Observational Research in Childhood Infectious Diseases (ORChID): a dynamic birth cohort study

Stephen Bernard Lambert, Robert S Ware, Anne L Cook, Frances A Maguire, David M Whiley, Seweryn Bialasiewicz, Ian M Mackay, David Wang, Theo P Sloots, Michael D Nissen, Keith Grimwood

ABSTRACT

Introduction: Even in developed economies infectious diseases remain the most common cause of illness in early childhood. Our current understanding of the epidemiology of these infections is limited by reliance on data from decades ago performed using low-sensitivity laboratory methods, and recent studies reporting severe, hospital-managed disease.

Methods and analysis: The Observational Research in Childhood Infectious Diseases (ORChID) study is an ongoing study enrolling a dynamic birth cohort to document the community-based epidemiology of viral respiratory and gastrointestinal infections in early childhood. Women are recruited antenatally, and their healthy newborn is followed for the first 2 years of life. Parents keep a daily symptom diary for the study child, collect a weekly anterior nose swab and dirty nappy swab and complete a burden diary when a child meets pre-defined illness criteria. Specimens will be tested for a wide range of viruses by real-time PCR assays. Primary analyses involves calculating incidence rates for acute respiratory illness (ARI) and acute gastroenteritis (AGE) for the cohort by age and seasonality. Control material from children when they are without symptoms will allow us to determine what proportion of ARIs and AGE can be attributed to specific pathogens. Secondary analyses will assess the incidence and shedding duration of specific respiratory and gastrointestinal pathogens.

Ethics and dissemination: This study is approved by The Human Research Ethics Committees of the Children’s Health Queensland Hospital and Health Service, the Royal Brisbane and Women’s Hospital and The University of Queensland.

Trial registration: clinicaltrials.gov NCT01304914.

ARTICLE SUMMARY

Article focus

- Infectious diseases are a common cause of morbidity in early childhood, even in developed economies.
- A diagnostic gap exists for common respiratory and gastrointestinal syndromes, with the likelihood that as yet undiscovered pathogens are involved.
- Existing knowledge about these common illnesses relies on research conducted before the rapid developments in molecular diagnostics of recent decades or focuses on disease at the severe end of the spectrum—hospitalisations— affecting a limited number of children and discounting the burden of more common, but less severe, community-managed illness.

Key messages

- This protocol outlines a dynamic birth cohort study that will allow for a detailed description of the epidemiology of respiratory and gastrointestinal viruses during the first 2 years of life.
- The large biobank of specimens to be collated will act as a rich source of material to answer targeted research questions, including the role of virus acquisition and shedding on clinical illness and the discovery of new infectious agents.

Strengths and limitations of this study

- As study procedures, including specimen collection and return, are conducted by parents, findings will be free from Hawthorne effects due to frequent interactions with study staff.
- Systematic weekly sampling will provide a control set of specimens for the individual and the cohort, allowing quantification of virus-specific attributable risk to illness.
- Non-random recruitment and enrolment requires awareness and assessment of potential bias and confounding prior to broad-based generalisation.
- Similar studies in the past have oversampled from higher socioeconomic households, and we attempt to avoid this by using a recruitment strategy that targets pregnant women in both public and private hospital settings.

INTRODUCTION

Even in developed economies where populations have high-quality housing, sanitation, secure food and drinking water supplies, good personal hygiene standards, widespread vaccine use and access to high-quality medical care, infectious diseases remain the most...
common cause of significant morbidity, and occasionally mortality, in early childhood.\textsuperscript{1,4–6} Our current understanding of the epidemiology of early childhood infections is limited by reliance on community-based data from decades ago using low-sensitivity diagnostic methods,\textsuperscript{7,9} and recent studies that primarily focus on severe, hospital-managed disease.\textsuperscript{10,11} Much of what we know, especially with newly discovered agents, originates from hospital-based prevalence studies where more than 80\% of cases are less than 2 years of age, representing the sickest 2–3\% of young children seen. Experience with influenza illustrates how easily disease burden can be underestimated by extrapolating from hospital data.\textsuperscript{12–14} Available community-based studies also have important methodological limitations, such as sampling from highly selected subject populations, lack of adequate control subjects, restricted sampling frequency and observation periods, small subject numbers and/or reporting on only a few or single agents.\textsuperscript{15–18} A key methodological issue is the use of home visits by healthcare workers or the requirement for clinic visits for specimen collection. Both are likely to be an imposition on busy families, regardless of the setting, leading to biased estimates of infection events and specimen availability.\textsuperscript{19}

The highest incidence rates of acute respiratory infections (ARI) are during the first 2 years of life where on average infants experience six to eight ARIs each year.\textsuperscript{20} Complication rates from acute otitis media (30\%) and sinusitis (8\%) are also highest in this age group,\textsuperscript{21} while 3–5\% of all infants are hospitalised for viral lower respiratory tract infections, including bronchiolitis, pneumonia, croup and secondary bacterial pneumonia.\textsuperscript{22} There is emerging evidence that infectious insults to the growing and developing lung during early childhood contribute to the pathogenesis of chronic pulmonary disorders in older children and adulthood, such as asthma,\textsuperscript{23} chronic obstructive pulmonary disease,\textsuperscript{24,25} and bronchiectasis.\textsuperscript{26} Young children are often household introducers, actively transmitting respiratory infections to other family members.\textsuperscript{27} Taken together, ARIs in children result in enormous current and future costs to the healthcare system, families and society.\textsuperscript{14}

Likewise, acute gastroenteritis (AGE) is still a major cause of childhood morbidity in high-income countries. An average of 3.2 diarrhoeal episodes per child per year has been reported globally, reaching as much as 12 episodes per child per year in some settings.\textsuperscript{1} Viruses remain important causes of gastroenteritis in children from developed countries, even after rotavirus vaccines have been introduced. In addition to rotaviruses, noroviruses, enteric adenoviruses and astroviruses are also important enteric pathogens.\textsuperscript{28} Nevertheless, a large diagnostic gap exists in developed countries where community-based studies of at least 1 year duration report no identifiable pathogen in approximately 40–60\% of young children with AGE.\textsuperscript{29–31} In contrast, recent work employing sensitive molecular techniques describes two healthy infants shedding a diverse range of enteric viruses almost continuously throughout the first 12–14 months of life.\textsuperscript{32} In order to first understand and then prevent early-childhood AGE such contradictory findings need to be resolved.

As well as pathogens traditionally associated with childhood infections, there are several recently identified respiratory and gastrointestinal agents, mostly viruses, which may have an important role in the epidemiology of these illnesses. Furthermore, there are likely to be undiscovered pathogens that contribute to the remaining diagnostic gap for ARI\textsuperscript{33} and AGE episodes.\textsuperscript{34}

We have designed this study—Observational Research in Childhood Infectious Diseases (ORChID)—to overcome identified methodological issues and collect data on the population-based epidemiology of respiratory and gastrointestinal infections during the first 2 years of life.

**METHODS AND ANALYSIS**

**Study design**

ORChID is a 5-year prospective, community-based, longitudinal, dynamic birth cohort study of ARI and AGE episodes and respiratory and gastrointestinal pathogen detection in children during the first 2 years of life (clinicaltrials.gov: NCT01304914).\textsuperscript{18} Recruitment will take place over a 2-year period, the last recruited child will be followed until their second birthday, and the final year of the study will be used to complete data entry and analysis, laboratory testing of specimens and to report study findings. The progressive 2-year recruitment plan allows for seasonal and year-to-year variation in pathogen activity.\textsuperscript{13,20}

The study is designed to allow families to be self-sufficient and avoid unnecessary contact with study staff, so as to minimise bias due to a Hawthorne effect. Parents are interviewed every 3 months to record immunisation status and changes in breast feeding and childcare attendance.

An initial visit is undertaken once a child is delivered, preferably while the mother and child are still in hospital. At this visit, consent for participation is confirmed, an initial anterior nose swab and nappy swab are collected from the study child, and parents are taught the process for collecting these specimens, nasal swabs are collected from parents, diaries and study paperwork are reviewed and arrangements are made to retrieve cord blood, if available.

The Human Research Ethics Committees of the Children’s Health Queensland Hospital and Health Service, the Royal Brisbane and Women’s Hospital and The University of Queensland approved the study.

**Study sample**

Pregnant women are approached for enrolment of their newborn infants at antenatal clinics in two hospitals: one public (Royal Women’s Hospital, Brisbane) and one private (Northwest Private Hospital, Everton Park). These hospitals serve communities in northern metropolitan Brisbane, a city of more than two million people,
and every year each has approximately 6100 and 1700 deliveries, respectively. Exclusion criteria for enrolment and ongoing participation include gestational age at birth of less than 36 weeks, major congenital abnormalities, chronic heart, respiratory (excluding asthma), gastrointestinal, neurological or immunological disorders, parents unable to converse in English, living outside the Brisbane metropolitan region or planning to move from the area within the next 2 years. As a dynamic cohort study, children can leave and rejoin the cohort as required.15

Outcomes to be measured
Following the initial visit, the study family performs the following tasks: (1) completion of the daily symptom diary, (2) weekly collection and return by mail to the research laboratory of separate anterior nose and nappy swabs and (3) when the study child has an illness that meets specified criteria, complete an impact diary.

The daily symptom diary consists of a day-by-day tick box framework, and has been modified from an ARI daily diary used in previous studies to include diarrhoea for capturing episodes of AGE.15 35 For an ARI, one or more category A features (fever, wheezing, shortness of breath, pulmonary congestion or moist cough or medically diagnosed oitis media and/or pneumonia) or at least two category B features (runny nose or nasal congestion, sore throat, cough, muscle aches, chills, headache, irritability, decreased activity or lethargy or any vomiting) trigger impact diary completion. For AGE, parents record daily number of vomiting episodes followed by the number of loose stools. Diarrhoea, dehydration, lethargy or parent concern of any illness can trigger impact diary completion. These data will allow for a gastroenteritis severity score (modified Vesikari score) validated for ambulatory settings,36 to be calculated. The impact diary collects information on healthcare visits, including hospitalisations, diagnostic investigations, use of antibiotics, missed childcare and parental time away from work or usual activities.15 35

Parents collect an anterior nose swab and nappy swab from the study child once a week using a plastic- shaft, rayon budded swab, which comes with a transport tube with a foam pad reservoir soaked with viral transport medium (Virotocul MW950, Medical Wire & Equipment, Wiltshire, England). For the nose specimen, a single swab is used to sample each nostril. Both specimens are sent to the research laboratory by surface mail57 58 where they are stored in a −80°C freezer.

Nose and nappy specimens are batch tested for viruses using previously validated and reported real-time PCR assays, or PCR assays developed specifically for this study (table 1). Reverse transcription precedes PCR for RNA viruses. Specimens are spiked with a known concentration of Equine Herpes Virus before extraction to assess extraction efficiency and for the presence of PCR inhibitors. The quality of respiratory and stool specimen collection are assessed by evaluating for the presence of a marker of human genomic DNA, ERV3.37 Appropriate positive and negative controls are included in every run.

When more than one pathogen is detected simultaneously in respiratory and stool specimens, we perform a semiquantitative analysis of individual viral nucleic acids based on the cycle threshold (Ct) value.55 56 To further differentiate between human rhinovirus (HRV) types a nested PCR-based typing system encompassing partial

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Specimen type, pathogens and assay details, ORChID study, Brisbane, Queensland, Australia</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nose swab</strong></td>
<td><strong>Organism</strong></td>
</tr>
<tr>
<td>Rhinoviruses</td>
<td>Lu et al40&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>HRSV-A and HRSV-B</td>
<td></td>
</tr>
<tr>
<td>Influenza A and influenza B</td>
<td></td>
</tr>
<tr>
<td>Adenoviruses</td>
<td></td>
</tr>
<tr>
<td>Parainfluenza viruses I, II, III</td>
<td></td>
</tr>
<tr>
<td>HMPV</td>
<td></td>
</tr>
<tr>
<td>Bocavirus</td>
<td>Tozer et al47&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>hPyV-WU</td>
<td>Bialasiewicz et al49&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>hPyV-KI</td>
<td>Gunson et al40&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Coronavirus OC43</td>
<td></td>
</tr>
<tr>
<td>Coronavirus 229E</td>
<td></td>
</tr>
<tr>
<td>Coronavirus NL63</td>
<td></td>
</tr>
<tr>
<td>Coronavirus HKU1</td>
<td></td>
</tr>
<tr>
<td>Bordetella pertussis</td>
<td></td>
</tr>
<tr>
<td>Mycoplasma pneumonia</td>
<td></td>
</tr>
<tr>
<td>Chlamydial species</td>
<td></td>
</tr>
<tr>
<td><strong>Nappy swab</strong></td>
<td><strong>Organism</strong></td>
</tr>
<tr>
<td>Rhinoviruses</td>
<td>As for respiratory viruses</td>
</tr>
<tr>
<td>Coronavirus</td>
<td></td>
</tr>
<tr>
<td>Bocavirus</td>
<td>Maunula et al42&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>Enteroviruses</td>
<td></td>
</tr>
<tr>
<td>Saffold viruses</td>
<td></td>
</tr>
<tr>
<td>Rotaviruses</td>
<td></td>
</tr>
<tr>
<td>Noroviruses</td>
<td>As for respiratory viruses</td>
</tr>
<tr>
<td>Adenoviruses</td>
<td>To be developed for this study</td>
</tr>
<tr>
<td>Saporovirus</td>
<td></td>
</tr>
<tr>
<td>Astroviruses</td>
<td></td>
</tr>
<tr>
<td>Parechoviruses</td>
<td></td>
</tr>
<tr>
<td>Cosaviruses</td>
<td></td>
</tr>
<tr>
<td>Klassevirus</td>
<td></td>
</tr>
</tbody>
</table>

HRSV, human respiratory syncytial virus; HMPV, human metapneumovirus; hPyV, human polyomavirus.
sequence of the 5′ untranslated region (5′UTR) is being used. It accurately groups existing and newly identified HRV and human enterovirus strains into clades, which reflect current species assignments. Sequencing the 540 bp internal amplicon provides a rapid method to assign new sequences to a genotype.57 58

Nasal and nappy samples from children during illness periods where a known agent cannot be identified will be used for further investigations for the presence of as yet unidentified organisms. Pathogen-negative specimens will be prioritised on the basis of disease severity (hospital admission, fever). The methods employed for new pathogen detection will include pan-viral DNA microarrays59 and ‘next generation’ high-throughput sequencing.60 61

The endpoints of the first phase of this study are (1) documentation of clinical and epidemiological data from children with a variety of syndromic illnesses which, along with exhaustive diagnostic testing using sensitive molecular methods, will allow for detailed identification and characterisation of pathogen–disease associations, (2) establishing a well-characterised collection of stored specimens with pathogen-testing results, which will be invaluable for new virus discovery and (3) determining the pathogenesis of newly identified agents from the human respiratory and gastrointestinal tracts and better defining the role of known pathogens, such as HRVs, detected in asymptomatic children.

Sample size and data analysis plan
As respiratory syncytial virus (RSV) is the most widely recognised respiratory agent associated with severe disease in this age group, we chose it to use for sample size calculations. To calculate the proportion of pathogens detected that are RSV to within ±2.5%, we assumed the current proportion to be approximately 12.5%, that conservatively each participant will have an average of eight episodes of ARI over the course of the study,13 20 62 the intraclass correlation coefficient within individuals is 0.07 (calculated from Lambert et al53) and that 90% of subjects will remain enrolled at study completion (previous large cohort studies conducted by the investigators had retention rates of 97–98%).13 63 Consequently, with α=0.05 and power of 80%, we were required to enrol 138 infants. Similarly, if each subject returns 80% of their weekly nasal swabs,13 we will be able to estimate the proportion of all swabs positive for RSV to within ±0.5%.

Primary analyses will concern the calculation of incidence rates for ARI and AGE in study children for the cohort as a whole, and by age and seasonality. The collection of control material from children when they are without symptoms will allow us to determine what proportion of ARIs and AGE can be attributed to the presence of specific pathogens. Secondary analyses will assess the incidence and shedding duration of respiratory and gastrointestinal pathogens. Analyses will be conducted with mixed effects models with random intercepts and slope (time effect) for each participant. This method controls for the non-independence in outcomes from the same participant at different times and allows for heterogeneity between participants.

DISCUSSION
Infections are responsible for a significant burden during the early childhood years. We have presented here the study protocol and data analysis plans for an ambitious observational study: ORChID. The study has been designed to overcome identified weaknesses in previous research, and to use modern molecular techniques to identify infecting agents. The real value of this study is in the establishment of a biobank of thousands of respiratory and gastrointestinal specimens linked with daily clinical data. As well as the testing we outline here, we will in future be able to use the biobank for rapid and detailed assessments of the clinical significance of newly identified pathogens using highly sensitive molecular testing.

Strengths of the study include that it is community-based and uses parent-collected specimens returned to the diagnostic laboratory by surface mail for highly sensitive and specific real-time PCR testing for a comprehensive range of viruses. Systematic weekly sampling provides a control set of specimens, from both the same study child and all study children, during asymptomatic periods that can act as a control for specimens collected during periods of illness. This will allow us to quantify the attributable risk of pathogen detection to illness.

It would not be logistically or economically possible to conduct a study on this scale without using parent-collected specimens that are returned to the laboratory by surface mail. We and others have demonstrated that collection of respiratory specimens by non-healthcare workers is feasible19 41 64–69 and, when combined with sensitive molecular diagnostics, does not result in any overall significant reduction in sensitivity for virus detection.19 41 65 68 In a small randomised controlled trial, parent collection was no worse than a home visit by a healthcare worker for both having a specimen collected during an illness and having a virus identified when specimens were tested; the trend for both measures was improved collection and identification with parent collection.19 Further, the swabs we are using have been shown to be similarly sensitive for the detection of influenza nucleic acid when compared with flocked swabs combined with a universal transport media.70 We have found that return of specimens using surface mail did not result in reduced sensitivity when compared to immediate and maintained freezing of specimens, even over long distances.37 Mailing specimens has also been shown to be an efficient means of community-based research in other settings.66 67

There remain diagnostic gaps in the identification of the causative agent in ARI and AGE.38 34 We have been involved in the discovery and fuller description of newly
identified pathogens from the human respiratory and gastrointestinal tract. \textsuperscript{40} \textsuperscript{47} \textsuperscript{49} \textsuperscript{60} \textsubscript{71–86} The ORChID study will allow us to link specimens negative for known pathogens with symptoms and illness impact data, enabling us to prioritise specimens for new pathogen discovery. As well as the real possibility of discovering new agents, the epidemiology of known viruses, including their interactions,\textsuperscript{87} will be analysed.

Like most observational studies, recruitment is non-random, requiring assessment of potential biases and confounders prior to considering the internal and external validity of any findings. Where reported, oversampling of families from higher socioeconomic households appears common in research similar to this study,\textsuperscript{14} \textsuperscript{63} \textsuperscript{88} although differences in incidence and cost of illnesses mean that any impact on burden assessment is likely to be minimal and biased towards the null. We have attempted to avoid this selection bias by using a large, public hospital as one of our two recruitment sites. The role household socioeconomic status plays on incidence and disease burden will be assessed in the study analysis.

Recruitment for the study started on 12 August 2010 with the first study baby born on 15 September 2010. Recruitment will continue for the 2 years planned and is expected to conclude in December 2012. To date (31 August 2012), there have been 165 pregnant women recruited and 148 children enrolled. We are currently receiving over 350 nasal and nappy swabs a month, for a total return of more than 6000 of each specimen type. Batch testing of the nasal swabs for respiratory viruses has started with 5200 nasal swabs tested for 17 respiratory viruses to date, with laboratory staff blind to symptom presence in study subjects.

**CONCLUSION**

The ORChID study is well underway having started recruitment of pregnant women and their newborns for 2 years of intensive respiratory and gastrointestinal specimen collection linked with daily clinical data in 2010. Over 12 000 specimens have been returned from 148 enrolled infants, with blinded real-time PCR started for respiratory viruses. The study will provide unique insights on the acquisition of viruses, duration of shedding and the relationship of these to symptomatic illness. As well as this, specimens will be available for targeted discovery of new pathogens, and will provide a rich source of material to define the epidemiology of new agents when described by others.

**Author affiliations**

\textsuperscript{1}Queensland Paediatric Infectious Diseases Laboratory, Queensland Children’s Medical Research Institute, The University of Queensland and the Royal Children’s Hospital, Brisbane, Queensland, Australia

\textsuperscript{2}Queensland Health Immunisation Program, Communicable Diseases Branch, Queensland Health, Brisbane, Queensland, Australia

\textsuperscript{3}School of Population Health and the Queensland Children’s Medical Research Institute, The University of Queensland, Brisbane, Queensland, Australia

\textsuperscript{4}Departments of Molecular Microbiology and Pathology & Immunology, Washington University, School of Medicine, St. Louis, Missouri, USA

\textsuperscript{5}Microbiology Division, Pathology Queensland Central Laboratory, Queensland Health, Brisbane, Queensland, Australia

**Acknowledgements** We thank the families participating in the study, Queensland Paediatric Infectious Diseases laboratory staff: Asma Alsaleh, Jane Gaydon, Rebecca Rockett, Lebo Mhango, Hannah Cox, Rebecca Holding, Kevin Jacob and Claire Wang, and study volunteer support: Lynne Grimwood and Patricia Sloots.

**Contributors** KG is the study principal investigator. SBL, RW, DW, TPS, MDN and KG developed the successful grant application which supports the study. ALC and FAM are study nurses responsible for subject recruitment, enrollment and cohort maintenance. DMW, SB and IMM provide technical expertise for diagnostic methods. SBL drafted the manuscript. All authors critically revised the manuscript for important intellectual content, contributed to and approved the final version of the manuscript for publication.

**Funding** Funding support for cohort establishment and the respiratory testing component of ORChID was received from the Australian National Health and Medical Research Council (NHMRC project grant number: 615700) and from the Children’s Health Foundation Queensland (program grant number: 50006). Applications for further funding support for the testing of nappy specimens for a range of gastrointestinal viruses are in progress. SBL is supported by an Early Career Fellowship (1036231) from the NHMRC.

**Competing interests** None.

**Ethics approval** The Human Research Ethics Committees of the Children’s Health Queensland Hospital and Health Service, the Royal Brisbane and Women’s Hospital and The University of Queensland.

**Provenance peer review** Not commissioned; internally peer reviewed.


