure.

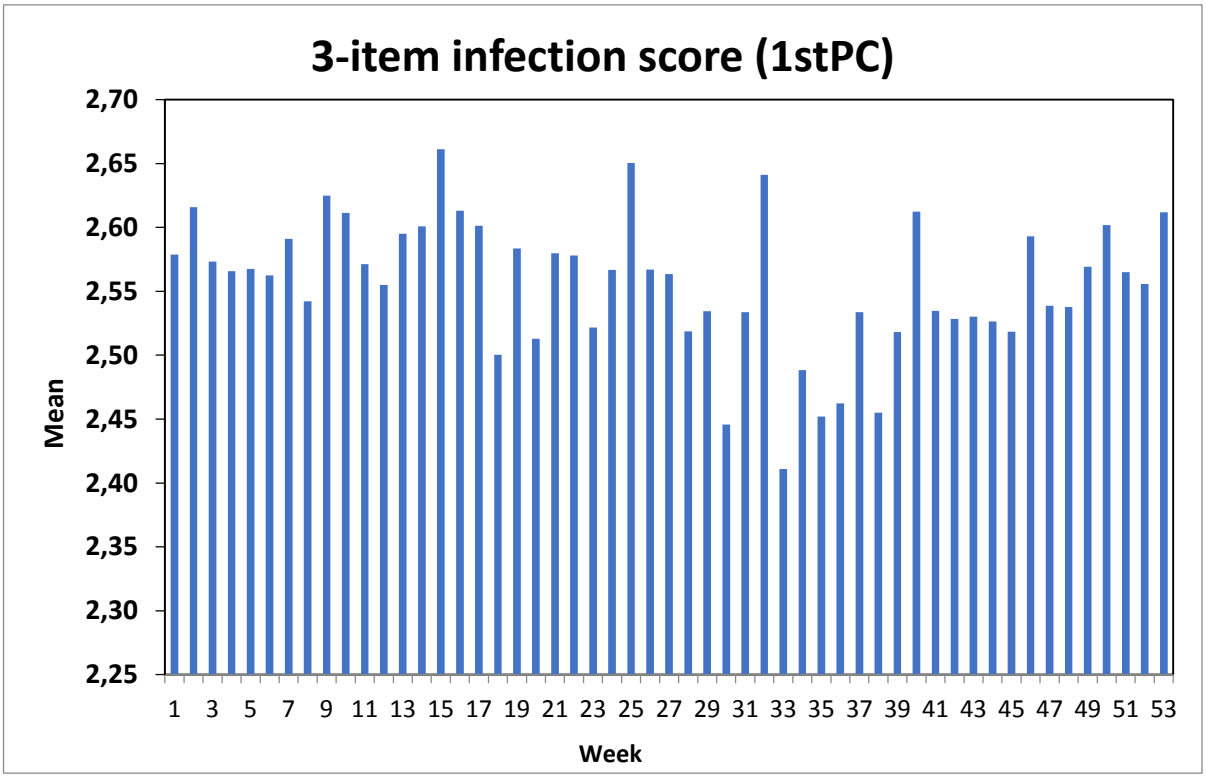

Figure.1. Scores of the PC-3 URTI-infection items at weekly intervals, with starting week 1st January.

***Notes**: High values are 'bad'. Peak is at late winter between 9-17 weeks with much semi-random oscillation from week to week (in effect, high-frequency noise).*

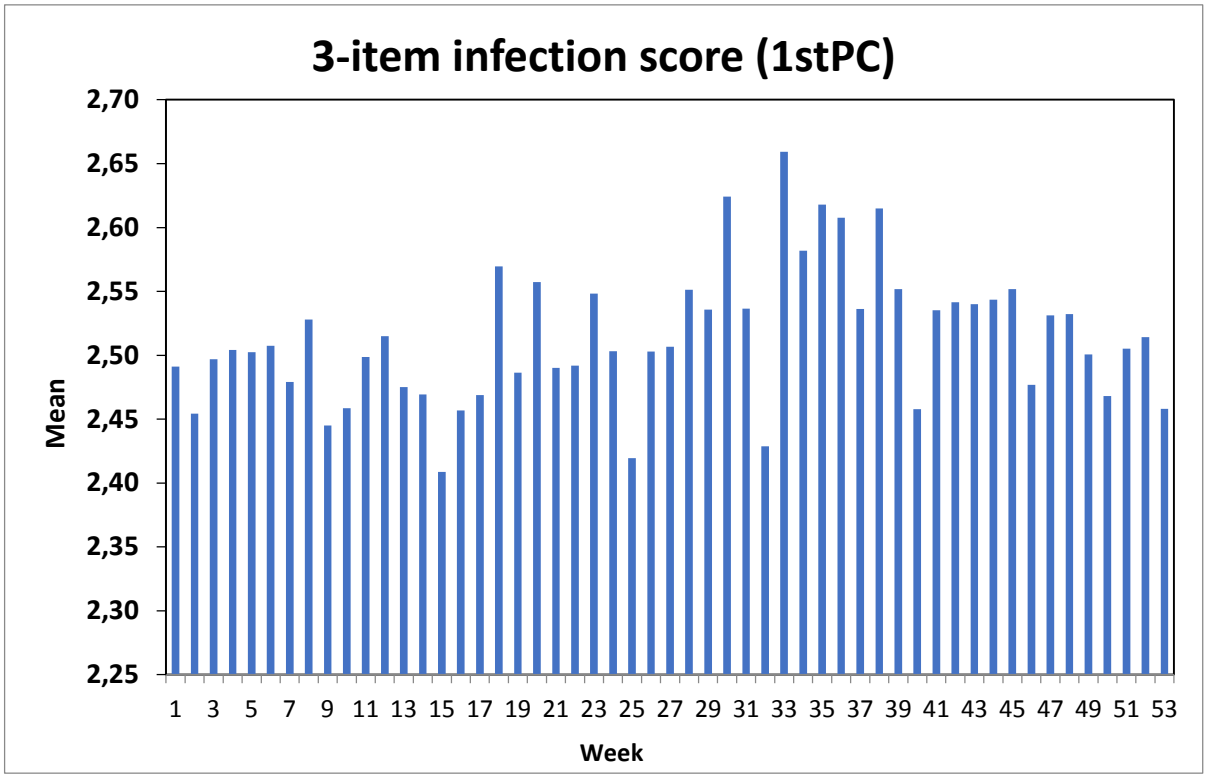

Figure.2. Same distribution as in Fg1 but the PC-3 URTI infection score is inverted: low is now 'bad'.

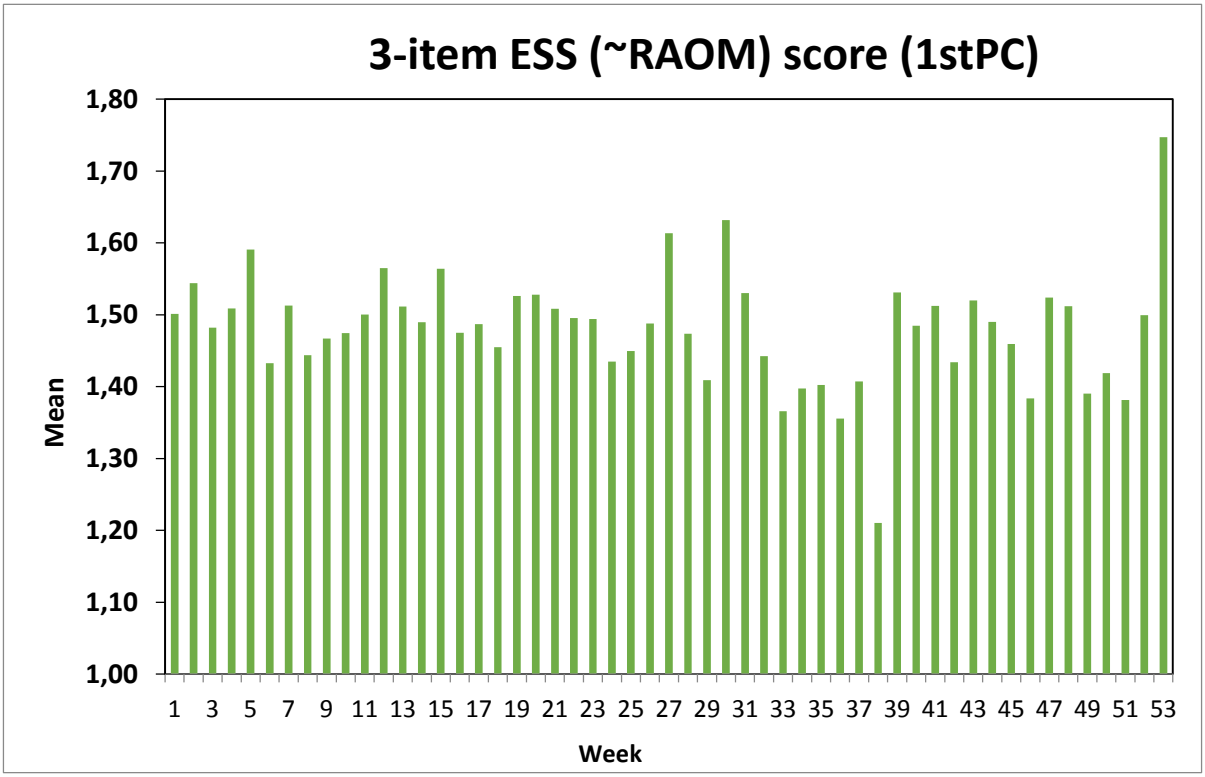

Figure.3. PC of 3-item ESS score peaking in 1st part of year, 'late winter' (early March, 10-16th calendar weeks).

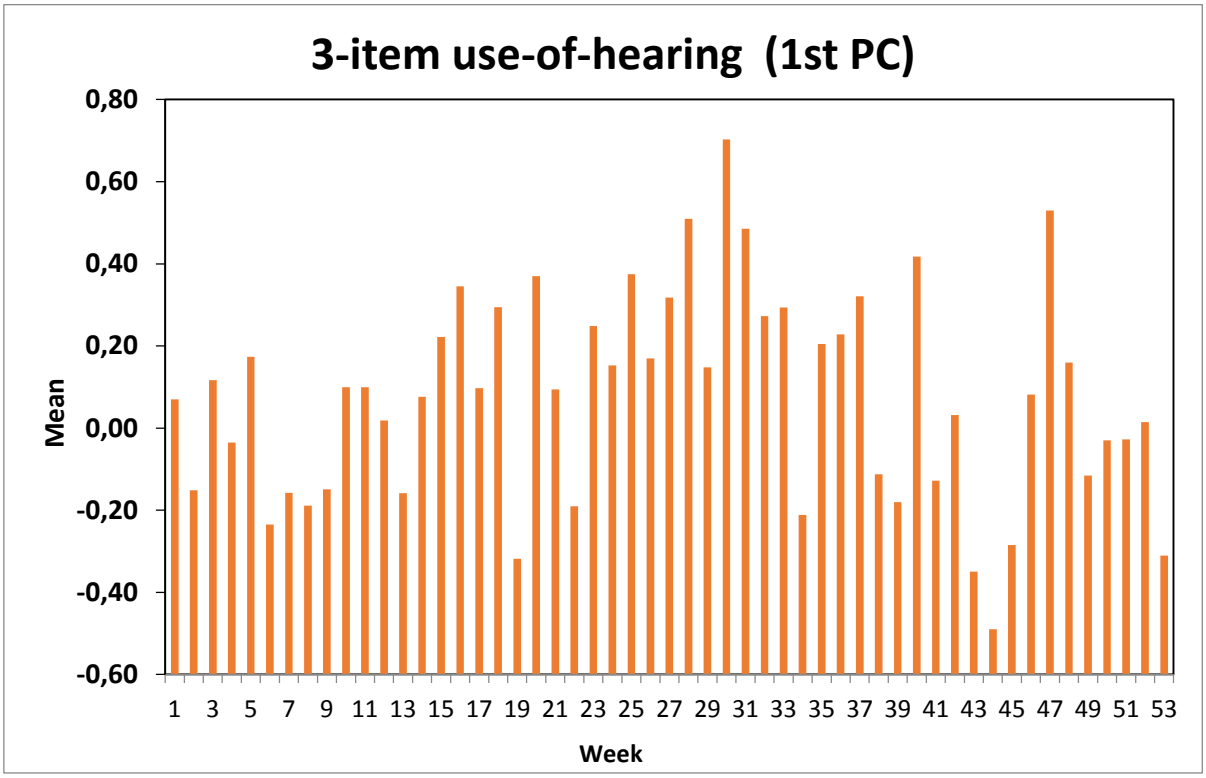

Figure.4. Weekly pattern for the PC of three use-of-hearing (communication) RHD items, showing increase in severity (high = 'bad') from 9th through 30th weeks, with late spring/early summer maximum.

The formal sinewave fitting procedure captures only the low-frequency ('long wave') structure. In Figure 4 is the statistically weakest of the 3 examples. The high-frequency i.e. week-to-week fluctuation is almost certainly mostly noise and can be disregarded by degrading the graphic (e.g. viewing through thin paper). However, it does not stretch the imagination, even though this example is marginal statistically (sinusoid $p = 0.028$), to see a high plateau from weeks 15 to 37 as corresponding to a close match with the negative cosine fit used previously. There is almost certainly no advantage in attempting any other degree of resolution in analysis e.g. 2-weekly, given that with the ability to add smoothing for graphical displays or via the fitting procedure, there is no loss in using weekly. The unit of 1 week is standard and understood.

It is much harder to be sure about the genuineness or noise status of the middle-frequency structure, and if it were to prove genuine, what its appropriate interpretation should be. Such middle-frequency structure is exemplified by dips around weeks 6-9 and 43-45 for RHD3 (or possibly even at weeks 37-45). From the point of view of the fit of sinusoids with a 1-year period, this remains handled as noise – to be considered error, and not part of the matched structural element. That does not mean that it has no structural validity at all -- just that we do not have a secure enough prediction of exactly where the rises and falls in this secondary patterning should be, except perhaps from contributions of school holiday periods. Some similar mid-frequency patterning is seen for infection but less for ESS, corresponding to RAOM diagnoses; there is perhaps a suggestion of reliable sub-annual patterning for infection and for RHD3. Such structure does not need to be rigidly periodic nor be close to multiples of the fundamental (i.e. integer fractions of the 1-year fundamental period.)

Incidence data show such minor dips during and briefly after school vacations with their reduction in children's exposure proximity to others in confined space (with few air changes per hour, for climatic reasons) is reduced. With delays of 1-2 months for RHD from the initiating infection score, identifying such reasons for the dips and shoulders seen in RHD is filled with uncertainties and difficulties of *post hoc* rationalisation. However for infection it is epidemiologically reasonable to predict reduction in severity as well as incidence due to school and nursery holidays and for a short period just after; this would occur from the 3rd week of December to the 3rd week in January, and again for a variable period in April. The eye of faith may see some evidence for this, but with most dips occurring for a single week it cannot be called convincing. Stronger fits could be had to these data for basis functions other than sinusoids, where the structure repeating annually is not sinusoidal in shape (i.e. the functions may have some 2nd, 3rd or 4th harmonic). But for the reasons cited against interpreting more complex patterns in monthly data, this could only be done by postulating such a function on half the data and testing it on the other half, or by some more complex computationally intensive re-sampling method extending this principle.

This thesis does not claim to have exhausted or even fully explored the issues in detailed delay fitting and viewing large-sample date data as a quasi-time-series (cross-sectional), only to have demonstrated a new method, showing promising coherence with expectations from pathogenetic understanding of sequences of signs and symptoms and with true time-series data.

APPENDIX II

Composition and distribution of 'impact' measures in OM8-30 and OMQ-14

This appendix illustrates some of the differences in scores derived from PC totals of discrete item sets (as in OM8-3) and from rotated factor scores using all items (OMQ-14). It also serves to specify the OM8-30 score for impact which has not been used before and was formulated in response to the need for a fuller criterion measure from OM8-30 to document the criterion validity of such a factor-based score in OMQ-14. With double the number of items of similar content, this fulfils the role of criterion measure for criterion validity. This work has the incidental useful by-product of showing that the OM Impact score in OM8-30 (1st PC of 14 items from all the items from the same facets as were sampled for OMQ-14 and emerged loading on the impact factor) has good distributional properties. The first pair of model residual distributions below (from models predicting these scores involving the full set of determining independent variables) shows essentially similar skew but slightly differing kurtosis. For OMQ14 impact there is skew 0.464 (SE .046) and kurtosis -0.233 (.091) and for OM8-30 the values are 0.469 (.053) and -0.431 (.105). The OM8-30 Impact distribution is more flat-topped and in this sense OMQ-14 has slightly the better of the two. Both samples are maximum cases with data in the models (N = 2,865 and 2,154 respectively). Model residuals are presented rather than raw distributions, to control for possible differences in sample composition (see Appendix on Balkan cases). The models adjust for centre, sex, SES, diagnosis (full 4-level version) length of history and age plus the three interactions SES*age, SES*history and age*history. For OM8-30 impact, although the two *SES interactions are only marginally significant; age*history is fairly strong (p = 0.000338; partial eta-squared 0.006), more so than is seen for RHD in Study II; the overall effect of length of history is also strong and in the expected direction (p = 0.000004; partial eta-squared 0.010). For OMQ-14 impact similar trends are seen, with similar model strength overall (adjusted Rsq 0.109) for OMQ-14 Impact factor and 0.100 for OM8-30 Impact aggregate. This suggests strongly that the development process for OMQ14 has captured much of the information available in the OM8-30 long form, complementing the validity correlations.

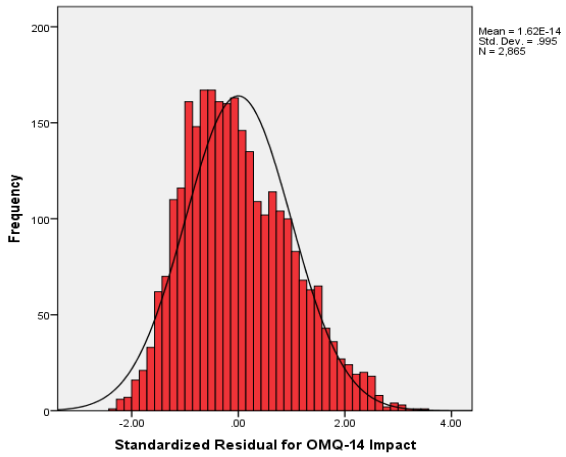

Figure.1. Residual distribution for impactQ14 score

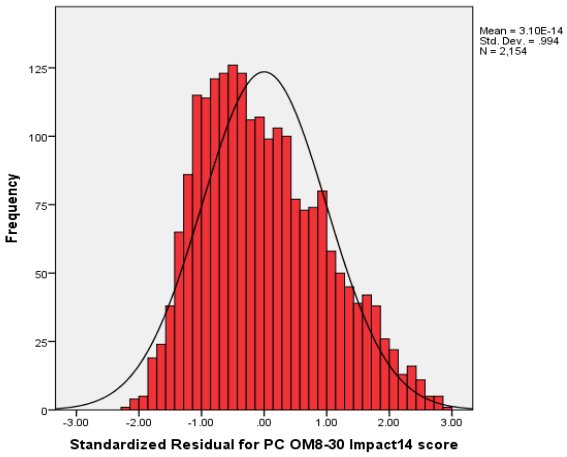

Figure.2. Distribution for OM8-30 impact PC score

Table 1. PC analyses and loadings on 2 components extracted for impact-related items in OM8-30

No	COMPONENT MATRIX	COMPONENT	
		1	2
1	Sitting still (e.g. at meal time, story time or at other times) he/she…	.485	-.174
2	**How often does he/she seek your attention unnecessarily?**	**.560**	**-.155**
3	How often does he/she whine or moan with little reason?	.584	-.193
4	**How often is he/she unhappy for no apparent reason?**	**.585**	**-214**
5	When you take him/her out somewhere, does he/she do what you ask?	.467	-.107
6	**How long can he/she concentrate on a game or task you have given him/her to do?**	**.441**	**.038**
7	**Has he/she mispronounced the beginnings or ends of words?**	**.535**	***.654***
8	**Has his/her speech been behind (less developed than) that of children of similar age?**	**.506**	***.705***
9	When trying to tell you something, does he/she have poor articulation?	.454	*.683*
10	**Have you often felt tired?**	**.507**	**-.223**
11	**Has your child needed more attention than other children?**	**.614**	**-.129**
12	Has your child been very demanding?	.625	-.273
13	Has it taken a lot of energy to cope?	.588	-.300
14	Would you agree that people wouldn't realise the effort involved until they had a child with ear and/or hearing problems?	.231	-.155

The two tables address whether the 14 impact items in OM8-30 would do better supporting two not one scores. Emboldened entries are items also in OMQ-14. The 2nd eigenvalue of 1.796 in Table 2 is well above the conventional minimum for postulating a 2nd

factor. Table 1 shows that the 2nd principle component is essentially the difference between the speech/language items and all others. Their high positive loadings (italicised) contrast with other loadings, which are negative, showing some inhomogeneity of the items. However separation of a speech/language factor on 14 items has two powerful arguments against it. It takes us back to the problem of reliability and validity with very few items that the impact aggregate is designed to avoid. Secondly the loadings on the first factor for the speech/language factor are not particularly low in the developmental impact aggregate, so something would be lost in reliability and generality by redefining the first factor as "impact apart from speech and language". The reasonable combined loadings suggest that the speech and language items may not be intrinsically poor, as was considered in interpreting their weak seasonality in study I, but rather that they lack some specificity of definition in the attribute(s) to which they refer, perhaps differing between different sub-populations and ideally requiring more than one factor solution and set of scoring formulae, e.g. broken down by age, an intriguing possibility beyond the present scope.

Table 2. PC eigenvalues and total variance explained for impact-related items in OM8-30

Component	Init. Eigenvalues		
	Total	% Var	Cum%
1	3.819	27.275	27.275
2	**1.796**	**12.827**	**40.102**
3	1.488	10.628	50.730
4	**1.182**	**8.446**	**59.177**
5	.822	5.869	65.046
6	**.753**	**5.379**	**70.425**
7	**.707**	**5.051**	**75.476**
8	**.647**	**4.621**	**80.097**
9	.594	4.241	84.339
10	**.511**	**3.648**	**87.986**
11	**.497**	**3.549**	**91.535**
12	.447	3.195	94.730
13	.389	2.780	97.510
14	.349	2.490	100.00

APPENDIX III

Specific details for cases from Serbia and Montenegro (Balkan Centres)

During the period between 26th February 2007 and 14th April 2014, data from five Balkan centres were collected as the region's contribution to Eurotitis-2. The Eurotitis-2 aim most relevant to centre differences is multicentre standardisation for OM data and optimisation of ways of recording them, to document covariation in symptom severities, and impacts. Balkan cases made nearly one third of all Eurotitis-2 cases (905/2865), and all Phase 2 cases (i.e. OMQ-14 only) came from the Balkans. As explained and defined in the General Method Chapter, three 'centres' are from Belgrade. The first two are from Clinical centre of Serbia, Clinic of ENT &Maxillofacial Surgery (Old Belgrade 1 centre with 325 cases and Old Belgrade 2 data with 325 cases), separated on the grounds of long time series and large sample size.. The second was the Institute for Mother and Child (New Belgrade 102). One centre was in Montenegro, the Clinical Centre of Podgorica (104) and one was a non-metropolitan Serbian Health Centre, Leskovac (44).

Table 1. PCOMQ14 total score in 905 Balkan cases from 2007 to 2014 year

Variables	P values	Direction of effect	Partial eta squared
Centre 11*	.114	-ve	.003
Centre 12*	.522	+ve	.000
Centre 17*	.014	-ve	.007
Centre 18*	.389	-ve	.001
Centre 19*	.	.	.
SES man/non man	.614/.143	.-ve/-ve	.000/.002
Diagnosis	.000	-ve	.015
D1	.000	+ve	.101
D2	.013	+ve	.007
D3	.	.	.
Length of history	.062	+ve	.004
age	.006	-ve	.009
Season/sine	.586	+ve	.000

Notes:

*The table shows the p-values for departure from reference centre 19 and partial η^2. Key to centre coding: Centre11 CCS Belgrade 1 Phase 1l *Centre 12 New Belgrade, Centre 17 CCS Belgrade 2, Centre18 Leskovac, and Centre19 Podgorica (Montenegro.) The centre differences are rather small compared to those in the study as whole. The adjusted Rsq is .174 (adjusted for df. 160).*

The difference between the two Old Belgrade 'centres' is in the questionnaires used. Old Belgrade in the phase distinguished as centre 1 used the OM830 questionnaire from 26/02/2007 to 30/12/2011, while Old Belgrade 2 centre used OMQ14 questionnaires from January 2012 till 12 March 2014. Verbal informed consent from parents was asked before examination of the child. Balkan cases have some specific characteristics that may be characteristic of the region – a topic that will be more fully explored outside this thesis, as the present aim is scientific generality. The questionnaires' characteristics and differences are presented in main Method of Study III as well as the statistical analyses and model which are the same as is explained under Method heading in Study III of this thesis.

The completeness of the data is globally slightly better than in the rest of the Europe with only 3.3% missing data, possibly due to the lower burden of the shorter questionnaire mostly used. The PC global variable (totalPC) is known from prior mapping work to be a fair approximation to generic quality of life of the child (QoL). The main independent variables used in the model are same as those used in Study III (age, gender, SES, history, diagnoses, season). The significance of variables and effect size and their direction are presented in Table 1. We did not find significant main (overall) effects of the independent variables (which can be seen as risk factors for severity) in the model except for two: diagnosis and age. Diagnosis is a very strong effect here, and in the direction that combined and OME cases (D1 and D2 diagnoses) have very strongly worse total score. The age effect is also strong, younger children being worse-affected. For the Balkans, the seasonal influence was not significant and history of OM problems marginal. The influence of centre is significant making it worth noting the directions, although not very strong (Table 1). New Belgrade and Montenegro cases are worse than others, while Old Belgrade 2 cases are the mildest. The effect of Old Belgrade 1 and Leskovac are of modest power. The difference between two Old Belgrade periods cases could be in part explained by a change of health policy and referral criteria published in 'Sluzbeni Glasnik' 06.01.2011.

Table 2. The difference between error variance of dependent variable between centres.

F	df1	df2	Sig.
1.593	43	861	.010

Notes:

Difference between groups (i.e. centres) is significant (Levene's test), meaning that the variation (e.g. range) also differs between centres but on such a large sample size very small differences in variance can be significant.

Using Bonferroni correction for minimising error rate and multiplying actual *p*-values by 10 (number of possible centre pairings) we obtain diagnosis still significant and age marginal. As often, it is not clear exactly what the Bonferroni multiplier should be as we would not particularly wish to explore particular pairings of centres as a scientific question. However 10 is close to the *df* for the centre variable plus its interactions and shows awareness of the need to be conservative if there is no particular hypothesis.

Comparing non-Balkan cases with Balkan cases using a dummy variable ('centre 1') we did not find large differences, but some particular points deserve attention. The difference between non Balkan cases (centre 1) in relation to Balkan cases (reference centre) is in the direction that other cases are worse than non-Balkan cases on PC total indicating relatively good access in the Balkans or relatively slight case triage in primary care. The most notable contrast between the tables is that history length within the Balkans is weak but in the remaining data it is very strong, a tenfold difference in partial eta squared. This would also be consistent with relatively good healthcare access. The general lack of significant interactions for Balkans/elsewhere with diagnosis suggests that the severity mildness difference is not specific to particular diagnoses, but with one exception where it is marginal, for the combined diagnosis ($p = 0.02$).

Table 3. Comparison of non Balkan cases and Balkan cases on OMQ14 PCtotal, p-values and $\acute{\eta}^2$.

Parameter	Sig.	95% Confidence Interval		Partial Eta Squared
		Lower Bound	Upper Bound	
Centre 1.*	.000	.317	.923	.006
Centre 2.	.	.	.	.
Sex missing	.000	-1.068	-.536	.012
Boys/girls	.236	-.027	.109	.000
SES missing	.125	-.036	.296	.001
SES manual/non manual	.000	.084	.223	.007
D0 (missing)	.000	-1.029	-.366	.006
D1 (combined cases)	.000	.361	.679	.014
D2 (OME)	.075	-.283	.014	.001
D3 (RAOM-reference)	.	.	.	.
Length of History	.000	.175	.248	.044
age	.000	-.009	-.005	.013
Season Sine	.021	.009	.106	.002
Centre 1*D0	.018	.089	.965	.002
Centre 1*D1§	.020	-.783	-.068	.002
Centre 1*D2	.592	-.239	.419	.000
Centre 1*D3	.	.	.	.
Centre2*D0	.	.	.	.
Centre2*D1	.	.	.	.
Centre 2*D2	.	.	.	.
Centre 2*D3	.	.	.	.

Notes:

**centre 1-non Balkan cases; centre2-Balkan cases;*

§ diagnoses D1 relatively worse on totalPC score in the Balkan cases.

Inspecting the direction of this interaction, the Balkan cases are generally milder than in rest of the Europe but children with combined diagnosis seem relatively worse in the Balkans than in other centres, so not milder in the way seen for other diagnoses. With 4 such comparisons, this finding does not survive Bonferroni correction (P → 0.08) and would need replication before making any strong claim. Changes in specifics of the referral policies in the Balkan region plus other local developments could contribute to the explanation of such a finding, so it is worth mentioning as exploratory to encourage replication to be sought. Professional discussion of the result and the use of this category could address the question: is it specifically informative diagnostically or is it being used in the region as a surrogate marker for general severity? The advantage of the large Balkan sample is that it may be possible to answer such questions using the database as resource. A specific publication on correlates of diagnosis in Eurotitis-2 will inform such discussions within and beyond the Balkans.

APPENDIX IV

Analysis of determinants of all scores in OMQ-14

This appendix provides summary data, emphasising partial $\acute{\eta}^2$, for total PC score from OMQ-14 and also for the three constituent factors. This is shown for two samples, the maximum OMQ-14 dataset of 2,865 and 1,866 cases with OMQ-14 and full hearing measures. Note that although the general concept of complete-data cases is the same as for the 1,400 complete cases in the analysis (Study II) of RHD and other hearing measures where a fourth RHD item from OM8-30 was required, the details and the numbers are different here. Partial eta squared is slightly stronger for most variables when comparing each of the supported OMQ-14 scores as dependent variables in the three constituent models and in the same direction. The SES effect (for PC total) is considerably stronger (partial $\acute{\eta}^2$ = 0.016 versus 0.008, comparing Table 1 with Table 2) in the better controlled subsample of complete cases than it is in maximum cases. So also is the effect of length of history, almost certainly for reasons of selection for cases in the TARGET sub-sample with longer histories, more hearing problems and fuller data, as discussed in the General Methods Chapter and elsewhere. A few smaller differences appear slightly to favour the larger sample, for which its greater numerical power would be sufficient explanation. The presence of a generally similar corresponding set of significant variables, across all models as in the maximal cases, counterpart confirms that the 1,866 sample achieves generally equivalent power to the larger sample, due to better measurement, as well as offering better control through completeness and inclusion of objective hearing measures. The difference between the two samples is restricted and relates chiefly to history and SES. The criterion validity correlations had to be run on the smaller set of complete-data cases aligning with use of the full OM8-30 item set and these analyses show that it is not seriously unrepresentative of the larger maximum-cases sample.

Table 1 Partial eta squared for OMQ-14 score (Total PC and the three rotated factors) in complete-data cases (N =1,866). Effects significant at p = 0.05 are in bold. The relation (second table) between HL level on a single concurrent occasion and general impact is very weak

	totalPCQ14	**Partial $\dot{\eta}^2$**	**ESSQ14**	**Partial $\dot{\eta}^2$**	**RHDQ14**	**Partial $\dot{\eta}^2$**	**impactQ14**	**Partial $\dot{\eta}^2$**
SES	Manual worse	**0.016**	Manual worse	0.0047	Manual worse	0.0020	Manual worse	**0.011**
Age	-ve: Older better	**0.010**	NS		+ve: Older worse	**0.005**	-ve: Older better	**0.042**
Sex	NS		Males better	0.0022	NS		Males worse	0.0030
History	+ve: Long worse	**0.031**	NS		+ve: Long worse	**0.030**	+ve: Long worse	**0.009**
Sine	+ve: Late winter max	0.0019	+ve: Late winter max	0.0022	NS		NS	
Cosine	-ve: Late spring max	**0.005**	-ve: Late spring max	0.0030	-ve: Late spring max	**0.0015**	-ve: Late spring max	0.0019
Diagnosis Missing	Better	**0.009**	Better	**0.039**	Better	0.0012	NS #	
Diagnosis Combined	Worse	**0.012**	Worse	0.0040	Worse	**0.014**	NS	
Diagnosis OME	Better	0.0032	Better	**0.077**	Worse	**0.015**	NS	

	totalPCQ14	**partial $\dot{\eta}^2$**	**ESSQ14**	**partial $\dot{\eta}^2$**	**RHDQ14**	**partial $\dot{\eta}^2$**	**impactQ14**	**partial $\dot{\eta}^2$**
HL	+ve: high worse	**0.029**	NS		+ve: high worse	**0.075**	+ve: high worse	0.0020
ACET	+ve: high worse	0.0049	+ve: high worse	**0.005**	+ve: high worse	**0.005**	NS	

***Notes**:*

 # For ImpactQ14, diagnosis is not significant overall and so dropped in the final model, and shown as non-significant in this table. However the missing category; level differs marginally from reference at p = 0.073 with partial $\dot{\eta}^2$ = 0.0017 in the model just before diagnosis drops out.

Table 2 Partial eta-squared for determinant variables in the maximum–cases sample (N = 2,865 cases), for total for PCQ14 and the three constituent rotated factors. HL and ACET contributions are not shown as they would be identical to those on the 1866 which have the relevant variables

	totalPCQ14	**Partial $\dot{\eta}^2$**	**ESSQ14**	**Partial $\dot{\eta}^2$**	**RHDQ14**	**Partial $\dot{\eta}^2$**	**impactQ14**	**Partial $\dot{\eta}^2$**
SES	Manual worse	**0.008**	Manual worse	0.0026	Manual worse	0.0002	Manual worse	**0.007**
Age	-ve: Older better	**0.012**	NS		+ve: Older worse	**0.011**	-ve: Older better	**0.059**
Sex	NS		NS		NS		Males worse	0.0035
History	+ve: Long worse	**0.023**	NS		+ve: Long worse	**0.019**	+ve: Long worse	**0.009**
Sine	+ve: Late winter max	0.0018	+ve: Late winter max	0.0042	NS		NS	
Cosine	-ve: Late spring max	0.0011	NS		NS		NS	
Diagnosis Missing	Better	**0.006**	Better	**0.029**	Better	0.0002	Worse	0.0015
Diagnosis Combined	Worse	**0.010**	Worse	0.0025	Worse	**0.010**	Worse	0.0010
Diagnosis OME	Better	0.0018	Better	**0.068**	Worse	**0.015**	Worse	0.0021

REFERENCES

Acuin, I. J. (2004). *Chronic suppurative otitis media: burden of ilness and management options.* Geneve: World Health Organization.

American Academy of Family Physicians. (2004). Otitis media with effusion. *Pediatrics, 113*(5), 1412 -1429.

American Academy of Pediatrics. (2013). The Diagnosis and Management of Acute Otitis Media. *Pediatrics*, e964.

American Speech-Language-Hearing Association (ASHA). (1997). Guidelines for screening infants and children for outer and middle ear disorders, birth through 18 years. In *Guidelines for audiologic screening* (pp. 15–22). Rockville, MD: Author.

Bardach, R. W., Buffler, P. A., & O'Donell, A. A. (1980). Epidemiology of otitis media: a community study. *American Journal of Public Health, 70*(6), 593-598.

Bardach, A., Ciapponi, A., Garcia-Marti, S., Glujovsky, D., Mazzoni, A., Fayad, A., Colindres, R.E., & Gentile, A. (2011). Epidemiology of acute otitis media in children of Latin America and the Caribbean: A systematic review and meta-analysis. *International Journal of Pediatric Otorhinolaryngology, 75*(9), 1062–1070. doi:10.1016/j.ijporl.2011.05.014

Bennett, K. E., & Haggard, M. P. (1999). Behaviour and cognitive outcomes from middle ear disease. *Archives of Disease in Childhood, 80*(1), 28–35. doi:10.1136/adc.80.1.28

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

Bess, F. H. & Humes, L. E. (2003). *Audiology: The Fundamentals.* Baltimore: Williams and Wilkins.

Biles, R. W., Buffler, P. A., & O'Donell, A. A. (1980). Epidemiology of otitis media: A community study. *American Journal of Public Health, 70*(6), 593–598. doi:10.2105/AJPH.70.6.593

Bloom-Feshbach, K., Alonso, W. J., Charu, V., Tamerius, J., Simonsen, L., Miller, M. A., & Viboud, C. (2013). Latitudinal variations in seasonal activity of influenza and respiratory syncytial virus (RSV): a global comparative review. *PLoS ONE, 8*(2), 3–4. doi:10.1371/journal.pone.0054445

Bluestone, C. D. (2004). Studies in otitis media: children's hospital of Pittsburgh-University of Pittsburgh progress report-2004. *The Laryngoscope, 114*(S105), 1–26. doi:10.1097/01.mlg.0000148223.45374.ec

Bluestone, C. D., & Klein. J. O. (2007). *Otitis Media in Infants and Children.* Philadelphia: PMPH-USA.

Bhutta, M. F. (2014). Epidemiology and pathogenesis of otitis media: Construction of a phenotype landscape. *Audiology and Neurotology*, *19*(3), 210–223. doi:10.1159/000358549

Bowerman, M. (1985). Beyond communicative adequacy: From piecemeal knowledge to an integrated system in the child's acquisition of language. In *Children's language* (pp. 369-398). Lawrence Erlbaum.

Brockson, M. E., Novotny, L. A., Jurcisek, J. A., McGillivary, G., Bowers, M. R., & Bakaletz, L. O. (2012). Respiratory syncytial virus promotes moraxella catarrhalis-induced ascending experimental otitis media. *PLoS ONE*, *7*(6). doi:10.1371/journal.pone.0040088

Brouwer, C. N. M., Schilder, A. G. M., van Stel, H. F., Rovers, M. M., Veenhoven, R. H., Grobbee, D. E., Grobbee, D. E., Sanders, E. A. M., & Maillé, A. R. (2007). Reliability and validity of functional health status and health-related quality of life questionnaires in children with recurrent acute otitis media. *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, *16*(8), 1357–1373. doi:10.1007/s11136-007-9242-0

Brody, R., Rosenfeld, R. M., Goldsmith, A. J., & Madell, J. R. (1999). Parents cannot detect mild hearing loss in children. *Otolaryngology - Head and Neck Surgery*, *121*(6), 681–686. doi:10.1053/hn.1999.v121.a101041

Butler, C. C., van der Linden, M. K., MacMillan, H. L., & van der Wouden, J. C. (2003). Should children be screened to undergo early treatment for otitis media with effusion? A systematic review of randomized trials. *Child: Care, Health & Development*, *29*(6), 425–432.

Calderon, R. (2000). Parental Involvement in deaf children's education programs as a predictor of child's language, early reading, and social-emotional development. *Journal of Deaf Studies and Deaf Education*, *5*(2), 140–155. doi:10.1093/deafed/5.2.140

Canadian Task Force on Preventive Health Care. (2001). Screening for otitis media with effusion: recommendation statement from the Canadian Task Force on Preventive Health Care. *CMAJ, 165*(8), 1092-1093.

Casselbrant, M. L., Mandel, E. M. (1999). Epidemiology. In R. M. Rosenfeld, & RM, C. D. Bluestone (Eds). *Evidence-based otitis media* (pp. 117-136). Hamilton, ON: BC Decker.

Cayé-Thomasen, P., Hermansson, A., Bakaletz, L., Hellstrøm, S., Kanzaki, S., Kerschner, J., Lim, D., Lin, J., Mason, K., & Spratley, J. (2013). Panel 3: Recent advances in anatomy, pathology, and cell biology in relation to otitis media pathogenesis. *Otolaryngology--Head and Neck Surgery*, *148*(4), E37–51. doi:10.1177/0194599813476257

Chiu, S. S., Lau, Y. L., Chan, K. H., Wong, W. H. S., & Peiris, J. S. M. (2002). Influenza-related hospitalizations among children in Hong Kong. *The New England Journal of Medicine*, *347*(26), 2097–2103. doi:10.1056/NEJMoa020546

Chonmaitree, T., & Heikkinen, T. (1997). Role of Viruses in Middle-ear Diseasea. *Annals of the New York Academy of Sciences, 830*(1), 143-157.

Chonmaitree, T., Revai, K., Grady, J. J., Clos, A., Patel, J. A, Nair, S., Nair, S., Fan. J., & Henrickson, K .J. (2008). Viral upper respiratory tract infection and otitis media complication in young children. *Clinical Infectious Disease: An Official Publication of the Infectious Diseases Society of America*, *46*(6), 815–823. doi:10.1086/528685

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd Ed.). Hillsdale, NJ: Erlbaum.

Cohen, S., Janicki-Deverts, D., Turner, R. B., Marsland, A. L., Casselbrant, M. L., Li-Korotky, H. S., Epel, E. S., & Doyle, W. J. (2013). Childhood socioeconomic status, telomere length, and susceptibility to upper respiratory infection. *Brain, behavior, and immunity*, *34*, 31-38.

Dakin, H., Petrou, S., Haggard, M., Benge, S., & Williamson, I. (2010). Mapping analyses to estimate health utilities based on responses to the OM8-30 otitis media questionnaire. *Quality of Life Research*, *19*(1), 65–80. doi:10.1007/s11136-009-9558-z

Daly, K. A., Brown, J. E., Lindgren, B. R., Meland, M. H., Le, C. T., & Giebink, G. S. (1999). Epidemiology of otitis media onset by six months of age. *Pediatrics, 103*(6), 1158-1166.

Davis, A., Bamford, J., Wilson, I., Ramkalawan, T., Forshaw, M., & Wright, S. (1996). A critical review of the role of neonatal hearing screening in the detection of congenital hearing impairment. *Health technology assessment* (Winchester, England), 1(10), i-iv.

Davis, E., Waters, E., Mackinnon, A., Reddihough, D., Kerr Graham, H., Mehmet-Radji, O., & Boyd, R. (2006). Paediatric quality of life instruments: a review of the impact of the conceptual framework on outcomes. *Developmental Medicine & Child Neurology*, *48*(04), 311. doi:10.1017/S0012162206000673

Dopico, X. C., Evangelou, M., Ferreira, R. C., Guo, H., Pekalski, M. L., Smyth, D. J., Cooper, N., Burren, O.S., Fulford, A.J., Hennig, B.J., Prentice, A.M., Ziegler, A., Ezio Bonifacio, E., Wallace, C., & Todd, J. A. (2015). Widespread seasonal gene expression reveals annual differences in human immunity and physiology. *Nature communications*, 6.

Dempster, J. H., & MacKenzie, K. (1991). Tympanometry in the detection of hearing impairments associated with otitis media with effusion. *Clinical Otolaryngology*, *16*(2), 157–159. doi:10.1111/j.1365-2273.1991.tb01967.x

Dettman, S. J., Pinder, D., Briggs, R. J. S., Dowell, R. C., & Leigh, J. R. (2007). Communication fevelopment in children who receive the cochlear implant younger than

12 months: risks versus benefits. *Ear and Hearing*, *28*, 11S–18S. doi:10.1097/AUD.0b013e31803153f8

Drummond, M. F., O'Brien, B., Stoddart, G. L., & Torrance, G. W. (1997). Basic types of economic evaluation. In M. F. Drummond, M. J. Sculpher, G. W. Torrance, B. J. O'Brien, & G. L. Stoddart (Eds.). *Methods for the economic evaluation of health care programmes* (pp.7-26). Oxford: Oxford University Press.

Dutt, S. N., Mcdermott, A., Jelbert, A., Reid, A. P., & Proops, D. W. (2002). The Glasgow benefit inventory in the evaluation of patient satisfaction with the bone-anchored hearing aid: quality of life issues. *The Journal of Laryngology & Otology*, *116*(28), 7–14.

Efron, B., & Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, *1*(1), 54–77.

Eiser, C., Mohay, H., & Morse, R. (2000). The measurement of quality of life in young children. *Child: Care, Health and Development*, *26*(5), 401–414. doi:10.1046/j.1365-2214.2000.00154.x

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

Fiellau-Nikolajsen, M., Lous, J., Pedersen, S. V., & Schousboe, H. H. (1977). Tympanometry in Three-Year-Old Children I. A Regional Prevalence Study on the Distribution of Tympanometric Results in a Non-selected Population of 3-Year-Old Children. *Scandinavian audiology*, *6*(4), 199-204.

Fidika, A., Salewski, C., & Goldbeck, L. (2013). Quality of life among parents of children with phenylketonuria (PKU). *Health and Quality of Life Outcomes*, *11*(1), 54. doi:10.1186/1477-7525-11-54

Field, D. L., & Haggard, M. P. (1989). Knowledge of hearing tactics: (I) Assessment by questionnaire and inventory. *British Journal of Audiology*, *23*(4), 349-354.

Filipovic, S. A, Haggard. M. P., Spencer, S., Zirk-Sadowksi, J., Milovanovic, J., Marchisio, P., & Eurotitis-2 Study Group. (2013). How physical and developmental aspects of OME influence parent quality of life. Presentation to *International Symposium on Otiitis Media Stockholm*.

Flynn, T., Möller, C., Jönsson, R., & Lohmander, A. (2009). The high prevalence of otitis media with effusion in children with cleft lip and palate as compared to children without clefts. *International journal of pediatric otorhinolaryngology*, *73*(10), 1441-1446.

Folsom, R. C., Weber, B. A., & Thompson, G. (1983). Auditory brainstem responses in children with early recurrent middle ear disease. *Annals of Otology, Rhinology & Laryngology*, *92*(3), 249-253.

Geers, A. E., & Nicholas, J. G. (2013). Enduring Advantages of Early Cochlear Implantation for Spoken Language Development. *Journal of Speech Language and Hearing Research, 56*(2), 643–653. doi:10.1044/1092-4388(2012/11-0347)

Giebink, G. S., Berzins, I. K., Marker, S. C., & Schiffman, G. (1980). Experimental otitis media after nasal inoculation of Streptococcus pneumoniae and influenza A virus in chinchillas. *Infection and Immunity, 30*(2), 445–450.

Grijalva, C. G., Nuorti, J. P., & Griffin, M. R. (2009). Antibiotic prescription rates for acute respiratory tract infections in US ambulatory settings. *Jama, 302*(7), 758-766.

Haggard, M., & Hughes, E. (1988). *Objectives, values and methods of screening children's hearing: a review of the literature.* London, England: Her Majesty's Stationery Office, 1991.

Haggard, M., Spencer, H., & Gregori, M. (2007). Internationally standardised clinical description: can it be achieved in ENT. *ENT News, 16*(1), 69-72.

Haggard, M., & MRC Multi-centre Otitis Media Study Group (2009). Air-conduction estimated from tympanometry (ACET): 2. The use of hearing level-ACET discrepancy (HAD) to determine appropriate use of bone-conduction tests in identifying permanent and mixed impairments. *International Journal of Pediatric Otorhinolaryngology, 73*(1), 43–55. doi:10.1016/j.ijporl.2008.09.023

Haggard, M. P., Smith, S. C., Nicholls, E. E. (2003) Quality of Life and Child Behavior. In R. M. Rosenfeld & C. D. Bluestone (Eds.). *Evidence-based otitis media* (pp. 400-428). B.C. Decker, Hamilton, Ontario.

Haggard, M. P., Zirk-Sadowski, J., Doniec, K. J., Filipovic, S. A., Trzpis, K., & Eurotitis-2 Study Group. (2015). Comprehensive evidence-based model for developmenmtal impact, linking facets in OM. Invited presentation to International *Society for Otitis Media, National Harbour*, Maryland, USA.

Haggard, M. P., Filipovic, S. A., Zhang, M. F., Spencer, H., & MRC Multi-centre Otitis Media Study Group. (in preparation). Precision-scored parental report questions and HL-scaled tympanometry as informative measures of hearing in otitis media 2: evidence-based indication for treatment beneift in a large randomised trial

Hall, A. J., Maw, R., Midgley, E., Golding, J., & Steer, C. (2014). Glue ear, hearing loss and IQ: An association moderated by the child's home environment. *PLoS ONE, 9*(2). doi: 10.1371/journal.pone.0087021

Harkness, P., & Topham, J. (1998). Classification of otitis media. *The Laryngoscope, 108*(10), 1539-1543.

Hammond, P. D., Gold, M. S., Wigg, N. R., & Volkmer, R. E. (1997). Preschool hearing screening: evaluation of a parental questionnaire. *Journal of Paediatrics and Child Health, 33*(6), 528–530.

Hartman, M., Rovers, M. M., Ingels, K., Zielhuis, G. A., Severens, J. L., & van der Wilt, G. J. (2001). Economic evaluation of ventilation tubes in otitis media with effusion. *Archives of Otolaryngology–Head & Neck Surgery, 127*(12), 1471-1476.

Hayes, A. F. (2013). *Introduction to Mediation, Moderation, and Conditional Process Analysis.* New York: The Guilford Press.

Heidemann, C. H., Godballe, C., Kjeldsen, A. D., Johansen, E. C. J., Faber, C. E., & Lauridsen, H. H. (2013). The Otitis Media-6 questionnaire: psychometric properties with emphasis on factor structure and interpretability. *Health and Quality of Life Outcomes, 11*(1), 201. doi:10.1186/1477-7525-11-201

Heikkinen, T., & Ruuskanen, O. (1995). Signs and symptoms predicting acute otitis media. *Archives of Pediatrics & Adolescent Medicine, 149*(1), 26–29.

Heikkinen, T., & Chonmaitree, T. (2003). Importance of respiratory viruses in acute otitis media. *Clinical Microbiology Reviews, 16*(2), 230–241. doi:10.1128/CMR.16.2.230-241.2003

Heinrich, J., & Raghuyamshi, V. S. (2004). Air pollution and otitis media: a review of evidence from epidemiologic studies. *Current allergy and asthma reports, 4*(4), 302-309.

Higson, J. & Haggard, M. (2008). An extension of the Jerger Classification of Tympanograms for Ventilation Tube Patency-Specification and Evaluation of Equivalent Ear-Canal Volume Criteria. *Ear & Hearing, 29*, 1-14.

Hind, S. E., Atkins, R. L., Haggard, M. P., Brady, D., & Grinham, G. (1999). Alternatives in screening at school entry: comparison of the childhood middle ear disease and hearing questionnaire (CMEDHQ) and the pure tone sweep test. *British Journal of Audiology, 33*(6), 403–414. doi:10.3109/03005364000000108

Ho, V., Daly, K. A, Hunter, L. L., & Davey, C. (2002). Otoacoustic emissions and tympanometry screening among 0-5 year olds. *The Laryngoscope, 112*(3), 513–519.

Hoa, M., Tomovic, S., Nistico, L., Hall-Stoodley, L., Stoodley, P., Sachdeva, L., Berk, R., & Coticchia, J. M. (2009). Identification of adenoid biofilms with middle ear pathogens in otitis-prone children utilizing SEM and FISH. *International journal of pediatric otorhinolaryngology, 73*(9), 1242-1248.

Hoberman, A., Marchant, C. D., Kaplan, S. L., & Feldman, S. (2002). Treatment of acute otitis media consensus recommendations. *Clinical pediatrics, 41*(6), 373-390.

Hoffman, H. J., Daly, K. A., Bainbridge, K. E., Casselbrant, M. L., Homøe, P., Kvestad, E., Kvaerner, K. J., & Vernacchio, L. (2013). Panel 1: Epidemiology, Natural History, and Risk Factors. *Otolaryngology--Head and Neck Surgery, 148*(4), E1-E25.

Hogan, S. C. M., & Moore, D. R. (2003). Impaired binaural hearing in children produced by a threshold level of middle ear disease. *JARO - Journal of the Association for Research in Otolaryngology*, *4*(2), 123–129. doi:10.1007/s10162-002-3007-9

Horsman, J., Furlong, W., Feeny, D., & Torrance, G. (2003). The Health Utilities Index (HUI): concepts, measurement properties and applications. *Health and Quality of Life Outcomes*, 1, 54. doi:10.1186/1477-7525-1-54

Hotomi, M., Sakai, K. F. A., Billal, D. S., Shimada, J., Suzumoto, M., & Yamanaka, N. (2006). Antimicrobial resistance in Haemophilus influenzae isolated from the nasopharynx among Japanese children with acute otitis media. *Actaotolaryngologica*, *126*(2), 130-137.

Howie, V. M., Ploussard, J. H., & Sloyer, J. (1975). The otitis-prone condition. *American Journal of Diseases of Children*, *129*(6), 676-678.

IBM Corp. Released 2013. *IBM SPSS Statistics for Windows, Version 22.0.* Armonk, NY: IBM Corp.

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, *2*(8), 0696–0701. doi:10.1371/journal.pmed.0020124

Jerger, J. (1970). Clinical experience with impedance audiometry. *Archives of Otolaryngology - Head and Neck Surgery*, *92*(4), 311–324. doi:10.1001/archotol.1970.04310040005002

Julien, G. I. L. L. E. S., Baxter, J. D., Crago, M. A. R. T. H. A., Ilecki, H. J., & Therien, F. R. A. N. Ç. O. I. S. (1987). Chronic otitis media and hearing deficit among native children of Kuujjuaraapik (Northern Quebec): a pilot project. *Canadian Journal of Public Health*, *78*(1), 57.

Kaczor, J. J., Ziolkowski, W., Popinigis, J., & Tarnopolsky, M. A. (2005). Anaerobic and aerobic enzyme activities in human skeletal muscle from children and adults. *Pediatric research*, *57*(3), 331-335.

Khalfa, S., Dubal, S., Veuillet, E., Perez-Diaz, F., Jouvent, R., & Collet, L. (2002). Psychometric normalization of a hyperacusis questionnaire. *ORL*, *64*(6), 436-442.

Kong, K., & Coates, H. L. (2009). Natural history, definitions, risk factors and burden of otitis media. *Medical Journal of Australia*, *191*(9), S39.

Kontiokari, T., Niemelä, M., & Uhari, M. (1998). Middle ear effusion among children diagnosed and treated actively for acute otitis media. *European Journal of Pediatrics*, *157*(9), 731–734. doi:10.1007/s004310050924

Kral, A., & Sharma, A. (2012). Developmental neuroplasticity after cochlear implantation. *Trends in Neurosciences*, *35*(2), 111–122. doi:10.1016/j.tins.2011.09.004

Kumar, V., Abbas, A. K., & Aster. J. C. (2014). Chapter III: Inflammation and Repair. In A. K. V. Kumar, *Robins and Cotran Pathologic bases of the diseases.* Philadelphia: Elselvier Sounders.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4*(11), 1–12. doi:10.3389/fpsyg.2013.00863

Landrot, I. D. R., Roche, F., Pichot, V., Teyssier, G., Gaspoz, J. M., Barthelemy, J. C., & Patural, H. (2007). Autonomic nervous system activity in premature and full-term infants from theoretical term to 7 years. *Autonomic Neuroscience, 136*(1), 105-109.

Leach, A. J., & Morris, P. S. (2001). Perspectives on infective ear disease in indigenous Australian children. *Journal of paediatrics and child health, 37*(6), 529-530.

Li, X., Driscoll, C., & Culbert, N. (2009). Investigating the Performance of a Questionnaire for Hearing Screening of Schoolchildren in China. *Australian and New Zealand Journal of Audiology, 31*(1), 45–52. doi:10.1375/audi.31.1.45

Light, R. W., Macgregor, M. I., Luchsinger, P. C., & Ball, W. C. (1972). Pleural effusions: the diagnostic separation of transudates and exudates. *Annals of Internal Medicine, 77*(4), 507-513.

Lo, P. S. Y., Tong, M. C. F., Wong, E. M. C., & Van Hasselt, C. A. (2006). Parental suspicion of hearing loss in children with otitis media with effusion. *European Journal of Pediatrics, 165*(12), 851–857. doi:10.1007/s00431-006-0181-5

Lous, J. (1987). Screening for secretory otitis media: evaluation of some impedance screening programs for long-lasting secretory otitis media in 7-year-old children. *International Journal of Pediatric Otorhinolaryngology, 13*(1), 85–97. doi:10.1016/0165-5876(87)90011-5

MacArthur, C. J., Wilmot, B., Wang, L., Schuller, M., Lighthall, J., & Trune, D. (2014). Genetic susceptibility to chronic otitis media with effusion: Candidate gene single nucleotide polymorphisms. *The Laryngoscope, 124*(5), 1229-1235.

MacIntyre, E.A., & Heinrich. J. (2012). Otitis media in infancy and development of asthma and atopic disease. *Current Allergy and Asthma Reports, 12*(6), 547-50.

Marchisio, P., Principi, N., Passali, D., Salpietro, D. C., Boschi, G., Chetri, G., Caramia, G., Longhi, R., Reali, E., Meloni, G., De Santis, A., Sacher, B., & Cupido, G. (1998). Epidemiology and treatment of otitis media with effusion in children in the first year of primary school. *Actaotolaryngologica, 118*(4), 557-562.

Maw, A. R., & Tiwari, R. S. (1988). Children with glue ear: how do they present? *Clinical Otolaryngology and Allied Sciences, 13*(3), 171–177. doi:10.1111/j.1365-2273.1988.tb01113.x

McFadden, D. M., Berwick, D. M., Feldstein, M. L., & Marter, S. S. (1985). Age-specific patterns of diagnosis of acute otitis media. *Clinical pediatrics, 24*(10), 571-575.

Miles, J., & Shevlin, M. (2001). *Applying regression and correlation: A guide for students and researchers.* London: SAGE.

Milovanovic, J., Filipovic, S. A., Haggard, M. P., Zhang, M. F. H. Spencer, H., & Eurotitis 2 study group. (accepted). Precision-scored parental report questions and HL-scaled tympanometry as informative measures of hearing in otitis media 1: large-sample evidence on determinants and complementarity to pure-tone audiometry. *International Journal of Pediatric Otorhinolaryngology*.

Mittal, R., Kodiyan, J., Gerring, R., Mathee, K., Li, J. D., Grati, M. H., & Liu, X. Z. (2014). Role of innate immunity in the pathogenesis of otitis media. *International Journal of Infectious Diseases, 29,* 259-267.

Morrison, A. S. Screening. (1998). In: K. J. Rothman & S. Greenland (Eds.). *Modern Epidemiology* (2nd ed, pp. 499-518). Philadelphia, Pa: Lippincott-Raven Publishers.

Morris, P. S., Leach, A. J., Halpin, S., Mellon, G., Gadil, G., Wigger, C., Mackenzie, G. Wilson, C., Gadil, E., & Torzillo, P. (2007). An overview of acute otitis media in Australian Aboriginal children living in remote communities. *Vaccine, 25*(13), 2389-2393.

MRC Multi-centre Otitis Media Study Group (2004). Speech reception in noise: An indicator of benefit from otitis media with effusion surgery. *Clinical Otolaryngology and Allied Sciences, 29*(5), 497–504. doi:10.1111/j.1365-2273.2004.00843.x

MRC Multi-centre Otitis Media Study Group (2009). Air-conduction estimated from tympanometry (ACET): 1. Relationship to measured hearing in OME. *International Journal of Pediatric Otorhinolaryngology*, *73*(1), 21–42. doi:10.1016/j.ijporl.2008.09.014

MRC Multicentre Otitis Media Study Group. (2012). Adjuvant adenoidectomy in persistent bilateral otitis media with effusion: Hearing and revision surgery outcomes through 2 years in the TARGET randomised trial. *Clinical Otolaryngology, 37*(2), 107–116. doi:10.1111/j.1749-4486.2012.02469.x

Myers, A. K., Ray, J., & Kulesza, R. J. (2012). Neonatal conductive hearing loss disrupts the development of the Cat-315 epitope on perineuronal nets in the rat superior olivary complex. *Brain Research, 1465,* 34–47. doi:10.1016/j.brainres.2012.05.024

Nelson, J. D. (1988). Chronic suppurative otitis media. *The Pediatric infectious disease journal, 7*(6), 446-448.

Nondahl, D. M., Cruickshanks, K. J., Wiley, T. L., Tweed, T. S., Klein, R., & Klein, B. E. (1998). Accuracy of self-reported hearing loss. *Audiology: Official Organ of the International Society of Audiology*, *37*(5), 295–301. doi:10.3109/00206099809072983

Niemelä, M., Uhari, M., & Möttönen, M. (1995). A pacifier increases the risk of recurrent acute otitis media in children in day care centers. *Pediatrics, 96*(5), 884-888.

O'Brien, M. A., Prosser, L. A., Paradise, J. L., Ray, G. T., Kulldorff, M., Kurs-Lasky, M., Hinrichsen, V. L., Mehta, J., Cplborn, D. K., & Lieu, T. A. (2009). New vaccines against otitis media: projected benefits and cost-effectiveness. *Pediatrics, 123*(6), 1452-1463.

Olusanya, B. (2001). Early Detection of Hearing impairment in a developing country: what options? *Audiology, 40*, 141–147.

Paparella, M. M., Schachern, P. A., Yoon, T. H., Abdelhammid, M. M., Sahni, R. A. N. A., & DaCosta, S. S. (1990). Otopathologic correlates of the continuum of otitis media. *The Annals of otology, rhinology & laryngology Supplement, 148*, 17-22.

Paradise, J. L., Rockette, H. E., Colborn, D. K., Bernard, B. S., Smith, C. G., Kurs-Lasky, M., & Janosky, J. E. (1997). Otitis media in 2253 Pittsburgh-Area infants: prevalence and risk factors during the first two years of life. *Pediatrics, 99*(3), 318–333. doi:10.1542/peds.99.3.318

Paradise, J. L., Dollaghan, C. A, Campbell, T. F., Feldman, H. M., Bernard, B. S., Colborn, D. K., Rockette, H. E., Janosky, J. E., Pitcairn, D. L., Sabo, D. L., Kurs-Lasky, M., & Smith, C. G. (2000). Language, speech sound production, and cognition in three-year-old children in relation to otitis media in their first three years of life. *Pediatrics, 105*(5), 1119–1130. doi:10.1542/peds.105.5.1119

Paradise, J. L., Dollaghan, C. A, Campbell, T. F., Feldman, H. M., Bernard, B. S., Colborn, D. K., Rockette, H. E., Janosky, J. E., Pitcairn, D. L., Kurs-Lasky, M., Sabo, D. L., & Smith, C. G. (2003). Otitis media and tympanostomy tube insertion during the first three years of life: developmental outcomes at the age of four years. *Pediatrics, 112*(2), 265–277. doi:10.1542/peds.112.2.265

Paramothayan, N. S., & Barron, J. (2002). New criteria for the differentiation between transudates and exudates. *Journal of clinical pathology, 55*(1), 69-71.

Pichichero, M. E., & Casey, J. R. (2007). Emergence of a multiresistant serotype 19A pneumococcal strain not included in the 7-valent conjugate vaccine as an otopathogen in children. *Jama, 298*(15), 1772-1778.

Preciado, D., Goyal, S., Rahimi, M., Watson, A. M., Brown, K. J., Hathout, Y., & Rose, M. C. (2010). MUC5B is the predominant mucin glycoprotein in chronic otitis media fluid. *Pediatric research, 68*(3), 231-236.

Revai, K., Dobbs, L. A., Nair, S., Patel, J. A., Grady, J. J., & Chonmaitree, T. (2007). Incidence of acute otitis media and sinusitis complicating upper respiratory tract infection: the effect of age. *Pediatrics, 119*(6), e1408-e1412.

Roberts, J. E., Rosenfeld, R. F., & Zeisel, S. A. (2004). Otitis media and speech and language: A meta-analysis of prospective studies. *The Journal of Pediatrics, 113*(3), 238–248. doi:10.1016/j.jpeds.2004.06.039

Robinette, M. S. & Glattke. T. J. (2007). *Otoacoustic emission* (3rd Ed.). New York: Thieme Medical Publisher.

Robertson, T., Batty, G. D., Der, G., Fenton, C., Shiels, P. G., & Benzeval, M. (2013). Is Socioeconomic Status Associated With Biological Aging as Measured by Telomere Length? *Epidemiologic Reviews*, *35*(1), 98–111. doi:10.1093/epirev/mxs001

Rood, S. R., & Doyle, W. J. (1978). Morphology of tensor veli palatini, tensor tympani, and dilatator tubae muscles. *Annals of Otology, Rhinology & Laryngology, 87*(2), 202-210.

Rosen, S., & Howell, P. (1991). *Signals and systems for speech and hearing*. London: Elsevier.

Rosenfeld, R. M., Goldsmith, A. J., & Madell, J. R. (1998). How accurate is parent rating of hearing for children with otitis media? *Archives of Otolaryngology--Head & Neck Surgery, 124*(9), 989–992.

Rutherford, A. (2012). *ANOVA and ANCOVA: A GLM Approach.* New York: John Wiley & Sons.

Rosenfeld, R. M., Schwartz, S. R., Pynnonen, M. A., Tunkel, D. E., Hussey, H. M., Fichera, J. S., Grimes, A. M., Hackell, H., Haynes, D. S., Kim, T. W., Lafreniere, D. C., LeBlanc, K., Mackey, W. L., Netterville, J. L., Pipan, M. E., Raol, N.P., & Schellhase, K. G. (2013). Clinical practice guideline: tympanostomy tubes in children -- executive summary. *Otolaryngology -- Head and Neck Surgery*, *149*(1), 8–16. doi:10.1177/0194599813490141

Rothman, R., Owens, T., & Simel, D. L. (2003). Does This Child Have Acute Otitis Media? *JAMA: The Journal of the American Medical Association*, *290*(12), 1633–1640. doi:10.1001/jama.290.12.1633

Rovers, M. M., Straatman, H., Ingels, K., van der Wilt, G. J., van den Broek, P., & Zielhuis, G. A. (2000). The effect of ventilation tubes on language development in infants with otitis media with effusion: A randomized trial. *Pediatrics*, *106*(3), E42. doi:10.1542/peds.106.3.e42

Rovers, M., Haggard, M., Gannon, M., Koeppen-Schomerus, G., & Plomin, R. (2002). Heritability of symptom domains in otitis media: a longitudinal study of 1,373 twin pairs. *American Journal of Epidemiology, 155*(10), 958-964.

Rovers, M. M., Schilder, A. G., Zielhuis, G. A, & Rosenfeld, R. M. (2004). Otitis media. *The Lancet*, *363*(9407), 465–473. doi:10.1016/S0140-6736(04)15495-0

Ryborg, C. T., Sondergaard, J., Lous, J., Munck, A., Larsen, P. V., & Thomsen, J. L. (2014). Quality of life in children with otitis media--a cohort study. *Family Practice*, *31*(1), 30–37. doi:10.1093/fampra/cmt066

Sangket, U., Vijasika, S., Noh, H., Chantratita, W., Klungthong, C., Yoon, I. K., Fernandez, S., & Rutvisuttinunt, W. (2015). SNPer: An R library for quantitative variant analysis on single nucleotide polymorphisms among influenza virus populations. *Plos One, 10*(4), e0122812. doi:10.1371/journal.pone.0122812

Sanes, D. H., & Kotak, V. C. (2011). Developmental plasticity of auditory cortical inhibitory synapses. *Hearing Research, 279*(1-2), 140–148. doi:10.1016/j.heares.2011.03.015

Sente, M., & Sente. R. (2000). Evaluation of the level of hearing loss based on the results of tympanometry in children under 5 years of age with eustachian tube dysfunction.

Senturia, B. H. (1963). Secretory otitis media. *Proceedings of the Royal Society of Medicine, 56*(8), 687.

Shaikh, N., Hoberman, A., Paradise, J. L., Rockette, H. E., Kurs-Lasky, M., Colborn, D. K., Kearney, D. H., RN, CCRC, & Zoffel, L. M. (2009). Responsiveness and construct validity of a symptom scale for acute otitis media. *The Pediatric Infectious Disease Journal, 28*(1), 9–12. doi:10.1097/INF.0b013e318185a3a0

Shaman, J., Pitzer, V. E., Viboud, C., Grenfell, B. T., & Lipsitch, M. (2010). Absolute humidity and the seasonal onset of influenza in the continental United States. *PLoS Biology, 8*(2). doi:10.1371/journal.pbio.1000316

Shanks, J., & Shohet, J. (2009). *Tympanometry in clinical practice* (pp. 157-188). Handbook of clinical audiology.

Sharma, S. K., & Pichichero, M. E. (2013). Cellular immune response in young children accounts for recurrent acute otitis media. *Current allergy and asthma reports, 13*(5), 495-500.

Shriberg, L. D., Friel-Patti, S., Flipsen, P. & Brown, R. L. (2000). Otitis media, fluctuant hearing loss, and speech-language outcomes: a preliminary structural equation model. *Journal of Speech, Language, and Hearing Research, 43,* 100–120.

Sichel, J. Y., Priner, Y., Barshtein, G., Weiss, S., Eliashar, R., Levi, H., & Elidan, J. (2003). Characteristics of the type B tympanogram can predict the magnitude of the air-bone gap in otitis media with effusion. *Annals of Otology, Rhinology & Laryngology, 112*(5), 450-454.

Smaldino, J., & Flexer, C. (2008). Classroom acoustics: Personal and soundfield FM and IR systems. In J. Madell & C. Flexer (Eds.), *Pediatric audiology: Diagnosis, technology, and management* (pp. 192-202). New York, NY: Thieme

Smith, C. G., Paradise, J. L., Sabo, D. L., Rockette, H. E., Kurs-Lasky, M., Bernard, B. S., & Colborn, D. K. (2006). Tympanometric findings and the probability of middle-ear effusion in 3686 infants and young children. *Pediatrics,* 118(1), 1-13.

Steptoe, A., Hamer, M., Butcher, L., Lin, J., Brydon, L., Kivimäki, M., Marmot, M., Blackburn, E., & Erusalimsky, J. D. (2011). Educational attainment but not measures of current socioeconomic circumstances are associated with leukocyte telomere length in healthy older men and women. *Brain, Behavior, and Immunity, 25*(7), 1292–1298. doi:10.1016/j.bbi.2011.04.010

Stevenson, A., & Waite. M. (2011). *Concise Oxford English dictionary* (pp. 1682), Oxford University Press, Oxford, UK.

Stewart, M. G., Chlms, L. A., Friedman, E. M., Sulek, M., N.O., D., Fernandez, D., & Bautista, M. H. (1999). Is parental perception an accurate predictor of childhood hearing loss? A prospective study. *Otolaryngology and Head and Neck, 120*(3), 340–344.

Tasker, A., Dettmar, P. W., Panetti, M., Koufman, J. A., Birchall, J. P., & Pearson, J. P. (2002). Reflux of gastric juice and glue ear in children. *The Lancet, 359*(9305), 493.

Teele, D. W., Klein, J. O., & Rosner, B. A. (1984). Otitis media with effusion during the first three years of life and development of speech and language. *Pediatrics, 74*(2), 282-287.

Teele, D. W., Klein, J. O., & Rosner, B. (1989). Epidemiology of otitis media during the first seven years of life in children in greater Boston: a prospective, cohort study. *Journal of infectious diseases, 160*(1), 83-94.

Terwee, C. B., Bot, S. D. M., de Boer, M. R., van der Windt, D. A. W. M., Knol, D. L., Dekker, J., Bouter, L.M., & de Vet, H. C. W. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*, *60*(1), 34–42. doi:10.1016/j.jclinepi.2006.03.012

Timmerman, A. A., Meesters, C. M. G., Speyer, R., & Anteunis, L. J. C. (2007). Psychometric qualities of questionnaires for the assessment of otitis media impact. *Clinical Otolaryngology*, *32*(6), 429–439. doi:10.1111/j.1749-4486.2007.01570.x

Timmerman, A. A., Meesters, C. M. G., Anteunis, L. J. C., Chenault, M. N., & Haggard, M. P. (2008). Psychometric evaluation of the OM8-30 questionnaire in Dutch children with otitis media. *European Archives of Oto-Rhino-Laryngology*, *265*(9), 1047–1056. doi:10.1007/s00405-008-0591-2

Ting, P. J., Lin, C. H., Huang, F. L., Lin, M. C., Hwang, K. P., Huang, Y. C., Chium C. H., Lin, T.Y., & Chen, P. Y. (2012). Epidemiology of acute otitis media among young children: a multiple database study in Taiwan. *Journal of Microbiology, Immunology and Infection, 45*(6), 453-458.

Tharpe, A. M., & Bess, F. H. (1991). Identification and management of children with minimal hearing loss. *International Journal of Audiology*, *21*(1), 41–50. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2037417

Thornton, R. B., Wiertsema, S. P., Kirkham, L. A., Rigby, P. J., Vijayasekaran, S., Coates, H. L., & Richmond, P. C. (2013). Neutrophil extracellular traps and bacterial biofilms in middle ear effusion of children with recurrent acute otitis media–a potential treatment target. *PLoS One, 8*(2), e53837.

Tomoda, K., Morii, S., Yamashita, T., & Kumazawa, T. (1984). Histology of human eustachian tube muscles: effect of aging. *Annals of Otology, Rhinology & Laryngology, 93*(1), 17-24.

Tsilis, N. S., Vlastarakos, P. V, Chalkiadakis, V. F., Kotzampasakis, D. S., & Nikolopoulos, T. P. (2013). Chronic otitis media in children: an evidence-based guide for diagnosis and management. *Clinical Pediatrics, 52*(9), 795–802. doi:10.1177/0009922813482041

Vergison, A., Dagan, R., Arguedas, A., Bonhoeffer, J., Cohen, R., DHooge, I., Hoberman, A., Liese, J., Marchisio, P., Palmu, A. A., Ray, G. T. Sanders, E. A. M., Simões, E. A. F., Uhari, M., van Eldere, J., & Pelton, S. I. (2010). Otitis media and its consequences: beyond the earache. *The Lancet Infectious Diseases, 10*(3), 195–203. doi:10.1016/S1473-3099(10)70012-8

Vernon-Feagans, L., Manlove, E. E., & Volling, B. L. (1996). Otitis media and the social behavior of day-care-attending children. Child Development, 67(4), 1528–1539.

Watkin, P. M., Baldwin, M., & Laoide, S. (1990). Parental suspicion and identification of hearing impairment. *Archives of Disease in Childhood*, *65*(8), 846–850. doi:10.1136/adc.65.8.846

Webster, D. B. (1977). Neonatal sound deprivation affects brain stem auditory nuclei. *Archives of Otolaryngology—Head & Neck Surgery*, *103*(7), 392.

Whelan, M. A., Hwan, W. H., Beausoleil, J., Hauck, W. W., & McGeady, S. J. (2006). Infants presenting with recurrent infections and low immunoglobulins: characteristics and analysis of normalization. *Journal of Clinical Immunology*, *26*(1), 7–11. doi:10.1007/s10875-006-8144-1

Whitton, J. P., & Polley, D. B. (2011). Evaluating the perceptual and pathophysiological consequences of auditory deprivation in early postnatal life: A comparison of basic and clinical studies. *Journal of the Association for Research in Otolaryngology*, *12*(5), 535–546. doi:10.1007/s10162-011-0271-6

Williams, C. J., & Jacobs, A. M. (2009). The impact of otitis media on cognitive and educational outcomes. *Otitis Media*, *191*(9), S69–S72.

Williamson, I., Benge, S., Barton, S., Petrou, S., Letley, L., Fasey, N., Haggard, M., Little, P. (2009). Topical intranasal corticosteroids in 4-11 year old children with persistent bilateral otitis media with effusion in primary care: double blind randomised placebo controlled trial. *BMJ (Clinical Research Ed.), 339*, b4984. doi:10.1136/bmj.b4984

Williamson, I., Vennik, J., Harnden, A., Voysey, M., Perera, R., Kelly, S., Yao, G., Raftery, J., David Mant, D., & Little, P. (2015). Effect of nasal balloon autoinflation in children with otitis media with effusion in primary care: an open randomized controlled trial. *Canadian Medical Association Journal*, *104*(7), 1–9. doi: 10.1503/cmaj.141608

Yiengprugsawan, V., Hogan, A., & Strazdins, L. (2013). Longitudinal analysis of ear infection and hearing impairment: findings from 6-year prospective cohorts of Australian children. *BMC pediatrics*, *13*(1), 28.

Zhang, Y., Xu, M., Zhang, J., Zeng, L., Wang, Y., & Zheng, Q. Y. (2014). Risk factors for chronic and recurrent otitis media–a meta-analysis. *PloS one*, *9*(1), e86397.

Zielhuis, G., Rach, G., & van Den Broek, P. (1989). Screening for otitis media with effusion in preschool children. *The Lancet*, *333*(8633), 311–314. doi:10.1016/S0140-6736(89)91317-2

Zielhuis, G. A, Rach, G. H., & van den Broek, P. (1990). The occurrence of otitis media with effusion in Dutch pre-school children. *Clinical Otolaryngology and Allied Sciences*, *15*(2), 147–153.

BIOGRAFIJA

Dr mr sci SNEŽANA ANDRIĆ FILIPOVIĆ rođena je u Sokobanji 1966 godine. Osnovno i srednje obrazovanje (molekulrana biologija sa bihemijom) završila je u svom rodnom gradu sa odličnim uspehom.

Medicinski fakultet Univerziteta u Beogradu upisala je 1984 godine i stekla zvanje Doktora medicine maja 1990 godine u svojoj 23 godini sa prosečnom ocenom 9.03. Godine 1993. upisuje specijalističke studije iz otorinolaringologije i zvanje specijaliste otorinolarngologije stiče decembra 1997. godine.Postdiplomske studije iz oblasti otorinolaringologije upisuje 1998, a zvanje Magistra nauka stiče 2003 godine.Zvanje specijaliste uže specijalizacije iz oblasti audiologijestekla je 2013 godine.

Od 1993 godine radila je u Kliničko-bolničkom centru Priština, u Klinici za otorinolarngologiju, a od 2002 godine stalno je zaposlena u Kliničkom Centru Srbije u Klinici za otorinolaringologiju i maksilofacijalnu hirurgiju, u Odseku audiološke rehabilitacije.

Više od dve decenije bavi se ranom dijagnostikom, amplifikacijom i rehabilitacijom dece oštećenog sluha. Fokus iszraživačkog rada, pored rane dijagnostike dece sa permanentnim oštećenjem sluha, su deca predškolskog uzrasta sa zapaljenskim promenama srednjeg uva. Postaje aktivni saradnik Eurotitis 2 projekta 1997godine čiji je osnivač Prof. Haggard sa Univerziteta Cambridge, UK. Aktivno učestvuje u projektu do 2014. godine sa ukupno 900 ispitanika, što predstavlja trećinu ukupne baze podataka.

Od 2014. godine je član projekta Horizon 2020, studije za evropsku standardizaciju iz oblasti skrininga vida i sluha.

Svoje rezultate u radu sa decom oštećenog sluha, prezentovala je na domaćim i evropskim kongresima i simpozijumima. Iskustvo i znanje iz oblasti audiologije koristila je za unapređenje organizacije rada Odseka za audiološku rehabilitaciju.

Prilog 1.

Izjava o autorstvu

Potpisani-a Snežana Andrić Filipović

broj upisa _____

Izjavljujem

da je doktorska disertacija pod naslovom

CORRELATION ANALYSES OF THE SEASONAL CHANGES AND QUALITY OF LIFE WITH THE CIINICAL PRESENTATIONS OF OTITIS MEDIA IN CHILDREN 3-8 YEARS OF AGE

АНАЛИЗА ПОВЕЗАНОСТИ СЕЗОНСКИХ ПРОМЕНА И КВАЛИТЕТА ЖИВОТА СА РАЗЛИЧИТИМ КЛИНИЧКИМ ПРЕЗЕНТАЦИЈАМА ЗАПАЉЕЊА СРЕДЊЕГ УВА КОД ДЕЦЕ УЗРАСТА 3-8 ГОДИНА

_____

_____

- rezultat sopstvenog istraživačkog rada,
- da predložena disertacija u celini ni u delovima nije bila predložena za dobijanje bilo koje diplome prema studijskim programima drugih visokoškolskih ustanova,
- da su rezultati korektno navedeni i
- da nisam kršio/la autorska prava i koristio intelektualnu svojinu drugih lica.

Potpis doktoranda

U Beogradu, 27.12.2015

_____

Prilog 2.

Izjava o istovetnosti štampane i elektronske verzije doktorskog rada

Ime i prezime autora Snežana Andrić Filipović

Broj upisa _____

Studijski program _____

Naslov rada :

CORRELATION ANALYSES OF THE SEASONAL CHANGES AND QUALITY OF LIFE WITH THE CIINICAL PRESENTATIONS OF OTITIS MEDIA IN CHILDREN 3-8 YEARS OF AGE

АНАЛИЗА ПОВЕЗАНОСТИ СЕЗОНСКИХ ПРОМЕНА И КВАЛИТЕТА ЖИВОТА СА РАЗЛИЧИТИМ КЛИНИЧКИМ ПРЕЗЕНТАЦИЈАМА ЗАПАЉЕЊА СРЕДЊЕГ УВА КОД ДЕЦЕ УЗРАСТА 3-8 ГОДИНА

Mentor: Prof. Dr Jovica Milovanović

Komentor: Prof. M. Haggard

Potpisani: Snežana Andrić Filipović

izjavljujem da je štampana verzija mog doktorskog rada istovetna elektronskoj verziji koju sam predao/la za objavljivanje na portalu **Digitalnog repozitorijuma Univerziteta u Beogradu.**

Dozvoljavam da se objave moji lični podaci vezani za dobijanje akademskog zvanja doktora nauka, kao što su ime i prezime, godina i mesto rođenja i datum odbrane rada.

Ovi lični podaci mogu se objaviti na mrežnim stranicama digitalne biblioteke, u elektronskom katalogu i u publikacijama Univerziteta u Beogradu.

Potpis doktoranda

U Beogradu, 27.12.15

_____

Prilog 3.

Izjava o korišćenju

Ovlašćujem Univerzitetsku biblioteku „Svetozar Marković“ da u Digitalni repozitorijum Univerziteta u Beogradu unese moju doktorsku disertaciju pod naslovom:

CORRELATION ANALYSES OF THE SEASONAL CHANGES AND QUALITY OF LIFE WITH THE CIINICAL PRESENTATIONS OF OTITIS MEDIA IN CHILDREN 3-8 YEARS OF AGE

АНАЛИЗА ПОВЕЗАНОСТИ СЕЗОНСКИХ ПРОМЕНА И КВАЛИТЕТА ЖИВОТА СА РАЗЛИЧИТИМ КЛИНИЧКИМ ПРЕЗЕНТАЦИЈАМА ЗАПАЉЕЊА СРЕДЊЕГ УВА КОД ДЕЦЕ УЗРАСТА 3-8 ГОДИНА

koja je moje autorsko delo.

Disertaciju sa svim prilozima predao/la sam u elektronskom formatu pogodnom za trajno arhiviranje.

Moju doktorsku disertaciju pohranjenu u Digitalni repozitorijum Univerziteta u Beogradu mogu da koriste svi koji poštuju odredbe sadržane u odabranom tipu licence Kreativne zajednice (Creative Commons) za koju sam se odlučio/la.

1. Autorstvo

2. Autorstvo - nekomercijalno

3. Autorstvo – nekomercijalno – bez prerade

4. Autorstvo – nekomercijalno – deliti pod istim uslovima

5. Autorstvo – bez prerade

6. Autorstvo – deliti pod istim uslovima

(Molimo da zaokružite samo jednu od šest ponuđenih licenci, kratak opis licenci dat je na poleđini lista).

Potpis doktoranda

U Beogradu, 27.12.15

1. **Autorstvo** - Dozvoljavate umnožavanje, distribuciju i javno saopštavanje dela, i prerade, ako se navede ime autora na način određen od strane autora ili davaoca licence, čak i u komercijalne svrhe. Ovo je najslobodnija od svih licenci.

2. **Autorstvo – nekomercijalno.** Dozvoljavate umnožavanje, distribuciju i javno saopštavanje dela, i prerade, ako se navede ime autora na način određen od strane autora ili davaoca licence. Ova licenca ne dozvoljava komercijalnu upotrebu dela.

3. **Autorstvo - nekomercijalno** – bez prerade. Dozvoljavate umnožavanje, distribuciju i javno saopštavanje dela, bez promena, preoblikovanja ili upotrebe dela u svom delu, ako se navede ime autora na način određen od strane autora ili davaoca licence. Ova licenca ne dozvoljava komercijalnu upotrebu dela. U odnosu na sve ostale licence, ovom licencom se ograničava najveći obim prava korišćenja dela.

4. **Autorstvo - nekomercijalno** – deliti pod istim uslovima. Dozvoljavate umnožavanje, distribuciju i javno saopštavanje dela, i prerade, ako se navede ime autora na način određen od strane autora ili davaoca licence i ako se prerada distribuira pod istom ili sličnom licencom. Ova licenca ne dozvoljava komercijalnu upotrebu dela i prerada.

5. **Autorstvo – bez prerade.** Dozvoljavate umnožavanje, distribuciju i javno saopštavanje dela, bez promena, preoblikovanja ili upotrebe dela u svom delu, ako se navede ime autora na način određen od strane autora ili davaoca licence. Ova licenca dozvoljava komercijalnu upotrebu dela.

6. **Autorstvo** - deliti pod istim uslovima. Dozvoljavate umnožavanje, distribuciju i javno saopštavanje dela, i prerade, ako se navede ime autora na način određen od strane autora ili davaoca licence i ako se prerada distribuira pod istom ili sličnom licencom. Ova licenca dozvoljava komercijalnu upotrebu dela i prerada. Slična je softverskim licencama, odnosno licencama otvorenog koda.