BMJ Open

Childhood respiratory illness presentation and service utilisation in primary care: a six-year cohort study in Wellington, New Zealand, using natural language processing (NLP) software

Anthony Dowell,[1] Ben Darlow,[1] Jayden Macrae,[2] Maria Stubbe,[1] Nikki Turner,[3] Lynn McBain[1]

To cite: Dowell A, Darlow B, Macrae J, *et al*. Childhood respiratory illness presentation and service utilisation in primary care: a six-year cohort study in Wellington, New Zealand, using natural language processing (NLP) software. *BMJ Open* 2017;**7**:e017146. doi:10.1136/bmjopen-2017-017146

► Prepublication history and additional material are available. To view these files please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2017-017146).

Received 4 April 2017
Revised 23 June 2017
Accepted 29 June 2017

[1]Department of Primary Health Care and General Practice, University of Otago, Wellington, New Zealand
[2]Datacraft Analytics, Wellington, New Zealand
[3]Department of General Practice and Primary Health Care, University of Auckland, Wellington, New Zealand

Correspondence to
Professor Anthony Dowell;
tony.dowell@otago.ac.nz

ABSTRACT

Objectives To identify childhood respiratory tract-related illness presentation rates and service utilisation in primary care by interrogating free text and coded data from electronic medical records.

Design Retrospective cohort study. Data interrogation used a natural language processing software inference algorithm.

Setting 36 primary care practices in New Zealand. Data analysed from January 2008 to December 2013.

Participants The records from 77 582 children enrolled were reviewed over a 6-year period to estimate the presentation of childhood respiratory illness and service utilisation. This cohort represents 268 919 person-years of data and over 650 000 unique consultations.

Main outcome measure Childhood respiratory illness presentation rate to primary care practice, with description of seasonal and yearly variation.

Results Respiratory conditions constituted 46% of all child-general practitioner consultations with a stable year-on-year pattern of seasonal peaks. Upper respiratory tract infection was the most common respiratory category accounting for 21.0% of all childhood consultations, followed by otitis media (12.2%), wheeze-related illness (9.7%), throat infection (7.4%) and lower respiratory tract infection (4.4%). Almost 70% of children presented to their general practitioner with at least one respiratory condition in their first year of life; this reduced to approximately 25% for children aged 10–17.

Conclusion This is the first study to assess the primary care incidence and service utilisation of childhood respiratory illness in a large primary care cohort by interrogating electronic medical record free text. The study identified the very high primary care workload related to childhood respiratory illness, especially during the first 2 years of life. These data can enable more effective planning of health service delivery. The findings and methodology have relevance to many countries, and the use of primary care 'big data' in this way can be applied to other health conditions.

Strengths and limitations of this study

- ► This study uses a novel and validated natural language processing software inference algorithm to identify childhood respiratory illness presentation rates and service utilisation using primary care Big Data.
- ► The presentation and burden of childhood respiratory diseases in primary care has not previously been estimated with such a high degree of accuracy.
- ► The algorithm was designed to maximise specificity, thereby generating a conservative estimate of the burden of childhood respiratory disease in primary care by keeping false positives to a minimum.
- ► The methodology has relevance to many Organisation for Economic Co-operation and Development countries, and the use of primary care natural language processing in this way can be applied to other health conditions.
- ► This study analysed normal hours primary care general practitioner consultations. The exclusion of nurse-only and out-of-hours consultations may result in an underestimation of primary care respiratory presentation rates.

INTRODUCTION

Childhood is a crucial period for development and well-being. A healthy start to life reduces adulthood morbidity and enhances participation in society.[1–5] Physical illness is an important risk factor for poor health outcomes.[6] Globally, primary care is used by all children,[7] but there are currently little published data of detailed morbidity and utilisation patterns in community settings.

Respiratory illness contributes substantially to childhood morbidity, yet despite the plethora of studies of general respiratory epidemiology few data exist describing the burden of respiratory tract-related illness in routine primary care. Children under five present up to six times a year with acute respiratory infections[8] and high prevalence rates are noted for asthma[9] and otitis media.[10] Such data are, however, mainly reliant on survey responses and parental report. These reports also lack precision regarding individual respiratory conditions, symptom severity, longitudinal patterns and variance related to age and seasonality. These data are needed to effectively plan primary healthcare service delivery. More detailed hospitalisation data are available[11]; however, these represent an unknown proportion of all cases and are based on diagnostic coding of uncertain accuracy.

International data suggest that respiratory tract-related conditions constitute 20%–25% of all general practitioner (GP) consultations, with higher rates in those under 25 years.[12 13] These data are based on GP self-report, and accuracy may be limited by the competing demands of reporting, meeting patient needs and practice management tasks. Wide variance has been reported in how GPs describe the reason for encounter.[13]

Improved understanding of primary care childhood respiratory illness presentation could enable more systematic approaches to care and resource allocation, and a context for exploring important social and ethnic variations in hospitalisation rates.[11 14] In the Organisation for Economic Co-operation and Development (OECD) countries, conditions such as bronchiolitis, asthma, upper respiratory tract infections (URTI) and pneumonia make up over 40% of ambulatory sensitive hospitalisation; admissions considered preventable through interventions delivered outside of hospitals, predominantly within primary care.[6 11 15 16]

More accurate assessment of illness presentation and service utilisation could be obtained by analysing consultation notes within electronic medical records (EMR) common in OECD primary care settings. While there has been some exploration of the potential for 'big data' assessment of general practice workload,[17] these data have not previously been used to analyse childhood respiratory service utilisation due to difficulties with extracting and analysing both structured and unstructured data available (primarily clinical consultation notes). The development of novel software has enabled the exploration of New Zealand (NZ) EMR data.[18 19]

This study aimed to interrogate data from EMR to identify primary care presentation and service utilisation related to common childhood respiratory tract conditions and their complications.

METHODS

Design

A natural language processing (NLP) software inference algorithm was developed to interrogate quantitative and qualitative cross-sectional and retrospective cohort data from EMR.[18 19]

Setting and participants

Figure 1 illustrates the creation and analysis of the data set. In NZ, there is universal enrolment with a primary care practice. As more fully described in an earlier publication outlining the development of the design,[17] the study was conducted in the Wellington region of NZ, a mixed urban and rural setting. It consisted of 36 consenting practices of 60 in total from two primary health organisations (PHOs). This comprised 75% of the total childhood population under 18 years of age of these combined PHOs. There was a total of 77 467 children enrolled in these practices over the study period between 1 January 2008 and 31 December 2013, including children who both joined and left this cohort during this period. Changes included births, deaths, turning 18 years or moving into or out of a practice. This cohort represented 268 919 person-years.

Data were collected directly from EMR using software which automates the extraction, and secure transmission of large data sets. The data set comprised records from consultations generated during both standard office hours and out-of-hours practice. Data were extracted from the EMR for all child-GP consultations at consenting practices during the study period (n=687 136). Each consultation record was connected to an individual's National Health Index number. The is a unique identifier assigned to every person who uses health services in NZ and enabled records to be matched between data sets. Consultations for which there were poor quality data (2439 consults from 256 children) were excluded. Out-of-hours consultations (n=34 584) were not analysed due to differing participation in out-of-hours services by the practices. All data were analysed within the PHO which has rigorous protocols in place to ensure patient confidentiality. The research team had no access to identifiable data.

Process

Each of the 650 123 clinical consultation notes was interrogated by a software inference algorithm and hierarchical classification system described previously.[19] The algorithm classified consultation records using: clinical information recorded by GPs and practice nurses, any recorded Read code diagnostic classifications and prescribing information. The first level of the hierarchy divided all consultations into either 'respiratory' or 'not respiratory.' Note that 'respiratory' here included all respiratory tract-related conditions and presentations and the associated complication of otitis media. The 'not respiratory' category included consultations where the primary presentation and diagnosis was for conditions such as injury or gastroenteritis, and consultations in which the respiratory system was examined and screened, but no signs, symptoms or diagnoses were recorded. These screening consultations were excluded so that the burden of respiratory tract illness estimate was not inflated by consultations which did not result from a respiratory tract illness or its complications.

The second level of the hierarchy subclassified consultations into one or more specific respiratory categories. These categories were determined by a group of clinical experts; consideration was given to the degree to which conditions could be mapped to high prevalence (that which is common) and/or responsible for

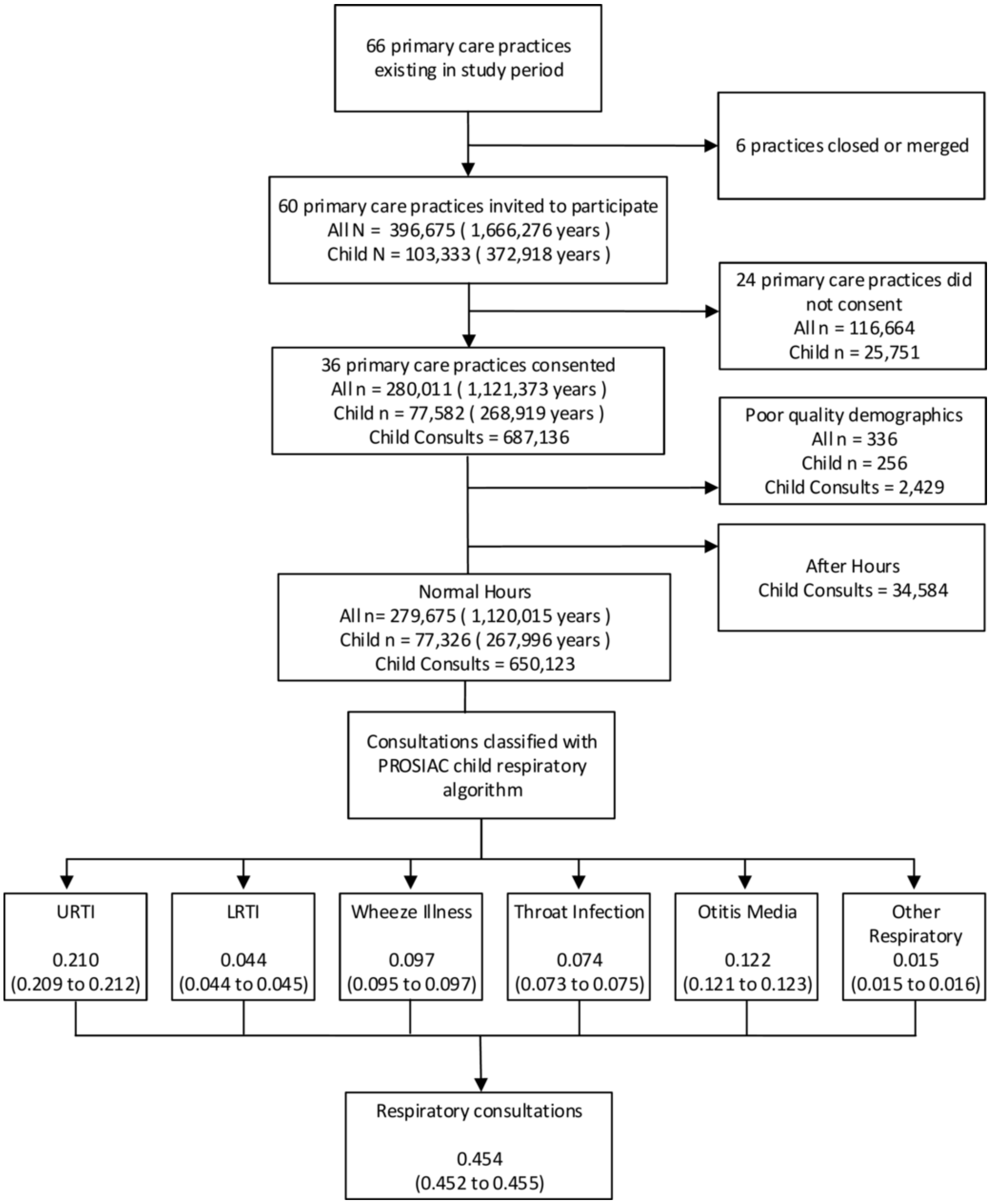
66 primary care practices existing in study period
6 practices closed or merged
60 primary care practices invited to participate
All N = 396,675 (1,666,276 years)
Child N = 103,333 (372,918 years)
24 primary care practices did not consent
All n = 116,664
Child n = 25,751
36 primary care practices consented
All n = 280,011 (1,121,373 years)
Child n = 77,582 (268,919 years)
Child Consults = 687,136
Poor quality demographics
All n = 336
Child n = 256
Child Consults = 2,429
After Hours
Child Consults = 34,584
Normal Hours
All n= 279,675 (1,120,015 years)
Child n = 77,326 (267,996 years)
Child Consults = 650,123
Consultations classified with PROSIAC child respiratory algorithm
URTI
0.210
(0.209 to 0.212)
LRTI
0.044
(0.044 to 0.045)
Wheeze Illness
0.097
(0.095 to 0.097)
Throat Infection
0.074
(0.073 to 0.075)
Otitis Media
0.122
(0.121 to 0.123)
Other Respiratory
0.015
(0.015 to 0.016)
Respiratory consultations
0.454
(0.452 to 0.455)

Figure 1 Selection of child-GP consultation notes and results from analysis. More than one respiratory condition can be classified in each consultation. GP, general practitioner; LRTI, lower respiratory tract infection; URTI, upper respiratory tract infection.

significant morbidity and hospitalisation (that which is important). The six categories were (1) URTI; (2) lower respiratory tract infections (LRTI); (3) wheeze-related illnesses; (4) throat infections; (5) otitis media; and (6) other respiratory conditions. The conditions included within each diagnostic category are presented in online supplementary appendix 1.

The algorithm was trained, tested and validated using three independent gold standard data sets of 1200 consultation records which had been independently classified by two general practice clinical experts (AD and LM). The algorithm was designed to replicate the judgements made by these clinical experts. Development aimed to optimise specificity while maximising sensitivity to minimise the occurrence of false positives. The algorithm's sensitivity, specificity, positive and negative predictive values, and F measure for each of the diagnostic categories against a gold standard validation set of 1200 consultation records have been published previously.[18]

Analysis

The demographic characteristics of age, gender, ethnicity (NZ indigenous Māori, Pacific, other) and NZ Deprivation Index (a measure of socioeconomic deprivation[20]) of the cohort (n=77 326) were compared with those of all children enrolled within the two PHOs (n=103 333) and the NZ population using national census data.

The proportion of primary care consultations for children aged 18 years and below which were related to the six specific respiratory conditions outlined above was obtained from the data set using the algorithm. The utilisation of services for these six conditions was analysed by demographic characteristics. Consultation rates are expressed per 1000 child-years observed due to the differing length of time individuals might be participants in the cohort. Patients were observed for the period in which they were enrolled in a participating practice; this was calculated from the date of a child's first visit to a practice until they were removed from the enrolment register. Both deprivation and ethnicity status were taken as the last ethnicity and deprivation recorded from the GP records. Consultation rates were adjusted for sensitivity and specificity of the algorithm[18]and a direct standardisation method was applied to level 2 ethnicity and socioeconomic deprivation quintiles against NZ Census 2013 data. Estimates of true rates were made using final test sensitivity and specificity results for each classification category using the method described by Rogan and Gladen.[21] All data aggregation, transformation, cleaning and storage were done in Microsoft SQL Server, and statistical analysis was undertaken in R using packages including boot, epiR, combinat, stats, tm, RWeka, slam, SnowballC and caret.[22]

STrengthening the Reporting of OBservational studies in Epidemiology (STROBE) guidelines were followed.

RESULTS

The demographic characteristics of the study cohort closely matched those of the enrolled population (online supplementary appendix 1). The age distribution of the study cohort also closely matched the national comparison data. Compared with national census data, the study cohort had a greater proportion of children from the least deprived quintile grouping (32% vs 25%) and a lower proportion of Māori (17% vs 22%).

From the 650123 consultations reviewed, the true rate of presentation for a respiratory tract condition or complication was calculated to be 45.4% of all consultations for children under 18 years of age (figure 1). URTI was the most common respiratory tract-related category represented in 21.0% (95% CI 20.9% to 21.2%) of all consultations, followed by otitis media (12.2% CI 12.1% to 12.3%), wheeze-related illness (9.7% CI 9.5% to 9.7%), throat infection (7.4% CI 7.3% to 7.5%) and LRTI (4.4% CI 4.4% to 4.5%). Other respiratory tract-related classifications accounted for just 1.5% of all consultations. One respiratory tract-related condition was classified in 27.6% of all consultations, two in 7.0%, three in 0.8% and greater than three in 0.1%. The rates of child respiratory tract condition or complication consultation were 1101 per 1000 person-years observed for all respiratory tract conditions and complications, 509 per 1000 for URTI, 107 per 1000 for LRTI, 235 per 1000 for wheeze illness, 180 per 1000 for throat infections, 296 per 1000 for otitis media and 36 per 1000 for other respiratory conditions. The incidence of both respiratory and non-respiratory tract-related consultations remained stable throughout the study period with a consistent pattern of seasonal peaks and troughs (figure 2). The respiratory tract-related consultation rate was highest in the Southern Hemisphere winter month of August, and lowest in January (figure 3). Non-respiratory consultations followed a similar pattern but with shallower peaks and troughs. Respiratory tract-related conditions explained 64.4% of the annual seasonal variation in child consultation rates. All respiratory tract-related conditions which were subclassified followed a similar pattern of peaks in August and troughs in January except for 'other' respiratory tract

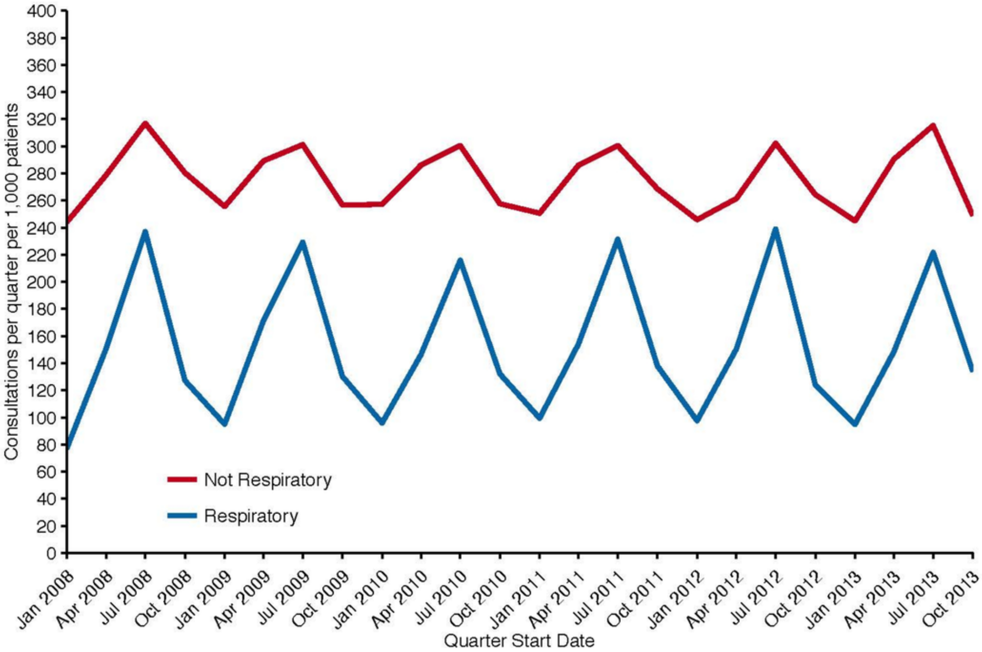

Figure 2 Respiratory tract-related and non-respiratory consultations per quarter per 1000 enrolled children, January 2008 to December 2013.

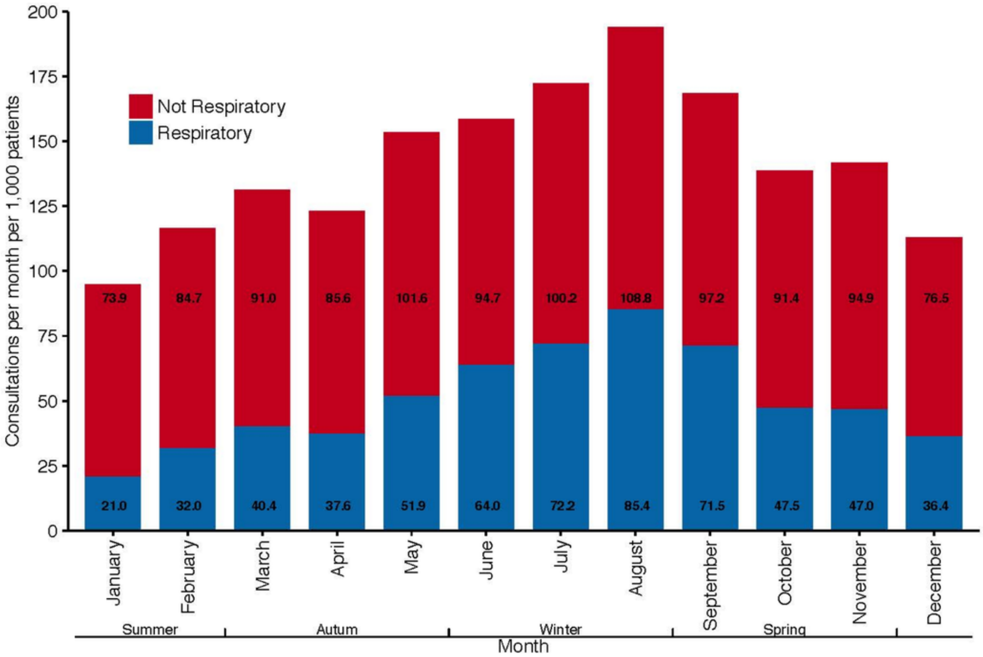

Figure 3 Mean respiratory tract-related and non-respiratory consultations per month per 1000 enrolled children. January 2008 to December 2013 demonstrating seasonal variation.

conditions which were highest in December and lowest in April (figure 4). Figures 4 and 5 present the annual variation in respiratory tract condition-related presentation for each classification category.

Respiratory tract-related consultations occurred throughout childhood, but at much greater frequency during the first 2 years of life (figure 6). During the first year of life, 73.5% of children presented to their GP with at least one respiratory tract-related condition. Following the second year of life, the presentation of all respiratory tract-related conditions decreased with increasing age. Of children aged 10–17 years, 22.5% presented with at least one respiratory condition (figure 6). The mean number of presentations for respiratory tract infection for an individual was 2.6 per year in those below 2 years, 2.1 per year in those aged 3–5 years and 1.5 per year in those over 15 years.

In figure 6, each facet includes children who were enrolled within that age band for a 12-month period (eg, from the day they turned 1 until the day before

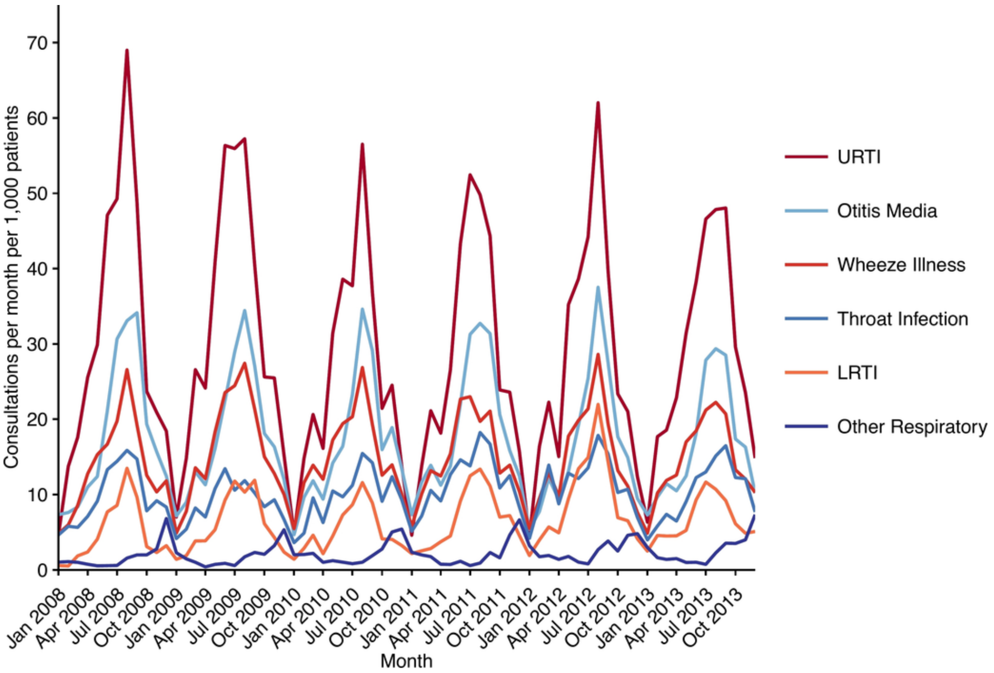

Figure 4 Yearly variation of consultations per month per 1000 enrolled children for each respiratory tract-related illness category—January 2008 to December 2013. LRTI, lower respiratory tract infection; URTI, upper respiratory tract infection.

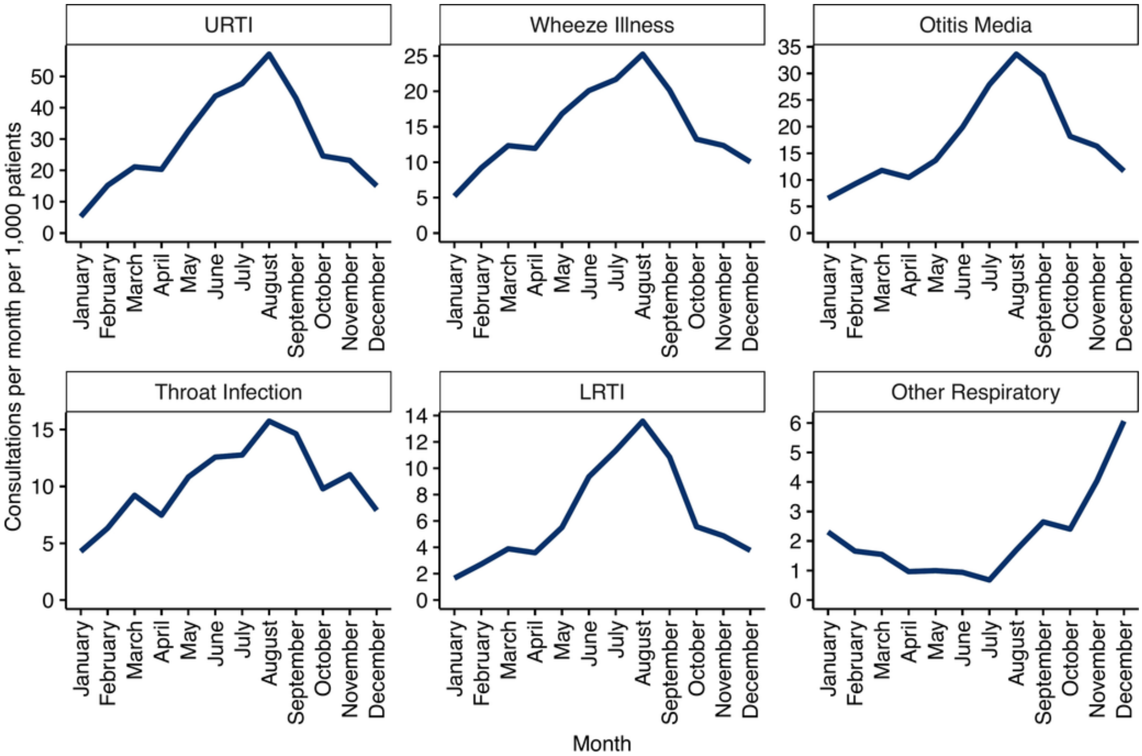

Figure 5 Mean consultations per month per 1000 enrolled children for each respiratory tract-related illness category—January 2008 to December 2013. LRTI, lower respiratory tract infection; URTI, upper respiratory tract infection.

they turned 2). The cohort of children under 1 is small because many children do not enrol until they are over 3 months old and may therefore only be enrolled for 9 months before turning 1.

DISCUSSION

This is the first study to assess the primary care incidence and service utilisation of childhood respiratory tract-related illness in such a large cohort observing over 250 000

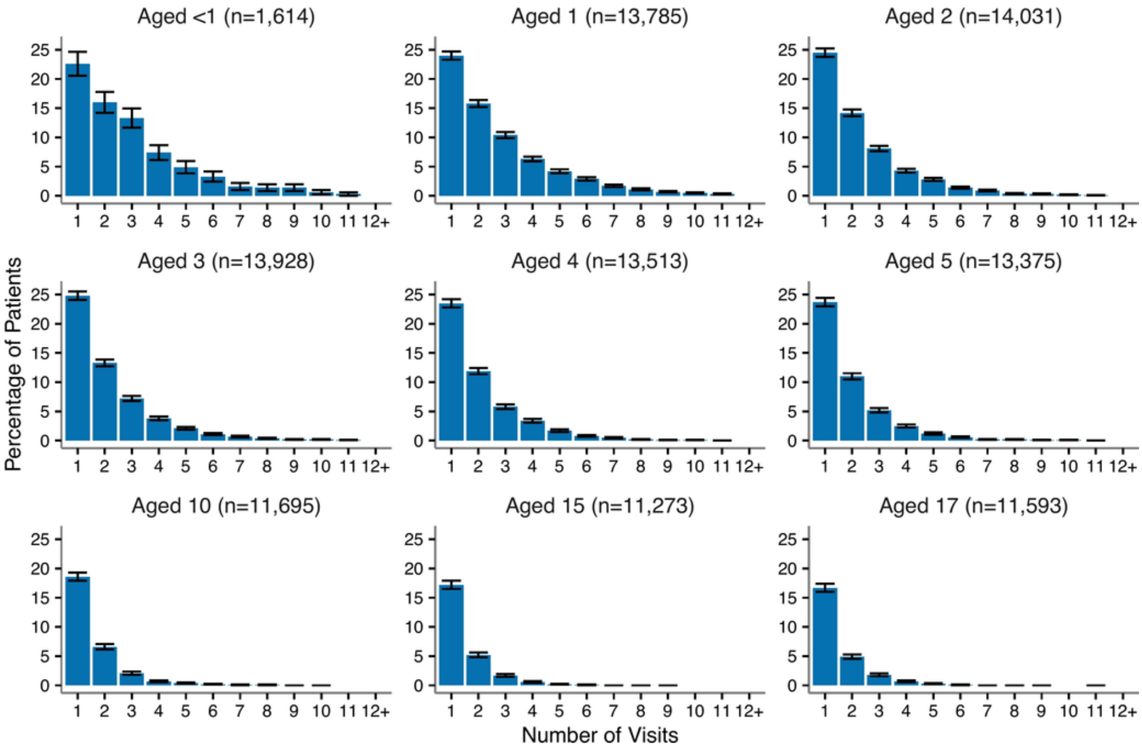

Figure 6 Respiratory tract-related consultation frequency by selected age cohort.

person-years and more than 650 000 unique consultations. Using a novel and validated method of interrogating EMR free text, this study found that respiratory tract-related conditions constituted 45.4% of all child-GP consultations. This quantifies the very high volume of childhood respiratory tract-related consultations and workload in general practice, especially during the first 2 years of life. These data can enable more effective planning of primary care service delivery and indicate areas in which to focus preventive programmes. The study also highlights the high presentation rates to primary care of those respiratory conditions which frequently present for hospital admission.

Comparison with other studies

The presentation rate of respiratory illness and pattern of seasonal peaks was remarkably stable across the 6 years included in this data set and was unchanged by events such as the H1N1 influenza pandemic of 2009. Consistent with findings from Australian,[8 23] and Chilean surveys,[24] the presentation of nearly all respiratory consultations more than doubled during the winter months, and providing a comparator with 'seasonal' changes between wet and dry seasons seen in tropical regions.[25] Respiratory consultations classified as 'other' had a different pattern with a peak in spring, consistent with seasonal allergies being the primary contributor to this classification group. The high presentation rates of wheeze-related illness highlight the importance of these conditions in primary care management, and align with the high community burden of wheeze identified from other cohort studies.[9 26] The prevalence of otitis media is consistent with other studies.[27]

While we could find no other studies that used an NLP methodology, our findings are consistent with the stated underestimate of respiratory illness prevalence reported from a recent primary care study in Ireland using Read code data only.[28]

Childhood respiratory conditions feature highly as a cause of hospital admissions thought to be amenable to preventive activity in primary care. These data suggest only small numbers of children are hospitalised compared with the high volume of respiratory conditions managed within general practice.[11 29] It is possible that paediatric hospitalisations thus represent appropriate care for children with severe respiratory illness, or significant socioeconomic difficulty rather than reflecting unmet need within primary care.

These data also provide information about consulting patterns across the childhood life course, highlighting the frequency of consultation in the early years and in particular during the first 2 years of life. While high consultation frequency in the earlier years has been recognised previously, data have usually been grouped within a birth to 5 years age band,[30] or focused on a single year of life.[31] This study highlights the degree of primary care contact children have in their first 2 years of life. Strategic management of clinical contact during this time may improve care delivery and enable a balance between preventive and acute care activity.

Strengths and limitations of study

This study examined a very large data set of child-GP consultations including clinical consultation notes, diagnostic codes and prescribing information by way of a software inference algorithm which performed with similar accuracy to clinical experts.[18] The algorithm was designed to maximise specificity, thereby generating a conservative estimate of the burden of childhood respiratory disease in primary care by keeping false positives to a minimum. The presentation and burden of childhood respiratory diseases in primary care has not previously been estimated with such a high degree of accuracy.

Computer algorithms using NLP have previously been found to be considerably more accurate than relying on diagnostic codes to make respiratory diagnoses.[32]

Data representing 75% of the child population enrolled within two large PHOs were analysed. The study data set included over 650 000 consultation records (representing over 260 000 person-years of data), and the age, ethnic and socioeconomic characteristics of children enrolled within participating practices were almost identical to those of children enrolled in practices which declined, and to the broader NZ population.

This study analysed normal hours primary care GP consultations. The exclusion of nurse-only and out-of-hours consultations may result in an underestimation of primary care respiratory tract-related presentation rates. Nurse-only consultations were excluded because only a small proportion of nursing records relate to direct clinical consultations and it was not possible for the algorithm to distinguish these from non-clinical records such as telephone calls.[18] The data set excluded out-of-hours consultations because out-of-hours care is also provided elsewhere to children from consenting practices, consequently PHO out-of-hours data were incomplete.

Although validation of the software algorithm against the gold standard of two expert clinicians' opinion indicated that it had excellent accuracy, particularly with respect to classification of consultations as respiratory tract related or non-respiratory, this methodology can only provide an estimation of the presentation of these respiratory conditions and resultant service utilisation. It would be impractical to manually check the several hundred thousand consultation records included in the full data set. Notwithstanding this, it is debatable whether manual record review would generate a more accurate estimation.[33 34]

The gold standard used for this study was the GP's stated diagnosis, matched to GP experts' assessment based on clinical data available. There is the potential for error in the GP decision making,[35 36] and is limited by the amount and detail of the recorded information by each GP. While recognising this limitation, accuracy of GP diagnosis was not the prime purpose of this study, the intention was to estimate illness and health service utilisation as identified by the GP records.

The algorithm requires common conditions with sufficient prevalence to allow effective training. Therefore, some important but less prevalent conditions (eg, croup, pertussis and pneumonia) required to be grouped. As a result, the study cannot give estimations of the burden of some diseases, which although relatively rare have considerable morbidity. The algorithm was not designed to differentiate between types of wheeze-related illness given the variation and debate among clinicians regarding the classification of wheeze presentations for younger children.[37]

Conclusions and policy implications

These data have demonstrated a clear and consistent pattern in general practice utilisation for children with respiratory tract-related illness. Results of this type can assist with general practice workforce planning, and inform debate about current presentation and triage models seen in primary care. The study also highlighted the burden of respiratory disease carried by the youngest members of society and reinforces calls to focus prevention and health promotion campaigns on early stages of the maternal and child health continuum.

The methodology used can be applied to provide similar estimates of respiratory and other conditions and workload across an entire population at all ages. The use of NLP software in this way also provides a tool for health service planning in primary care which would have increasing application across a wide range of countries.

Acknowledgements The authors gratefully acknowledge the primary care practices which consented to their consultation records being included in the study data set, and the primary health organisations which permitted use of proprietary software and resources.

Contributors AD, JM, MS, LM and NT conceived the study. All authors contributed to the development of the overall study methodology. AD, LM and NT provided clinical input into the algorithm design. JM designed and built the natural language processing tools. JM programmed and trained the algorithm. AD and LM classified the consultation records in the gold standard sets. BD and AD were the principal writers of the manuscript. All authors reviewed and revised the manuscript and approved its final version. All authors had full access to all of the data (including statistical reports and tables) in the study and can take responsibility for the integrity of the data and the accuracy of the data analysis.

Funding This work was supported by a New Zealand Lotteries Health Research Grant. The funding body had no role in the collection or analysis of data or the preparation of this manuscript.

Competing interests All authors have completed the Unified Competing Interest form at www.icmje.org/coi_disclosure.pdf (available upon request from the corresponding author) and declare support from New Zealand Lotteries Health Research for the submitted work. LM is a director of Compass Health Wellington Trust that might have an interest in the submitted work. No other relationships or activities could appear to have influenced the submitted work.

Patient consent No identifiable patient data used in the output of the study. Patient consent for collection and utilisation of non-identifiable data is contained within the enrolment forms collected by General Practices when patients enrol with their GP.

Ethics approval This study was approved by the University of Otago Ethics Committee (H13/044).

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data are available.

REFERENCES

1. Hertzman C, Siddiqi A, Hertzman E, *et al*. Tackling inequality: get them while they're young. *BMJ* 2010;340:346–8.
2. Lynch JW, Kaplan GA, Cohen RD, *et al*. Childhood and adult socioeconomic status as predictors of mortality in Finland. *Lancet* 1994;343:524–7.
3. Marmot MG, Allen JL, Goldblatt P, *et al*. *Fair society healthy lives: Strategic review of health inequalities in England post*. 2010.
4. Walker SP, Wachs TD, Grantham-McGregor S, *et al*. Inequality in early childhood: risk and protective factors for early child development. *Lancet* 2011;378:1325–38.
5. Bethell CD, Newacheck P, Hawes E, *et al*. Adverse childhood experiences: assessing the impact on health and school engagement and the mitigating role of resilience. *Health Aff* 2014;33:2106–15.
6. *The Green Paper for Vulnerable Children. Every child thrives, belongs, achieves*. Wellington, NZ: New Zealand Government, 2011.
7. Ministry of Health. *New Zealand Health survey: annual update of key findings 2012/13*. Wellington: Ministry of Health, 2013.
8. Chen Y, Kirk MD. Incidence of acute respiratory infections in Australia. *Epidemiol Infect* 2014;142:1355–61.
9. Asher MI, Montefort S, Björkstén B, *et al*. Worldwide time trends in the prevalence of symptoms of asthma, allergic rhinoconjunctivitis, and eczema in childhood: ISAAC phases one and three repeat multicountry cross-sectional surveys. *Lancet* 2006;368:733–43.
10. Gribben B, Salkeld LJ, Hoare S, *et al*. The incidence of acute otitis media in New Zealand children under five years of age in the primary care setting. *J Prim Health Care* 2012;4:205–12.
11. Craig E, Anderson P, Jackson G, *et al*. Measuring potentially avoidable and ambulatory care sensitive hospitalisations in New Zealand children using a newly developed tool. *N Z Med J* 2012;125:38-50.
12. Davis P, Suaalii-Sauni T, Lay-Yee R, *et al*. *Pacific patterns in Primary Health Care: a comparison of Pacific and all patient visits to doctors: the National Primary Medical Care survey (NatMedCa): 2001/02*. Wellington: Ministry of Health, 2005.
13. Britt H, Britt H, Miller G, *et al*. *General Practice Activity in Australia 2011-12: BEACH,Bettering the Evaluation And Care of Health*. Sydney University Press, 2012.
14. Telfar Barnard L, Baker M, Pierse N, *et al*. *The impact of respiratory disease in New Zealand: 2014 update*. Wellington: The Asthma Foundation, 2015.
15. Dowell A, Turner N. Child health indicators: from theoretical frameworks to practical reality? *Br J Gen Pract* 2014;64:608–9.
16. Gill PJ, Goldacre MJ, Mant D, *et al*. Increase in emergency admissions to hospital for children aged under 15 in England, 1999-2010: national database analysis. *Arch Dis Child* 2013;98:328–34.
17. Hobbs FR, Bankhead C, Mukhtar T, *et al*. *The Lancet* 2016Clinical workload in UK primary care: a retrospective analysis of 100 million consultations in England;14.
18. MacRae J, Darlow B, McBain L, *et al*. Accessing primary care Big Data: the development of a software algorithm to explore the rich content of consultation records. *BMJ Open* 2015;5:e008160.
19. MacRae J, Love T, Baker MG, *et al*. Identifying influenza-like illness presentation from unstructured general practice clinical narrative using a text classifier rule-based expert system versus a clinical expert. *BMC Med Inform Decis Mak* 2015;15:1.
20. Salmond C, Crampton P, King P, *et al*. NZiDep: a New Zealand index of socioeconomic deprivation for individuals. *Soc Sci Med* 2006;62:1474–85.
21. Rogan WJ, Gladen B. Estimating prevalence from the results of a screening test. *Am J Epidemiol* 1978;107:71–6.
22. Team RC. R Foundation forStatistical Computing. *R: A language and environment for statistical computing*. Vienna, Austria, 2013. ISBN 3-900051-07-0.
23. Chen Y, Williams E, Kirk M. *Risk factors for acute respiratory infection in the Australian Community*. 2014.

24. Astudillo P, Mancilla P, Olmos C, *et al*. [Epidemiology of pediatric respiratory consultations in Santiago de Chile, from 1993 to 2009]. *Rev Panam Salud Publica* 2012;32:56–61.
25. Rosa AM, Ignotti E, Botelho C, *et al*. Respiratory disease and climatic seasonality in children under 15 years old in a town in the brazilian Amazon. *J Pediatr* 2008;84:543–9.
26. Taussig LM, Wright AL, Holberg CJ, *et al*. Tucson Children's Respiratory Study: 1980 to present. *J Allergy Clin Immunol* 2003;111:661–75.
27. Mahadevan M, Navarro-Locsin G, Tan HK, *et al*. A review of the burden of disease due to otitis media in the Asia-Pacific. *Int J Pediatr Otorhinolaryngol* 2012;76:623–35.
28. Molony D, Beame C, Behan W, *et al*. 70,489 primary care encounters: retrospective analysis of morbidity at a primary care centre in Ireland. *Ir J Med Sci* 2016;185:805–11.
29. Craig E, Adams JA, Oben G, *et al*. *The health status of children and young people in the Hutt Valley and Capital and Coast DHBs*. Dunedin, New Zealand: NZ Child and Youth Epidemiology Service, 2014.
30. Saxena S, Majeed A, Jones M. Socioeconomic differences in childhood consultation rates in general practice in England and Wales: prospective cohort study. *BMJ* 1999;318:642–6.
31. Golenko XA, Shibl R, Scuffham PA, *et al*. Relationship between socioeconomic status and general practitioner visits for children in the first 12 months of life: an Australian study. *Aust Health Rev* 2015;39:136–45.
32. Wu ST, Sohn S, Ravikumar KE, *et al*. Automated chart review for asthma cohort identification using natural language processing: an exploratory study. *Ann Allergy Asthma Immunol* 2013;111:364–9.
33. McColm D, Karcz A. Comparing manual and automated coding of physicians quality reporting initiative measures in an ambulatory EHR. *J Med Pract Manag* 2010;26:6–12.
34. Gorelick MH, Knight S, Alessandrini EA, *et al*. Lack of agreement in pediatric emergency department discharge diagnoses from clinical and administrative data sources. *Acad Emerg Med* 2007;14:646–52.
35. Peabody JW, Luck J, Jain S, *et al*. Assessing the accuracy of administrative data in health information systems. *Med Care* 2004;42:1066–72.
36. O'Malley KJ, Cook KF, Price MD, *et al*. Measuring diagnoses: ICD code accuracy. *Health Serv Res* 2005;40:1620–39.
37. Brand PL, Baraldi E, Bisgaard H, *et al*. Definition, assessment and treatment of wheezing disorders in preschool children: an evidence-based approach. *Eur Respir J* 2008;32:1096–110.