BMJ Open

Quantifying the incidence and burden of herpes zoster in New Zealand general practice: a retrospective cohort study utilising a natural language processing software inference algorithm.

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-021241
Article Type:	Research
Date Submitted by the Author:	18-Dec-2017
Complete List of Authors:	Turner, Nikki; University of Auckland, General Practice and Primary Health Care MacRae, Jayden; Datacraft Analytics Nowlan, Mary; University of Auckland Tamaki Campus, General Practice and Primary Health Care McBain, Lynn; University of Otago, Wellington, Stubbe, Maria; University of Otago, Primary Health Care and General Practice Dowell, Anthony; Wellington School of Medicine and Health Sciences,
Keywords:	herpes zoster, PRIMARY CARE, BIOTECHNOLOGY & BIOINFORMATICS, VIROLOGY

Quantifying the incidence and burden of herpes zoster in New Zealand general practice: a retrospective cohort study utilising a natural language processing software inference algorithm.

N M Turner, Associate Professor[1],

J MacRae, Chief Scientist[2]

M L Nowlan, Project Manager[1],

L McBain, Senior Lecturer[3],

M Stubbe, Senior Lecturer[3]

A Dowell, Professor[3]

1. Department of General Practice and Primary Care, University of Auckland, Private Bag 92019, Victoria Street West, Auckland 1142, New Zealand .
2. DataCraft Analytics Limited, PO Box 17271, Karori, Wellington 6147
3. Department of Primary Care and General Practice, University of Otago Wellington, PO Box 7343 Newtown Wellington 6242 New Zealand

Correspondence to Dr. N Turner; n.turner@auckland.ac.nz

Abstract

Objective: To investigate the incidence of primary care presentations for herpes zoster (zoster) in a representative New Zealand population and to evaluate the utilisation of primary health care services following zoster diagnosis.

Design: A cross-sectional retrospective cohort study used a natural language processing software inference algorithm to identify general practice consultations for zoster by interrogating 22 million electronic medical record (EMR) transactions routinely recorded from January 2005 to December 2015. Data-linking enabled analysis of the demographics of each case. The frequency of doctor visits were assessed prior to and after the first consultation diagnosing zoster to determine health service utilisation.

Setting: General practice, using EMRs from two primary health organisations located in the lower North Island, New Zealand.

Participants: Thirty-nine general practices consented interrogation of their EMRs to access de-identified records for all enrolled patients. Out-of-hours and practice nurse consultations were excluded.

Main outcome measures: The incidence of first and repeated zoster-related visits to the doctor across all age groups and associated patient demographics. To determine whether zoster affects workload in general practice.

Results: Overall, for 6,189,019 doctor consultations, the incidence of zoster was 48.6 per 10,000 patient years (95% CI 47.6 to 49.6). Incidence increased from the age of 50 years to a peak rate of 128 per 10,000 in the 80-90 years age group and was significantly higher in females than males ($p < 0.001$). Over this 11 year period, incidence increased gradually, notably in those aged 80-85 years. Only 19% of patients had one or more follow-up zoster consultations within 12 months of a zoster index consultation. The frequency of consultations, for any reason, did not change between periods before and after the diagnosis.

Conclusions: Zoster consultations in general practice are rare and the burden of these cases on overall general practice caseload is low.

Article summary

Strengths and limitations of the study

- This study used a novel and validated natural language processing software inference algorithm to identify herpes zoster presentation rates and service utilisation using primary care electronic medical records over an eleven year period.
- Despite a low frequency of zoster cases, the large data set enabled analysis of rates of zoster incidence by age bands and different demographics across the whole time period.
- The algorithm was designed to maximise specificity and accuracy, thereby generating a conservative estimate of the burden of zoster presentations in primary care by keeping false positives to a minimum.
- The gold standard for this study was based on doctor decision making, and the algorithm is limited by the quantity and detail of the recorded information in each consultation.
- 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 rates.

Introduction¶

Infection with varicella-zoster virus (VZV) establishes life-long persistence in sensory nerve ganglia. When the immune system is unable to maintain the suppression of the virus, it manifests as a clinical syndrome known as herpes zoster (zoster; shingles).[1] About one third of the population experiences zoster, with greater incidence in older age groups.[2] The major risk factor for zoster is a decline in VZV immunity as cell-mediated immunity wanes with age.[3,4] Cohort studies, from at least 22 countries, give incidence of zoster in the general population from 3 to 5 per 1000 person-years [5–10], and a lifetime incidence of 20 - 30%.[2] Much higher incidence is reported in older adults, increasing from 50 – 60 years of age, and ranging from 8 – 12 per 1000 person-years.[2,7] Around half of the zoster cases over 70 years of age will develop postherpetic neuralgia (PHN).[11] The main rationale for the use of a zoster vaccine is to prevent the long-lasting effects of PHN. A live attenuated zoster vaccine is internationally available for adults over the age of 50 years, and is on the national schedule in some countries, notably, the United Kingdom and Australia at the age of 70 years, and recommended in the USA from age 60 years.[1]

Several studies have reported increasing incidence rates over time across all age groups [6], in countries both with [12–14] and without childhood varicella immunisation programmes.[2,15,16] The World Health Organization recommends, when considering the optimal age and dosing schedule for zoster vaccines, that countries should take into consideration the age-dependent burden of disease, vaccine effectiveness, duration of protection and cost effectiveness of the vaccine.[17] While there is a body of research around zoster and the burden of zoster disease, this is almost exclusively based on observational studies from administrative databases or health records, which will underestimate cases not coming to medical attention.[6] Decisions around vaccination strategies for countries remain hindered by a more complete understanding of the burden of the disease for a population.

The objective of this study was to interrogate data from general practice electronic medical records (EMR) to identify zoster-related primary care presentations and service utilisation associated with consultations for zoster-related conditions. Previous studies have shown the potential of utilising a novel software algorithm to enable the exploration of consultation notes in EMR systems used commonly for OECD primary care records.[18] To date this software has shown the ability to analyse service utilisation for H1N1 influenza and childhood respiratory diseases while eliminating reliance on clinical coding.[19][20]

Methods¶

A natural language processing (NLP) software inference algorithm was developed to interrogate quantitative and qualitative cross-sectional, and retrospective cohort data from EMRs.

Setting and participants

The development of the algorithm methodology has been earlier described.[18] New Zealand has universal enrolment of the population with a primary care (general) practice, and universal computerised recording of general practice doctor (GP) consultations. The study was conducted

in the lower North Island in a mixed urban and rural setting. It consisted of 39 consenting general practices from two primary health organisations (PHOs) giving a total unique enrolled population of 391,000 over the study period between 1 January 2005 to 31 December 2015. This included individuals who both joined and left the cohort during this period. The cohort totalled 2,366,870.5 person-years and contained over 22 million medical record transactions representing 6,189,019 doctor consultations.

Figure 1 shows the selection process, and steps taken through the development and analysis of the data set. Data was extracted directly from the EMR using software that automates the collection and secure transmission of large datasets. The complete dataset was filtered to identify doctor consultations generated during standard office hours. Practice nurse and out-of-hours consultations were excluded. Each consultation record was linked to the individual's unique National Health Index number. This individual unique identifier is assigned to every person who uses health services in New Zealand to enable records to be matched between datasets. All data were analysed on the premises of the PHO utilising rigorous protocols to ensure patient confidentiality, and no member of the research team accessed identifiable data. Rates were age-standardised, where appropriate, using the direct method and exact confidence intervals (CI) were calculated.[21]

Process

The software algorithm was designed to replicate the diagnosis and assessment made by clinical experts.[18] It was trained, validated and tested using three independent data sets of 800 doctor consultation records. Each data set was stratified to contain 600 randomly chosen records and a further 200 random records from all that contained a simple keyword related to zoster. Clinical records from any single practice or provider could only exist in a single data set so as to avoid any training bias during validation or test procedures. Each record in all data sets was independently classified by two general practice clinical experts (LM and AD). At the completion of coding all records, the clinical experts reviewed, discussed and reached consensus on any records where they showed a discrepancy in coding. The focus of algorithm development was to maximise specificity and positive predictive values to reduce false positives because of the expected relative rarity of zoster consultations.

A zoster index consultation was defined as the initial zoster consultation occurring for an individual within the past 12 months. A follow-up visit was any visit identified as related to zoster occurring within 12 months of any prior zoster consultation.

Analysis

The demographic characteristics of the study cohort was measured by age, gender, patient-identified ethnicity, and socioeconomic deprivation quintile via the New Zealand Deprivation Index (a small census-area measure of socioeconomic deprivation).[22,23] These were compared with those of all patients enrolled with the PHO (N=3,014,598 person-years) and the 2013 New Zealand Census data (N=4,353,198 people). Patients were observed for the period they were enrolled in a participating practice over the study period. In order to maintain a

consistency of analysis across the study period, both deprivation and ethnicity were determined from the most recently recorded information available for each patient. Where appropriate, consultation rates were adjusted for algorithm sensitivity, specificity and age-adjusted for the 2013 New Zealand Census population using direct-standardisation. All data aggregation, transformation, cleaning and storage were done in Microsoft SQL Server, and statistical analysis was undertaken in R. Confidence intervals for crude rates were calculated using a bootstrap over 1000 replicate rounds with resampling. Confidence intervals on age-standardised rates were made using the method described by Fay and Feuer [21]. Hypothesis tests were conducted using the Fischer exact test for cohort analysis and the Wilcoxon signed rank test for self-controlled case series.

Results

Study cohort demographics

The age and gender of the study cohort approximated the national census profile. There was a larger proportion of 18-23 year olds in the local enrolled population than in the study cohort, which is likely to be due to self-exclusion of practices providing student health services to the universities within the catchment area. While the demographic characteristics of the study cohort closely matched those of the enrolled population for deprivation, due to the geocoding system used for deprivation, 6% of both the enrolled and study cohort populations with incorrectly documented address information were not allocated deprivation scores. The study cohort had a higher proportion of people in the least deprived socioeconomic quintile with a correspondingly lower proportion in the two most deprived. As compared with the enrolled population, the study cohort had fewer with Pacific Island ethnicity (3% versus 5%) and more 'Other' ethnicities, which included New Zealand European and Asian (88% versus 86%).

Accuracy of herpes zoster identification

The natural language algorithm had a positive-predictive value (PPV) of 0.82 (95% CI 0.72 to 0.92), specificity of 0.9998 (95% CI 0.9997 to 0.9999) and sensitivity of 0.84 (95% CI 0.74 to 0.92). This was more accurate than using keywords only (PPV 0.66, specificity 0.9994, and sensitivity 1.0), or using a single clinical expert (PPV 0.53, specificity 0.9991, and sensitivity 0.93).

Herpes zoster consultations

The overall age-adjusted apparent rate of zoster index consultations was 42.7 per 10,000 person-years observed (95% CI 41.9 to 43.5), with an estimated true rate of 48.6 (95% CI 47.6 to 49.6). There were 10,316 index consultations for zoster and 3,060 zoster-related follow-up consultations. The apparent rate for zoster index consultations was 16.7 per 10,000 doctor consultations (95% CI 16.3 to 17.0) with an estimated true rate of 17.5 (95% CI 17.1 to 17.9). This was the equivalent to one in 571 doctor consultations.

The rate of consultations were much higher in older age groups, as shown in Figure 2, with the highest rate in the 80 – 90 year age group at 128 consultations per 10,000 person-years.

There was a significant increase in rate of zoster consultations over time (odds ratio 0.86 [95% CI 0.78 to 0.94]; p = 0.0015). The rate of increase over time was particularly noticeable in the older age groups. This trend is shown in Figure 3.

Gender

When comparing age-standardised data by gender, females experienced a statistically significant higher rate of zoster over the study period compared to males (F=48.3, M=36.6, p< 0.001) as shown in Figure 4.

While the age-standardised rate of zoster increased over time generally, there was no statistically significant difference in the rate of increase between males or females.

Ethnicity and socioeconomic status

There were fewer index zoster consultations of those with Pacific Island ethnicity (age-adjusted rate 29.1 per 10,000 patient-years [95% CI 25.6 to 33.1]; p < 0.01) and Māori (New Zealand indigenous; rate 38.9 [36.3 to 41.6]; p = 0.019) than those from other ethnicities (rate 42.3 [41.4 to 43.2]). There was no significant difference in the rate of zoster index consultations across the different socioeconomic quintiles.

Health service utilisation

Herpes zoster-related consultations

Of the 10,316 zoster index consultations identified by the cohort analysis, most had no zoster follow-up consultations. Of these, 19% had a zoster follow-up consultation and only 5.8% consulted their doctor more than once in relation to zoster (Table 1).

Table 1: Distribution of follow-up consultations following zoster index cases

Follow-ups	0	1	2	3	4	5	6	7	8	9	10	11	12
Total number	8354	1365	372	100	73	16	14	8	2	6	2	3	1
Percentage	80.98	13.23	3.61	0.97	0.71	0.16	0.14	0.08	0.02	0.06	0.02	0.03	0.01

With increasing age, particularly from 45 years onwards, there was an increasing likelihood of follow-up consultations per episode as shown in Figure 5.

Distribution of non-herpes zoster consultations around index cases

To assess any correlation between zoster consultations and any changes in overall (non-zoster) consultations to general practice, we undertook a self-controlled case series analysis, consisting of 6,823 of the zoster index cases, and measured the variance in the number of all non-zoster consultations for 12 months prior to and 12 months after each zoster-index consultation occurred (a total of 27 months observation for each index consultation). This is expressed as the variance between the number of non-zoster consultations before and after the index consultation as a proportion of their sum. Positive proportions represented those with more consultations after the index consultation, while negative proportions represented those with more consultations before. Those with no consultations before or after were considered to have a proportion of zero as in Figure 6. There were no statistically significant differences in the distributions at a threshold of 0.001.

Data sharing

No additional data are available.

Discussion

Consistent with previously reported rates of incidence of zoster in primary care presentations, across many countries [5–8,10], this study found that the overall rate of incidence of zoster is 48.6 cases per 10,000 patient-years, with higher rates in the elderly [7] and a 32% higher rate in females than males [9]. The peak age of incidence seen in this study was in the 80 to 84 years age group, which is older than previously reported. An Australian study using Medicare Benefits Schedule items reported a peak age of 60-69 years.[24] The importance of showing a peak at an older age is that it will affect modelling for decision-making around the ideal age for zoster vaccine introduction.

Similar to other international findings, the overall incidence of zoster has increased from 2005 to 2016 across all age groups [6]. New Zealand did not have a childhood varicella vaccination programme over this period, supporting previous commentary that this increase is unlikely to be related to a decline in circulating varicella virus.[12]

Previous literature has reported differences between ethnic groups, most notably, a reduced self-reported occurrence was seen in US Blacks [25]. Our New Zealand -based study reported a lower age-adjusted incidence for those of Pacific Island ethnicity at 29.1 per 10,000 patient-years (95% CI 25.6 to 33.1) and a rate of 38.9 per 10,000 patient-years (95% CI 36.3 to 41.6]) for New Zealand indigenous Māori when compared to other ethnicities. This is in contrast to almost all other important health statistics for older Pacific Island and Māori populations in New Zealand, for which there is a consistent equity gap, and poorer age-adjusted health outcomes are associated with these groups.[26] There does not appear to be any socioeconomic link as there was no significant difference between different levels of socioeconomic deprivation. A different burden of childhood varicella seems unlikely, because Māori are not a migrant group. This raises the question as to whether there are significant differences in rates across other ethnic groups internationally.

An important original finding from this study was the lack of evidence for increased burden of utilisation of health services at the primary care level. Utilising a large primary care data-set over an 11 year period, we have demonstrated an equivalent of one zoster-related consultation in every 571 general practice consultations. Furthermore, the burden of subsequent consulting was very low with 80% of zoster-related presentations requiring no follow-up and 13% requiring only a single follow-up consultation. While there was an increasing likelihood of follow-up consultations following a zoster episode with increasing age, particularly from 45 years onwards, the burden on general practice consultation rates was not significantly affected, overall. An episode of zoster is reported to frequently reduce overall quality of health particularly in older age groups, most likely related to the prolonged effects of post herpetic neuralgia.[27,28] International literature reports hospitalisation in adults aged over 50 years at a yearly rate of 28/100,000 zoster related hospitalisations as primary diagnosis (ranging from 6.1 in the 50-54 year age group to 95.8/100,000 persons in the over 80 years age group).[29] In New Zealand, during 2015 there were 361 hospitalisations with a primary diagnosis of zoster across all ages (Ministry of Health data). The burden of zoster admissions to hospital is severe when length of stay, cost and mortality are considered.[30] However, our study has demonstrated that these complications are not translating through to increasing utilisation of general practice services. This indicates that overall burden on the primary care services is less than has previously been suggested, and that zoster contributes very little to the overall utilisation burden in general practice.

Study Limitations

The gold standard for this study was based on doctor decision making, and the algorithm is limited by the quantity and detail of the recorded information in each consultation. In particular, repeat consultations may underreport ongoing zoster-related symptoms when the primary reason for a visit is in relation to other comorbidities. However this may not be a major concern as, for medico-legal reasons, significant clinical observations about ongoing zoster complications or progress are likely to be recorded. While recognising this, the overall incidence found in this study matches other international data, suggesting good concordance.

The number of less deprived individuals in the population studied was higher than that of the regional and national populations, but was otherwise well matched. Due to the self-exclusion of student health centres, the rate of zoster incidence in younger age-groups may be underreported, although, it is unlikely to affect overall incidence rates since cases are predominantly in those over 50 year olds. This study did not include out-of-hours presentations and patients not enrolled with a general practice. However, the main purpose of this study was to report on service utilization burden in routine general practice, not periodic acute presentations in after-hours clinics. New Zealand has very high registration in general practice, particularly of the elderly population.[31]

Study Strengths

In this study, a very large data set of doctor consultations were examined by the way of a software inference algorithm. This methodology interrogated the free-text electronic medical

records of over 6 million unique doctor consultations over an 11 year period. The large data set enabled analysis of rates of zoster incidence by age bands and different demographic measures, across the whole time period, despite the low frequency of zoster cases.

The strength of the NLP algorithm-based method used in our study is improved accuracy, when compared with the use of clinical codes or a simple keyword search, or review by a single clinical expert.[19,32] This methodology, as opposed to a simple keyword search, is able to identify the context in which pertinent terms are being used in clinical narrative. For example, a keyword may be used by a clinician to express either the presence or absence of the disease, which impacts the specificity and positive predictive value of that approach and ultimately over-estimating disease. A single clinical expert is prone to make errors, and has previously been shown to perform worse than a simple keyword search or NLP algorithm. In our study, two independent clinical coders reached concordance to provide a robust gold-standard comparison.

Using this NLP algorithm methodology, we have demonstrated its ability to review the burden of low frequency conditions, such as zoster, in primary care, and follow changes in patterns over times with large numbers.

Unanswered questions and future research

This study only focused on burden at the primary care level. Future research is needed on the comparative burden of disease across the full spectrum of health services: community, primary care and hospital inpatient care. Further important questions include disaggregation of the burden of disease by comorbidities, the effect of the use of antivirals and other treatment modalities and why the rate of zoster is continuing to increase over time. In addition, studies internationally have shown significant burden of disease on the quality of life of individuals, particularly those with severe or prolonged disease complications.[28] A qualitative study could provide insight into the effect shingles has on individuals and the potential value of a vaccine in reducing this burden. The low level of utilisation of general practice services suggests, however, that the burden of more severe disease falls on a small number of individuals and our results may prompt further discussion around modelling for whom to introduce zoster vaccines to in a population.

Conclusions

Overall, the rate of general practice doctor consultations related to zoster showed that this condition is rare in primary care and, while repeat visits following a zoster episode became increasingly common with age, the disease does not represent a significant burden on overall general practice workload. The peak age for consultations was older than has previously been reported, and there appear to be significant differences between ethnic groups, unrelated to socioeconomic circumstances. These are important findings particularly when considering the introduction of zoster vaccines across national immunisation schedules.

Use of a novel software algorithm to enable the exploration of consultation notes in EMRs is an efficient and effective way of identifying conditions, patterns, changes over time and the burden

of disease on primary care services. We have also shown the methodology has the ability to review the burden of low frequency conditions in primary care.

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

Ethics approval: This study was approved by the University of Auckland Human Participants Ethics Committee Ref. 017617 on 25 Jul 2016¶

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

Contributors: NT, MN, AD conceived the study. All authors contributed to the development of the overall study methodology. MS and MN managed the ethical approval and consent processes. AD, LMB and NT provided clinical input into the algorithm design. JMR designed and built the natural language processing tools, programmed and trained the algorithm, and conducted the data analyses. AD and LM classified the consultation records in the gold standard sets. NT and MN 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 in the study and can take responsibility for the integrity of the data and accuracy of the analysis.

Competing Interests. All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare no other support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years' no other relationships or activities that could appear to have influenced the submitted work.

Transparency declaration The lead author (NT) affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

References¶

1 Eshleman E, Shahzad A, Cohrs RJ. Varicella zoster virus latency. *Future Virol* 2011;**6**:341–55.

2 Brisson M, Edmunds WJ, Law B, *et al.* Epidemiology of varicella zoster virus infection in Canada and the United Kingdom. *Epidemiol Infect* 2001;**127**:305–14.

3 Wehrhahn MC, Dwyer DE. Herpes zoster: epidemiology, clinical features, treatment and prevention. *Australian Prescriber* 2012;**35**:143–7.

4 Schmader K, Gnann JW, Jr., *et al.* The Epidemiological, Clinical, and Pathological Rationale for the Herpes Zoster Vaccine. *J Infect Dis* 2008;**197**:S207–15.

5 Gauthier A, Breuer J, Carrington D, *et al.* Epidemiology and cost of herpes zoster and post-herpetic neuralgia in the United Kingdom. *Epidemiol Infect* 2009;**137**:38–47.

6 Yawn BP, Gilden D. The global epidemiology of herpes zoster. *Neurology* 2013;**81**:928–30.

7 Kawai K, Gebremeskel BG, Acosta CJ. Systematic review of incidence and complications of herpes zoster: towards a global perspective. *BMJ Open* 2014;**4**:e004833.

8 Pinchinat S, Cebrián-Cuenca AM, Bricout H, *et al.* Similar herpes zoster incidence across Europe: results from a systematic literature review. *BMC Infect Dis* 2013;**13**:170.

9 Yawn BP, Saddier P, Wollan PC, Sauver JLS, Kurland MJ, Sy LS. A population-based study of the incidence and complication rates of herpes zoster before zoster vaccine introduction. In: *Mayo Clinic Proceedings*. Mayo Clinic 2007. 1341–9.

10 Reid JS, Ah Wong B. Herpes zoster (shingles) at a large New Zealand general practice: incidence over 5 years. *N Z Med J* 2014;**127**:56–60.

11 Johnson RW. Herpes zoster and postherpetic neuralgia. *Expert Rev Vaccines* 2010;**9**:21–6.

12 Hales CM, Harpaz R, Joesoef MR, *et al.* Examination of links between herpes zoster incidence and childhood varicella vaccination. *Ann Intern Med* 2013;**159**:739–45.

13 Jumaan AO, Yu O, Jackson LA, *et al.* Incidence of Herpes Zoster, Before and After Varicella-Vaccination–Associated Decreases in the Incidence of Varicella, 1992–2002. *J Infect Dis* 2005;**191**:2002–7.

14 Leung J, Harpaz R, Molinari N-A, *et al.* Herpes zoster incidence among insured persons in the United States, 1993-2006: evaluation of impact of varicella vaccination. *Clin Infect Dis* 2011;**52**:332–40.

15 Toyama N, Shiraki K, Society of the Miyazaki Prefecture Dermatologists. Epidemiology of herpes zoster and its relationship to varicella in Japan: A 10-year survey of 48,388 herpes zoster cases in Miyazaki prefecture. *J Med Virol* 2009;**81**:2053–8.

16 Russell ML, Schopflocher DP, Svenson L, *et al.* Secular trends in the epidemiology of shingles in Alberta. *Epidemiol Infect* 2007;**135**:908–13.

17 Varicella and herpes zoster vaccines: WHO position paper, June 2014. *Wkly Epidemiol Rec* 2014;**89**:265–87.

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**:78.

20 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.

21 Fay MP, Feuer EJ. Confidence intervals for directly standardized rates: a method based on the gamma distribution. *Stat Med* 1997;**16**:791–801.

22 Houghton F. Degrees of Deprivation in New Zealand — An Atlas of Socioeconomic Difference: By P. Crampton, C. Salmond, R. Kirkpatrick, R. Scarborough and C. Skelly. *N Z Geog* 2001;**57**:62–62.

23 Salmond C, Crampton P, Sutton F. NZDep91: A New Zealand index of deprivation. *Aust N Z J Public Health* 1998;**22**:835–7.

24 Karki S, Newall AT, Raina MacIntyre C, *et al.* Healthcare Resource Utilisation Associated with Herpes Zoster in a Prospective Cohort of Older Australian Adults. *PLoS One* 2016;**11**:e0160446.

25 Schmader K, George LK, Burchett BM, *et al.* Racial differences in the occurrence of herpes zoster. *J Infect Dis* 1995;**171**:701–4.

26 Anderson I, Crengle S, Kamaka ML, *et al.* Indigenous health in Australia, New Zealand, and the Pacific. *Lancet* 2006;**367**:1775–85.

27 Serpell M, Gater A, Carroll S, *et al.* Burden of post-herpetic neuralgia in a sample of UK residents aged 50 years or older: findings from the Zoster Quality of Life (ZQOL) study. *Health Qual Life Outcomes* 2014;**12**:92.

28 Gater A, Abetz-Webb L, Carroll S, *et al.* Burden of herpes zoster in the UK: findings from the zoster quality of life (ZQOL) study. *BMC Infect Dis* 2014;**14**:402.

29 Stein AN, Britt H, Harrison C, *et al.* Herpes zoster burden of illness and health care resource utilisation in the Australian population aged 50 years and older. *Vaccine* 2009;**27**:520–9.

30 Hobbelen PHF, Stowe J, Amirthalingam G, *et al.* The burden of hospitalisation for varicella and herpes zoster in England from 2004 to 2013. *J Infect* 2016;**73**:241–53.

31 Enrolment in a primary health organisation. Ministry of Health NZ. http://www.health.govt.nz/our-work/primary-health-care/about-primary-health-organisations/enrolment-primary-health-organisation (accessed 27 Oct 2017).

32 Wu M-F, Yang Y-W, Lin W-Y, *et al.* Varicella zoster virus infection among healthcare workers in Taiwan: seroprevalence and predictive value of history of varicella infection. *J Hosp Infect* 2012;**80**:162–7.

Figure legends

Figure 1: Recruitment Process

Figure 2: Herpes zoster index consultation rate by age group (bars represent 95% CIs)

Figure 3: Herpes zoster index consultation rate over time from age of 50 years, by five year age groups (n is mean patient years per year, with 95% confidence intervals and linear regression shown in blue).

Figure 4: Herpes zoster consultation rate of index case over time by gender, age adjusted (95% CIs)

Figure 5: Proportion of herpes zoster consultations with or without follow-up visits by age group showing 95% CIs.

Figure 6 Distribution of overall consultation visits occurring prior to and after a herpes zoster index case by age group

Table 1: Distribution of follow-up consultations following zoster index cases

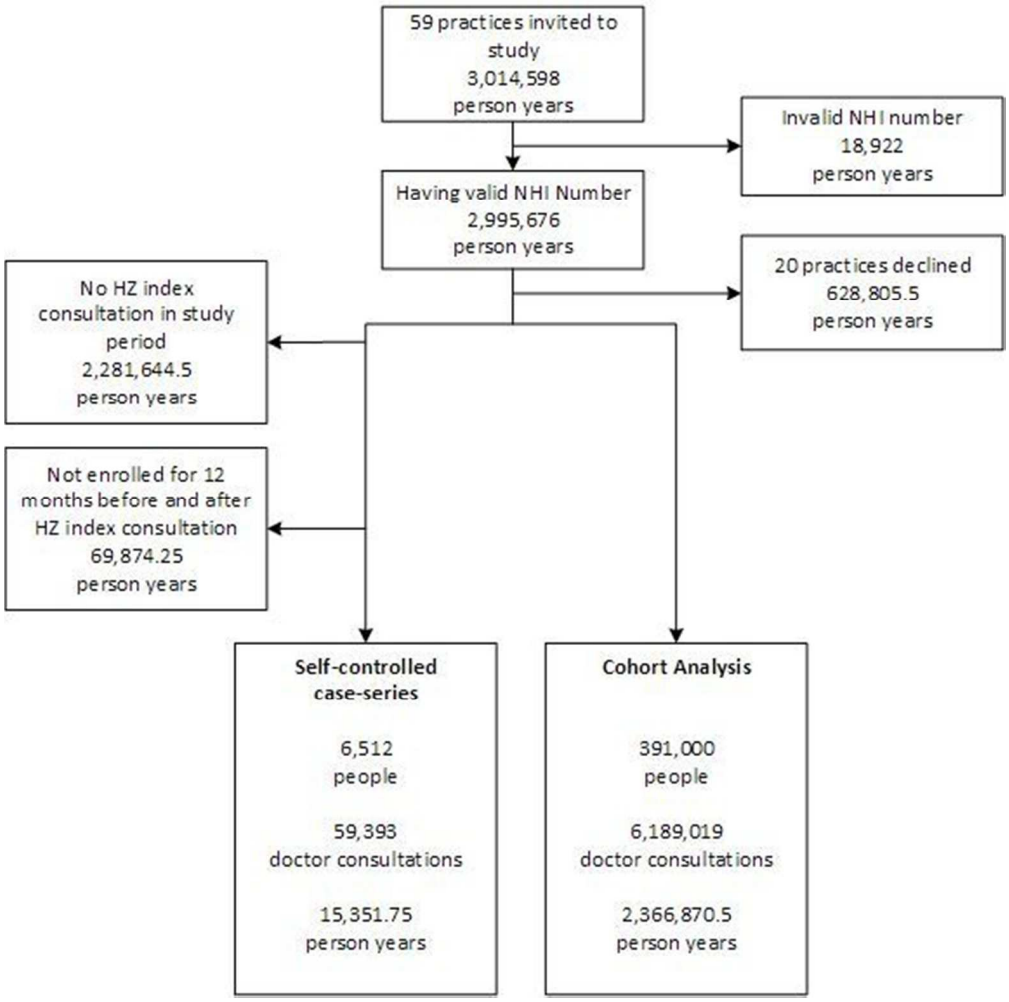

Figure 1: Recruitment Process

48x48mm (300 x 300 DPI)

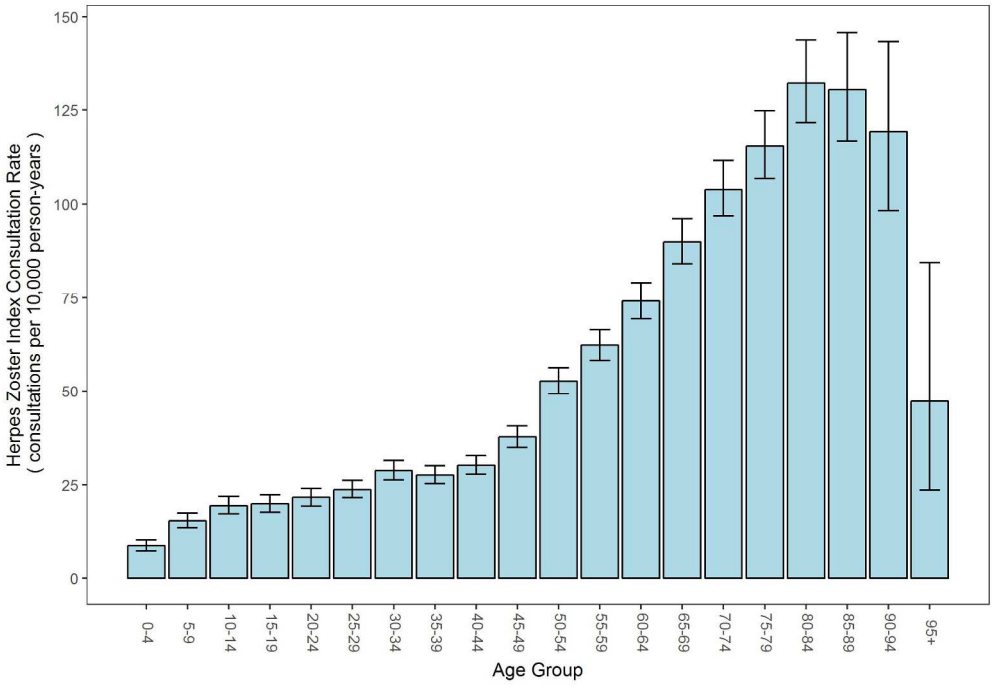

Figure 2: Herpes zoster index consultation rate by age group (bars represent 95% CIs)

210x145mm (300 x 300 DPI)

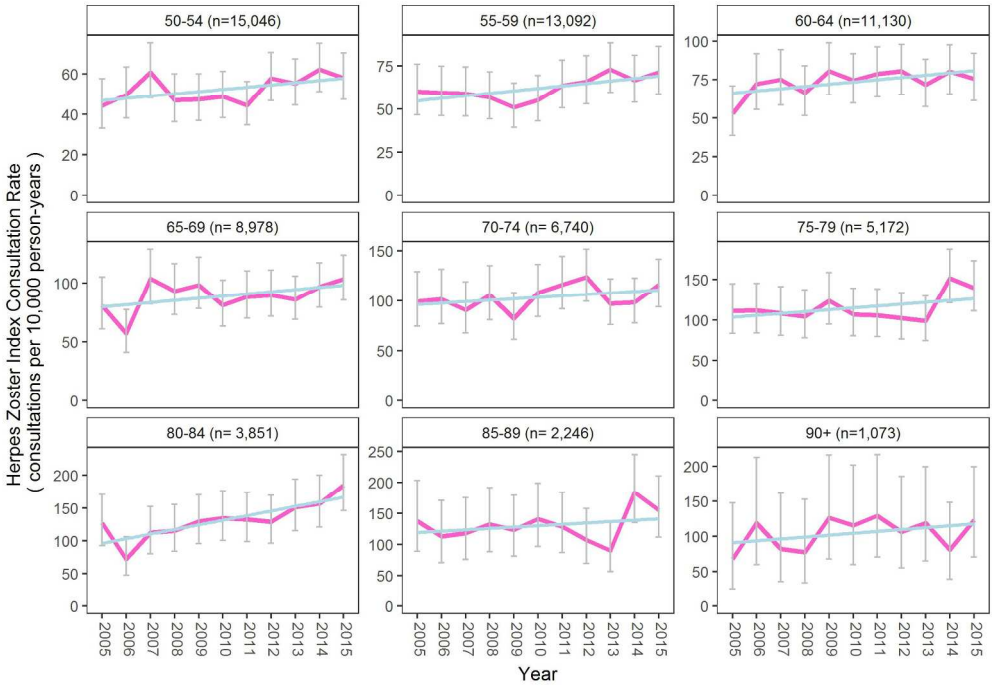

Figure 3: Herpes zoster index consultation rate over time from age of 50 years, by five year age groups (n is mean patient years per year, with 95% confidence intervals and linear regression shown in blue).

210x145mm (300 x 300 DPI)

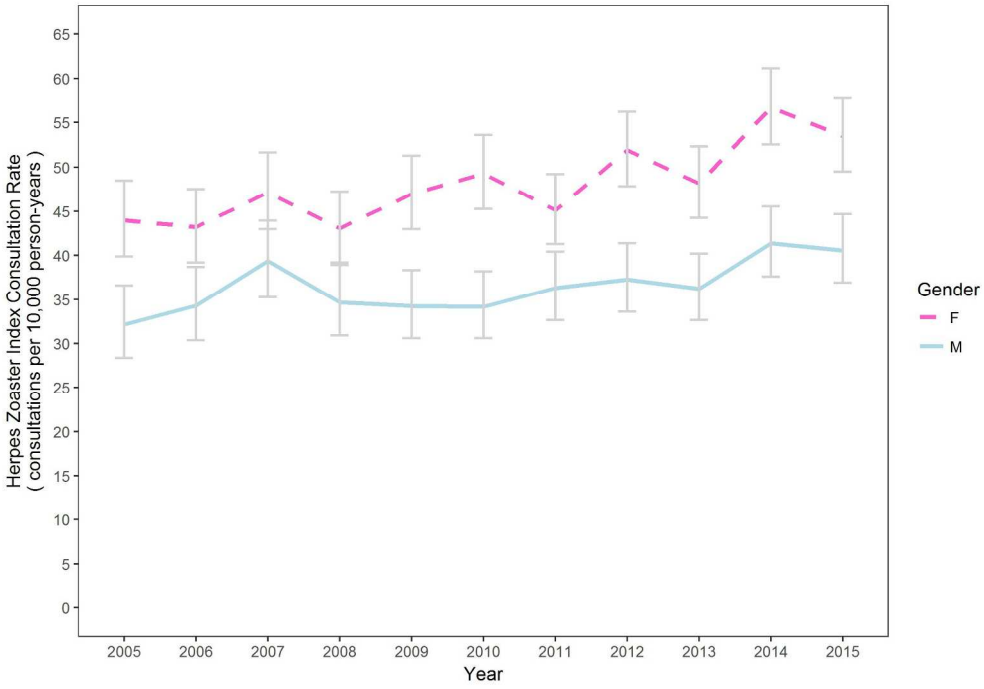

Figure 4: Herpes zoster consultation rate of index case over time by gender, age adjusted (95% CIs)

210x145mm (300 x 300 DPI)

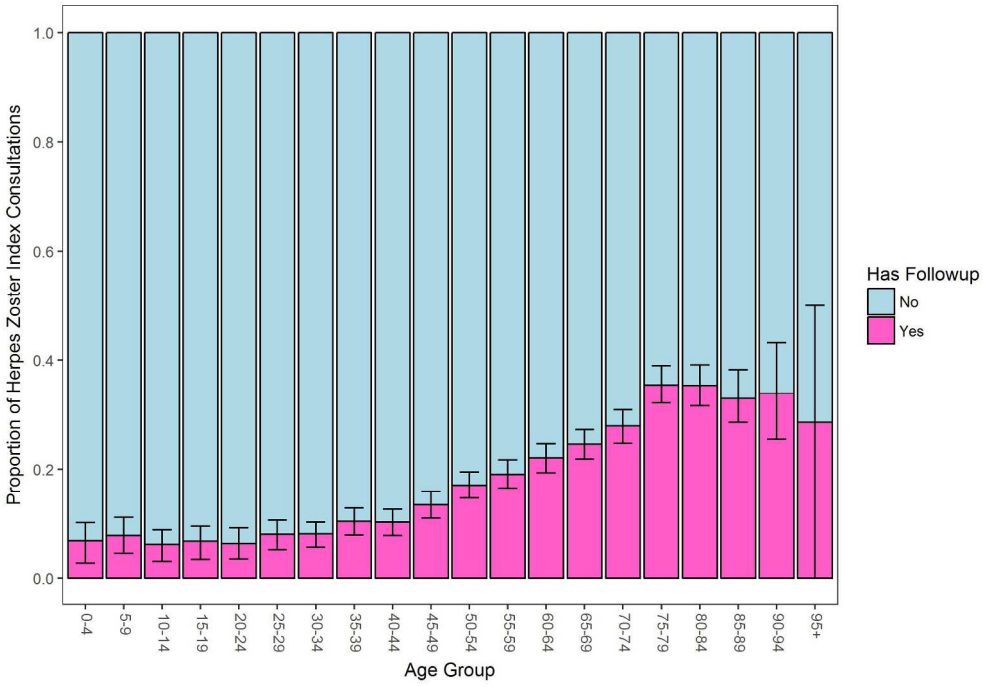

Figure 5: Proportion of herpes zoster consultations with or without follow-up visits by age group showing 95% CIs.

210x145mm (300 x 300 DPI)

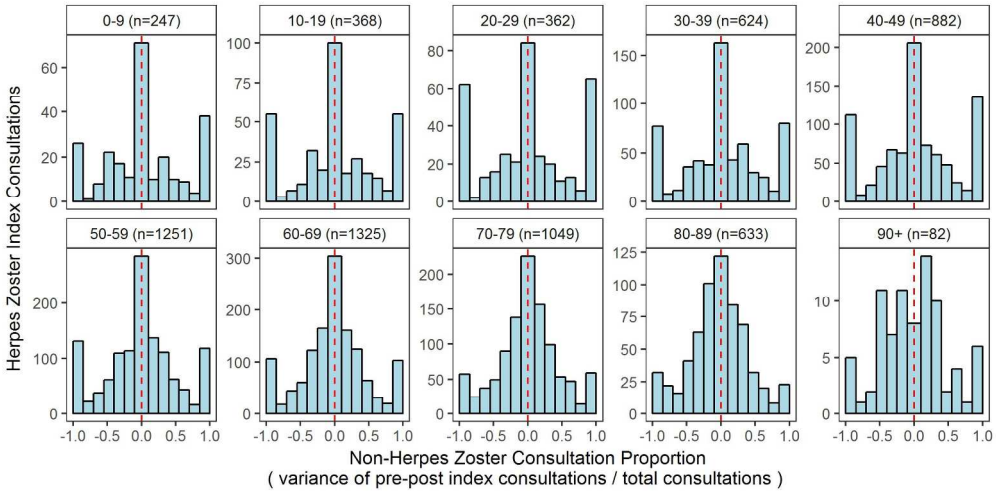

Figure 6 Distribution of overall consultation visits occurring prior to and after a herpes zoster index case by age group

211x105mm (300 x 300 DPI)

ORD statement – checklist of items, extended from the STROBE statement, that should be reported in observational stud collected health data.

	Item No.	STROBE items	Location in manuscript where items are reported	RECORD items	Location manuscri where ite reported
bstract					
	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found	Title - line 6 -7 Abstract line 10	RECORD 1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included. RECORD 1.2: If applicable, the geographic region and timeframe within which the study took place should be reported in the title or abstract. RECORD 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.	Title - lin Abstract - Title: line Abstract - 14, line 1 Abstract - 14
on					
d	2	Explain the scientific background and rationale for the investigation being reported	page 4 paragraph 1 and 2		
	3	State specific objectives, including any prespecified hypotheses	page 4 paragraph 3 - line 36		
gn	4	Present key elements of study design early in the paper	page 4 paragraph 1 line 47-48		
	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	page 5 line 53-		
s	6	*(a) Cohort study* - Give the	page 5 line 11	RECORD 6.1: The methods of study	Flow char

		eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study* - Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study* - Give the eligibility criteria, and the sources and methods of selection of participants *(b) Cohort study* - For matched studies, give matching criteria and number of exposed and unexposed *Case-control study* - For matched studies, give matching criteria and the number of controls per case		population selection (such as codes or algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided. RECORD 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided. RECORD 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data linkage process, including the number of individuals with linked data at each stage.	figure 1 Page 5 lin Reference Page 5 lin Figure 1 - chart Page 5, li line 49
	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.		RECORD 7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided.	Process, p line 42 an reference
es/ nt	8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	flow chart Figure 1 analysis page 4 line 12, page 5 lines 50 and 54 Not applicable		
	9	Describe any efforts to address potential sources of bias	page 5 line 34		

	10	Explain how the study size was arrived at	settings and participants from page 4 line 56 to page 5 line 9 Figure 1 - flow chart		
e	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why			Analysis line 49 – line 7
	12	(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) *Cohort study* - If applicable, explain how loss to follow-up was addressed *Case-control study* - If applicable, explain how matching of cases and controls was addressed *Cross-sectional study* - If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses	Page 6 line 6 Page 5 line 22 page 6 paragraph 1 page 5 line 38 page 6 line 38		
s and ethods		..	analysis page 6 line 8	RECORD 12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population.	Settings a participan page 5, li

				RECORD 12.2: Authors should provide information on the data cleaning methods used in the study.	analysis P line 8
		..		RECORD 12.3: State whether the study included person-level, institutional-level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.	Settings a participan 5 line 18
s	13	(a) Report the numbers of individuals at each stage of the study (*e.g.*, numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed) (b) Give reasons for non-participation at each stage. (c) Consider use of a flow diagram		RECORD 13.1: Describe in detail the selection of the persons included in the study (*i.e.,* study population selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in the text and/or by means of the study flow diagram.	figure 1 - chart
e data	14	(a) Give characteristics of study participants (*e.g.*, demographic, clinical, social) and information on exposures and potential confounders (b) Indicate the number of participants with missing data for each variable of interest (c) *Cohort study* - summarise follow-up time (*e.g.*, average and total amount)			Study coh demograp page 6 lin Accuracy identifica page 6 lin
ata	15	*Cohort study* - Report numbers of outcome events or summary measures over time *Case-control study* - Report			herpes zo consultati page 6 lin

		numbers in each exposure category, or summary measures of exposure *Cross-sectional study* - Report numbers of outcome events or summary measures			healthcare utilisation from line Table 1
ts	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period			Confidenc intervals f statistical analyses a presented figures. Methods, page 6 lin
yses	17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses			Accuracy herpes zo identifica page 6 lin
	18	Summarise key results with reference to study objectives			paragraph
s	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias		RECORD 19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.	study limi page 8 lin page 8 lin
on	20	Give a cautious overall interpretation of results			page 9 lin

		considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence			
bility	21	Discuss the generalisability (external validity) of the study results			conclusio 10 line 47 line 53
rmation					
	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based			funding, p line 21
ty of aw ng		..		RECORD 22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code.	not applic

: Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM, the RECORD
. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Medi*

BMJ Open

Quantifying the incidence and burden of herpes zoster in New Zealand general practice: a retrospective cohort study utilising a natural language processing software inference algorithm.

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-021241.R1
Article Type:	Research
Date Submitted by the Author:	06-Mar-2018
Complete List of Authors:	Turner, Nikki; University of Auckland, General Practice and Primary Health Care MacRae, Jayden; Datacraft Analytics Nowlan, Mary; University of Auckland Tamaki Campus, General Practice and Primary Health Care McBain, Lynn; University of Otago, Wellington, Stubbe, Maria; University of Otago, Primary Health Care and General Practice Dowell, Anthony; Wellington School of Medicine and Health Sciences,
Primary Subject Heading:	General practice / Family practice
Secondary Subject Heading:	General practice / Family practice, Health informatics, Health services research, Infectious diseases, Research methods
Keywords:	herpes zoster, PRIMARY CARE, VIROLOGY, Health informatics < BIOTECHNOLOGY & BIOINFORMATICS, Infectious diseases & infestations < DERMATOLOGY

Quantifying the incidence and burden of herpes zoster in New Zealand general practice: a retrospective cohort study utilising a natural language processing software inference algorithm.

N M Turner, Associate Professor[1],

J MacRae, Chief Scientist[2]

M L Nowlan, Project Manager[1],

L McBain, Senior Lecturer[3],

M Stubbe, Senior Lecturer[3]

A Dowell, Professor[3]

1. Department of General Practice and Primary Care, University of Auckland, Private Bag 92019, Victoria Street West, Auckland 1142, New Zealand .
2. DataCraft Analytics Limited, PO Box 17271, Karori, Wellington 6147
3. Department of Primary Care and General Practice, University of Otago Wellington, PO Box 7343 Newtown Wellington 6242 New Zealand

Correspondence to Dr. N Turner; n.turner@auckland.ac.nz

Abstract

Objective: To investigate the incidence of primary care presentations for herpes zoster (zoster) in a representative New Zealand population and to evaluate the utilisation of primary health care services following zoster diagnosis.

Design: A cross-sectional retrospective cohort study used a natural language processing software inference algorithm to identify general practice consultations for zoster by interrogating 22 million electronic medical record (EMR) transactions routinely recorded from January 2005 to December 2015. Data-linking enabled analysis of the demographics of each case. The frequency of doctor visits were assessed prior to and after the first consultation diagnosing zoster to determine health service utilisation.

Setting: General practice, using EMRs from two primary health organisations located in the lower North Island, New Zealand.

Participants: Thirty-nine general practices consented interrogation of their EMRs to access de-identified records for all enrolled patients. Out-of-hours and practice nurse consultations were excluded.

Main outcome measures: The incidence of first and repeated zoster-related visits to the doctor across all age groups and associated patient demographics. To determine whether zoster affects workload in general practice.

Results: Overall, for 6,189,019 doctor consultations, the incidence of zoster was 48.6 per 10,000 patient years (95% CI 47.6 to 49.6). Incidence increased from the age of 50 years to a peak rate of 128 per 10,000 in the 80-90 years age group and was significantly higher in females than males ($p < 0.001$). Over this 11 year period, incidence increased gradually, notably in those aged 80-85 years. Only 19% of patients had one or more follow-up zoster consultations within 12 months of a zoster index consultation. The frequency of consultations, for any reason, did not change between periods before and after the diagnosis.

Conclusions: Zoster consultations in general practice are rare and the burden of these cases on overall general practice caseload is low.

Article summary

Strengths and limitations of the study

- This study used a novel and validated natural language processing software inference algorithm to identify herpes zoster presentation rates and service utilisation using primary care electronic medical records over an eleven year period.

- Despite a low frequency of zoster cases, the large data set enabled analysis of rates of zoster incidence by age bands and different demographics across the whole time period.

- The algorithm was designed to maximise specificity and accuracy, thereby generating a conservative estimate of the burden of zoster presentations in primary care by keeping false positives to a minimum.

- The gold standard for this study was based on doctor decision making, and the algorithm is limited by the quantity and detail of the recorded information in each consultation. Details of the reason for each visit were not determined, just that it was identified as zoster related.

- 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 rates.

Introduction¶

Infection with varicella-zoster virus (VZV) establishes life-long persistence in sensory nerve ganglia. When the immune system is unable to maintain the suppression of the virus, it manifests as a clinical syndrome known as herpes zoster (zoster; shingles).[1] About one third of the population experiences zoster, with greater incidence in older age groups.[2] The major risk factor for zoster is a decline in VZV immunity as cell-mediated immunity wanes with age.[3,4] Cohort studies, from at least 22 countries, give incidence of zoster in the general population from 3 to 5 per 1000 person-years [5–10], and a lifetime incidence of 20 - 30%.[2] Much higher incidence is reported in older adults, increasing from 50 – 60 years of age, and ranging from 8 – 12 per 1000 person-years.[2,7] Around half of the zoster cases over 70 years of age will develop postherpetic neuralgia (PHN).[11] The main rationale for the use of a zoster vaccine is to prevent the long-lasting effects of PHN. A live attenuated zoster vaccine is internationally available for adults over the age of 50 years, and is on the national schedule in some countries, notably, the United Kingdom and Australia at the age of 70 years, and recommended in the USA from age 60 years.[1]

Several studies have reported increasing incidence rates over time across all age groups [6], in countries both with [12–14] and without childhood varicella immunisation programmes.[2,15,16] The World Health Organization recommends, when considering the optimal age and dosing schedule for zoster vaccines, that countries should take into consideration the age-dependent burden of disease, vaccine effectiveness, duration of protection and cost effectiveness of the vaccine.[17] While there is a body of research around zoster and the burden of zoster disease, this is almost exclusively based on observational studies from administrative databases or health records, which will underestimate cases not coming to medical attention.[6] Decisions around vaccination strategies for countries remain hindered by a more complete understanding of the burden of the disease for a population.

The objective of this study was to interrogate data from general practice electronic medical records (EMR) to identify zoster-related primary care presentations and service utilisation associated with consultations for zoster-related conditions. Previous studies have shown the potential of utilising a novel software algorithm to enable the exploration of consultation notes in EMR systems used commonly for OECD primary care records.[18] To date this software has shown the ability to analyse service utilisation for H1N1 influenza and childhood respiratory diseases while eliminating reliance on clinical coding.[19][20]

Methods¶

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

Setting and participants

The development of the algorithm methodology has been earlier described.[18] New Zealand has universal enrolment of the population with a primary care (general) practice, and universal computerised recording of general practice doctor (GP) consultations. The study was conducted

in the lower North Island in a mixed urban and rural setting. It consisted of 39 consenting general practices from two primary health organisations (PHOs) giving a total unique enrolled population of 391,000 over the study period between 1 January 2005 to 31 December 2015. This included individuals who both joined and left the cohort during this period. The cohort totalled 2,366,870.5 person-years and contained over 22 million medical record transactions representing 6,189,019 doctor consultations.

Figure 1 shows the selection process, and steps taken through the development and analysis of the data set. Data was extracted directly from the EMR using software that automates the collection and secure transmission of large datasets. The complete dataset was filtered to identify doctor consultations generated during standard office hours. Practice nurse and out-of-hours consultations were excluded. Each consultation record was linked to the individual's unique National Health Index number. This individual unique identifier is assigned to every person who uses health services in New Zealand to enable records to be matched between datasets. All data were analysed on the premises of the PHO utilising rigorous protocols to ensure patient confidentiality, and no member of the research team accessed identifiable data. Rates were age-standardised, where appropriate, using the direct method and exact confidence intervals (CI) were calculated.[21]

Patient and public involvement

This retrospective observational study on general practice doctor notes did not directly involve patients.

Process

The software algorithm was designed to replicate the diagnosis and assessment made by clinical experts.[18] It was trained, validated and tested using three independent data sets of 800 doctor consultation records. Each data set was stratified to contain 600 randomly chosen records and a further 200 random records from all that contained a simple keyword related to zoster. Clinical records from any single practice or provider could only exist in a single data set so as to avoid any training bias during validation or test procedures. Each record in all data sets was independently classified by two general practice clinical experts (LMB and AD). At the completion of coding all records, the clinical experts reviewed, discussed and reached consensus on any records where they showed a discrepancy in coding. The focus of algorithm development was to maximise specificity and positive predictive values to reduce false positives because of the expected relative rarity of zoster consultations.

A zoster index consultation was defined as the initial zoster consultation occurring for an individual within the past 12 months. A follow-up visit was any visit identified as related to zoster occurring within 12 months of any prior zoster consultation.

Analysis

The demographic characteristics of the study cohort was measured by age, gender, patient-identified ethnicity, and socioeconomic deprivation quintile via the New Zealand Deprivation Index (a small census-area measure of socioeconomic deprivation).[22,23] These were compared with those of all patients enrolled with the PHO (N=3,014,598 person-years) and the 2013 New Zealand Census data (N=4,353,198 people). Patients were observed for the period they were enrolled in a participating practice over the study period. In order to maintain a consistency of analysis across the study period, both deprivation and ethnicity were determined from the most recently recorded information available for each patient. Where appropriate, consultation rates were adjusted for algorithm sensitivity, specificity and age-adjusted for the 2013 New Zealand Census population using direct-standardisation. All data aggregation, transformation, cleaning and storage were done in Microsoft SQL Server, and statistical analysis was undertaken in R. Confidence intervals for crude rates were calculated using a bootstrap over 1000 replicate rounds with resampling. Confidence intervals on age-standardised rates were made using the method described by Fay and Feuer [21]. Hypothesis tests were conducted using the Fischer exact test for cohort analysis and the Wilcoxon signed rank test for self-controlled case series.

Results

Study cohort demographics

The age and gender of the study cohort approximated the national census profile. There was a larger proportion of 18-23 year olds in the local enrolled population than in the study cohort, which is likely to be due to self-exclusion of practices providing student health services to the universities within the catchment area. While the demographic characteristics of the study cohort closely matched those of the enrolled population for deprivation, due to the geocoding system used for deprivation, 6% of both the enrolled and study cohort populations with incorrectly documented address information were not allocated deprivation scores. The study cohort had a higher proportion of people in the least deprived socioeconomic quintile with a correspondingly lower proportion in the two most deprived. As compared with the enrolled population, the study cohort had fewer with Pacific Island ethnicity (3% versus 5%) and more 'Other' ethnicities, which included New Zealand European and Asian (88% versus 86%).

Accuracy of herpes zoster identification

The natural language algorithm had a positive-predictive value (PPV) of 0.82 (95% CI 0.72 to 0.92), specificity of 0.9998 (95% CI 0.9997 to 0.9999) and sensitivity of 0.84 (95% CI 0.74 to 0.92). This was more accurate than using keywords only (PPV 0.66, specificity 0.9994, and sensitivity 1.0), or using a single clinical expert (PPV 0.53, specificity 0.9991, and sensitivity 0.93).

Herpes zoster consultations

The overall age-adjusted apparent rate of zoster index consultations was 42.7 per 10,000 person-years observed (95% CI 41.9 to 43.5), with an estimated true rate of 48.6 (95% CI 47.6 to 49.6). There were 10,316 index consultations for zoster and 3,060 zoster-related follow-up consultations. The apparent rate for zoster index consultations was 16.7 per 10,000 doctor consultations (95% CI 16.3 to 17.0) with an estimated true rate of 17.5 (95% CI 17.1 to 17.9). This was the equivalent to one in 571 doctor consultations.

The rate of consultations were much higher in older age groups, as shown in Figure 2, with the highest rate in the 80 – 90 year age group at 128 consultations per 10,000 person-years.

There was a significant increase in rate of zoster consultations over time (odds ratio 0.86 [95% CI 0.78 to 0.94]; p = 0.0015). The rate of increase over time was particularly noticeable in the older age groups. This trend is shown in Figure 3.

Gender

When comparing age-standardised data by gender, females experienced a statistically significant higher rate of zoster over the study period compared to males (F=48.3, M=36.6, p< 0.001) as shown in Figure 4.

While the age-standardised rate of zoster increased over time generally, there was no statistically significant difference in the rate of increase between males or females.

Ethnicity and socioeconomic status

There were fewer index zoster consultations of those with Pacific Island ethnicity (age-adjusted rate 29.1 per 10,000 patient-years [95% CI 25.6 to 33.1]; p < 0.01) and Māori (New Zealand indigenous; rate 38.9 [36.3 to 41.6]; p = 0.019) than those from other ethnicities (rate 42.3 [41.4 to 43.2]). There was no significant difference in the rate of zoster index consultations across the different socioeconomic quintiles.

Health service utilisation

Herpes zoster-related consultations

Of the 10,316 zoster index consultations identified by the cohort analysis, most had no zoster follow-up consultations. Of these, 19% had a zoster follow-up consultation and only 5.8% consulted their doctor more than once in relation to zoster (Table 1).

Table 1: Distribution of follow-up consultations following zoster index cases

Follow-ups	0	1	2	3	4	5	6	7	8	9	10	11	12
Total number	8354	1365	372	100	73	16	14	8	2	6	2	3	1

Percentage	80.98	13.23	3.61	0.97	0.71	0.16	0.14	0.08	0.02	0.06	0.02	0.03	0.01

With increasing age, particularly from 45 years onwards, there was an increasing likelihood of follow-up consultations per episode as shown in Figure 5.

Distribution of non-herpes zoster consultations around index cases

To assess any correlation between zoster consultations and any changes in overall (non-zoster) consultations to general practice, we undertook a self-controlled case series analysis, consisting of 6,823 of the zoster index cases, and measured the variance in the number of all non-zoster consultations for 12 months prior to and 12 months after each zoster-index consultation occurred (a total of 27 months observation for each index consultation). This is expressed as the variance between the number of non-zoster consultations before and after the index consultation as a proportion of their sum. Positive proportions represented those with more consultations after the index consultation, while negative proportions represented those with more consultations before. Those with no consultations before or after were considered to have a proportion of zero as in Figure 6. There were no statistically significant differences in the distributions at a threshold of 0.001.

Data sharing

No additional data are available.

Discussion

Consistent with previously reported rates of incidence of zoster in primary care presentations, across many countries [5–8,10], this study found that the overall rate of incidence of zoster is 48.6 cases per 10,000 patient-years, with higher rates in the elderly [7] and a 32% higher rate in females than males [9]. The peak age of incidence seen in this study was in the 80 to 84 years age group, which is older than previously reported. An Australian study using Medicare Benefits Schedule items reported a peak age of 60-69 years.[24] The importance of showing a peak at an older age is that it will affect modelling for decision-making around the ideal age for zoster vaccine introduction.

Similar to other international findings, the overall incidence of zoster has increased from 2005 to 2016 across all age groups [6]. New Zealand did not have a childhood varicella vaccination programme over this period, supporting previous commentary that this increase is unlikely to be related to a decline in circulating varicella virus.[12]

Previous literature has reported differences between ethnic groups, most notably, a reduced self-reported occurrence was seen in US Blacks [25]. Our New Zealand -based study reported a lower age-adjusted incidence for those of Pacific Island ethnicity at 29.1 per 10,000 patient-years (95% CI 25.6 to 33.1) and a rate of 38.9 per 10,000 patient-years (95% CI 36.3 to 41.6]) for New Zealand indigenous Māori when compared to other ethnicities. This is in contrast to almost all other important health statistics for older Pacific Island and Māori populations in New

Zealand, for which there is a consistent equity gap, and poorer age-adjusted health outcomes are associated with these groups.[26] There does not appear to be any socioeconomic link as there was no significant difference between different levels of socioeconomic deprivation. A different burden of childhood varicella seems unlikely, because Māori are not a migrant group. This raises the question as to whether there are significant differences in rates across other ethnic groups internationally.

An important original finding from this study was the lack of evidence for increased burden of utilisation of health services at the primary care level. Utilising a large primary care data-set over an 11 year period, we have demonstrated an equivalent of one zoster-related consultation in every 571 general practice consultations. Furthermore, the burden of subsequent consulting was very low with 80% of zoster-related presentations requiring no follow-up and 13% requiring only a single follow-up consultation. While there was an increasing likelihood of follow-up consultations following a zoster episode with increasing age, particularly from 45 years onwards, the burden on general practice consultation rates was not significantly affected, overall. An episode of zoster is reported to frequently reduce overall quality of health particularly in older age groups, most likely related to the prolonged effects of post herpetic neuralgia.[27,28] International literature reports hospitalisation in adults aged over 50 years at a yearly rate of 28/100,000 zoster related hospitalisations as primary diagnosis (ranging from 6.1 in the 50-54 year age group to 95.8/100,000 persons in the over 80 years age group).[29] In New Zealand, during 2015 there were 361 hospitalisations with a primary diagnosis of zoster across all ages (Ministry of Health data). The burden of zoster admissions to hospital is severe when length of stay, cost and mortality are considered.[30] However, our study has demonstrated that these complications are not translating through to increasing utilisation of general practice services. This indicates that overall burden on the primary care services is less than has previously been suggested, and that zoster contributes very little to the overall utilisation burden in general practice.

Study Limitations

The gold standard for this study was based on doctor decision making, and the algorithm is limited by the quantity and detail of the recorded information in each consultation. In particular, repeat consultations may underreport ongoing zoster-related symptoms when the primary reason for a visit is in relation to other comorbidities. We did not examine in detail the reason for the zoster-related visit to assess the incidence of zoster complications. However this may not be a major concern as, for medico-legal reasons, significant clinical observations about ongoing zoster complications or progress are likely to be recorded. While recognising this, the overall incidence found in this study matches other international data, suggesting good concordance.

The number of less deprived individuals in the population studied was higher than that of the regional and national populations, but was otherwise well matched. Due to the self-exclusion of student health centres, the rate of zoster incidence in younger age-groups may be underreported, although, it is unlikely to affect overall incidence rates since cases are predominantly in those over 50 year olds. This study did not include out-of-hours presentations and patients not enrolled with a general practice. However, the main purpose of this study was

to report on service utilization burden in routine general practice, not periodic acute presentations in after-hours clinics. New Zealand has very high registration in general practice, particularly of the elderly population.[31]

Study Strengths

In this study, a very large data set of doctor consultations were examined by the way of a software inference algorithm. This methodology interrogated the free-text electronic medical records of over 6 million unique doctor consultations over an 11 year period. The large data set enabled analysis of rates of zoster incidence by age bands and different demographic measures, across the whole time period, despite the low frequency of zoster cases.

The strength of the NLP algorithm-based method used in our study is improved accuracy, when compared with the use of clinical codes or a simple keyword search, or review by a single clinical expert.[19,32] This methodology, as opposed to a simple keyword search, is able to identify the context in which pertinent terms are being used in clinical narrative. For example, a keyword may be used by a clinician to express either the presence or absence of the disease, which impacts the specificity and positive predictive value of that approach and ultimately over-estimating disease. A single clinical expert is prone to make errors, and has previously been shown to perform worse than a simple keyword search or NLP algorithm. In our study, two independent clinical coders reached concordance to provide a robust gold-standard comparison.

Using this NLP algorithm methodology, we have demonstrated its ability to review the burden of low frequency conditions, such as zoster, in primary care, and follow changes in patterns over times with large numbers.

Unanswered questions and future research

This study only focused on burden at the primary care level. Future research is needed on the comparative burden of disease across the full spectrum of health services: community, primary care and hospital inpatient care. Further important questions include disaggregation of the burden of disease by comorbidities, the effect of the use of antivirals and other treatment modalities and why the rate of zoster is continuing to increase over time.In addition, studies internationally have shown significant burden of disease on the quality of life of individuals, particularly those with severe or prolonged disease complications.[28] A qualitative study could provide insight into the effect shingles has on individuals and the potential value of a vaccine in reducing this burden. The low level of utilisation of general practice services suggests, however, that the burden of more severe disease falls on a small number of individuals and our results may prompt further discussion around modelling for whom to introduce zoster vaccines to in a population.

Conclusions

Overall, the rate of general practice doctor consultations related to zoster showed that this condition is rare in primary care and, while repeat visits following a zoster episode became increasingly common with age, the disease does not represent a significant burden on overall general practice workload. The peak age for consultations was older than has previously been reported, and there appear to be significant differences between ethnic groups, unrelated to socioeconomic circumstances. These are important findings particularly when considering the introduction of zoster vaccines across national immunisation schedules.

Use of a novel software algorithm to enable the exploration of consultation notes in EMRs is an efficient and effective way of identifying conditions, patterns, changes over time and the burden of disease on primary care services. We have also shown the methodology has the ability to review the burden of low frequency conditions in primary care.

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

Ethics approval: This study was approved by the University of Auckland Human Participants Ethics Committee Ref. 017617 on 25 Jul 2016¶

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

Contributors: NT, MN, AD conceived the study. All authors contributed to the development of the overall study methodology. MS and MN managed the ethical approval and consent processes. AD, LMB and NT provided clinical input into the algorithm design. JMR designed and built the natural language processing tools, programmed and trained the algorithm, and conducted the data analyses. AD and LMB classified the consultation records in the gold standard sets. NT and MN 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 in the study and can take responsibility for the integrity of the data and accuracy of the analysis.

Competing Interests. All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare no other support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years' no other relationships or activities that could appear to have influenced the submitted work.

Transparency declaration The lead author (NT) affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Exclusive licence "I, Nicola Turner, The Corresponding Author of this article contained within the original manuscript which includes any diagrams & photographs within and any related or stand-alone film submitted (the Contribution") has the right to grant on behalf of all authors and does grant on behalf of all authors, a licence to the BMJ Publishing Group Ltd and its licencees,

to permit this Contribution (if accepted) to be published in the BMJ and any other BMJ Group products and to exploit all subsidiary rights, as set out in our licence set out at: http://www.bmj.com/about-bmj/resources-authors/forms-policies-and-checklists/copyright-open-access-and-permission-reuse."

References¶

1 Eshleman E, Shahzad A, Cohrs RJ. Varicella zoster virus latency. *Future Virol* 2011;**6**:341–55.

2 Brisson M, Edmunds WJ, Law B, *et al.* Epidemiology of varicella zoster virus infection in Canada and the United Kingdom. *Epidemiol Infect* 2001;**127**:305–14.

3 Wehrhahn MC, Dwyer DE. Herpes zoster: epidemiology, clinical features, treatment and prevention. *Australian Prescriber* 2012;**35**:143–7.

4 Schmader K, Gnann JW, Jr., *et al.* The Epidemiological, Clinical, and Pathological Rationale for the Herpes Zoster Vaccine. *J Infect Dis* 2008;**197**:S207–15.

5 Gauthier A, Breuer J, Carrington D, *et al.* Epidemiology and cost of herpes zoster and post-herpetic neuralgia in the United Kingdom. *Epidemiol Infect* 2009;**137**:38–47.

6 Yawn BP, Gilden D. The global epidemiology of herpes zoster. *Neurology* 2013;**81**:928–30.

7 Kawai K, Gebremeskel BG, Acosta CJ. Systematic review of incidence and complications of herpes zoster: towards a global perspective. *BMJ Open* 2014;**4**:e004833.

8 Pinchinat S, Cebrián-Cuenca AM, Bricout H, *et al.* Similar herpes zoster incidence across Europe: results from a systematic literature review. *BMC Infect Dis* 2013;**13**:170.

9 Yawn BP, Saddier P, Wollan PC, Sauver JLS, Kurland MJ, Sy LS. A population-based study of the incidence and complication rates of herpes zoster before zoster vaccine introduction. In: *Mayo Clinic Proceedings*. Mayo Clinic 2007. 1341–9.

10 Reid JS, Ah Wong B. Herpes zoster (shingles) at a large New Zealand general practice: incidence over 5 years. *N Z Med J* 2014;**127**:56–60.

11 Johnson RW. Herpes zoster and postherpetic neuralgia. *Expert Rev Vaccines* 2010;**9**:21–6.

12 Hales CM, Harpaz R, Joesoef MR, *et al.* Examination of links between herpes zoster incidence and childhood varicella vaccination. *Ann Intern Med* 2013;**159**:739–45.

13 Jumaan AO, Yu O, Jackson LA, *et al.* Incidence of Herpes Zoster, Before and After Varicella-Vaccination–Associated Decreases in the Incidence of Varicella, 1992–2002. *J Infect Dis* 2005;**191**:2002–7.

14 Leung J, Harpaz R, Molinari N-A, *et al.* Herpes zoster incidence among insured persons in the United States, 1993-2006: evaluation of impact of varicella vaccination. *Clin Infect Dis* 2011;**52**:332–40.

15 Toyama N, Shiraki K, Society of the Miyazaki Prefecture Dermatologists. Epidemiology of herpes zoster and its relationship to varicella in Japan: A 10-year survey of 48,388 herpes zoster cases in Miyazaki prefecture. *J Med Virol* 2009;**81**:2053–8.

16 Russell ML, Schopflocher DP, Svenson L, *et al.* Secular trends in the epidemiology of shingles in Alberta. *Epidemiol Infect* 2007;**135**:908–13.

17 Varicella and herpes zoster vaccines: WHO position paper, June 2014. *Wkly Epidemiol Rec* 2014;**89**:265–87.

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**:78.

20 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.

21 Fay MP, Feuer EJ. Confidence intervals for directly standardized rates: a method based on the gamma distribution. *Stat Med* 1997;**16**:791–801.

22 Houghton F. Degrees of Deprivation in New Zealand — An Atlas of Socioeconomic Difference: By P. Crampton, C. Salmond, R. Kirkpatrick, R. Scarborough and C. Skelly. *N Z Geog* 2001;**57**:62–62.

23 Salmond C, Crampton P, Sutton F. NZDep91: A New Zealand index of deprivation. *Aust N Z J Public Health* 1998;**22**:835–7.

24 Karki S, Newall AT, Raina MacIntyre C, *et al.* Healthcare Resource Utilisation Associated with Herpes Zoster in a Prospective Cohort of Older Australian Adults. *PLoS One* 2016;**11**:e0160446.

25 Schmader K, George LK, Burchett BM, *et al.* Racial differences in the occurrence of herpes zoster. *J Infect Dis* 1995;**171**:701–4.

26 Anderson I, Crengle S, Kamaka ML, *et al.* Indigenous health in Australia, New Zealand, and the Pacific. *Lancet* 2006;**367**:1775–85.

27 Serpell M, Gater A, Carroll S, *et al.* Burden of post-herpetic neuralgia in a sample of UK residents aged 50 years or older: findings from the Zoster Quality of Life (ZQOL) study. *Health Qual Life Outcomes* 2014;**12**:92.

28 Gater A, Abetz-Webb L, Carroll S, *et al.* Burden of herpes zoster in the UK: findings from the zoster quality of life (ZQOL) study. *BMC Infect Dis* 2014;**14**:402.

29 Stein AN, Britt H, Harrison C, *et al.* Herpes zoster burden of illness and health care resource utilisation in the Australian population aged 50 years and older. *Vaccine* 2009;**27**:520–9.

30 Hobbelen PHF, Stowe J, Amirthalingam G, *et al.* The burden of hospitalisation for varicella and herpes zoster in England from 2004 to 2013. *J Infect* 2016;**73**:241–53.

31 Enrolment in a primary health organisation. Ministry of Health NZ. http://www.health.govt.nz/our-work/primary-health-care/about-primary-health-organisations/enrolment-primary-health-organisation (accessed 27 Oct 2017).

32 Wu M-F, Yang Y-W, Lin W-Y, *et al.* Varicella zoster virus infection among healthcare workers in Taiwan: seroprevalence and predictive value of history of varicella infection. *J Hosp Infect* 2012;**80**:162–7.

Figure legends

Figure 1: Recruitment Process

Figure 2: Herpes zoster index consultation rate by age group (bars represent 95% CIs)

Figure 3: Herpes zoster index consultation rate over time from age of 50 years, by five year age groups (n is mean patient years per year, with 95% confidence intervals and linear regression shown in blue).

Figure 4: Herpes zoster consultation rate of index case over time by gender, age adjusted (95% CIs)

Figure 5: Proportion of herpes zoster consultations with or without follow-up visits by age group showing 95% CIs.

Figure 6 Distribution of overall consultation visits occurring prior to and after a herpes zoster index case by age group

Table 1: Distribution of follow-up consultations following zoster index cases

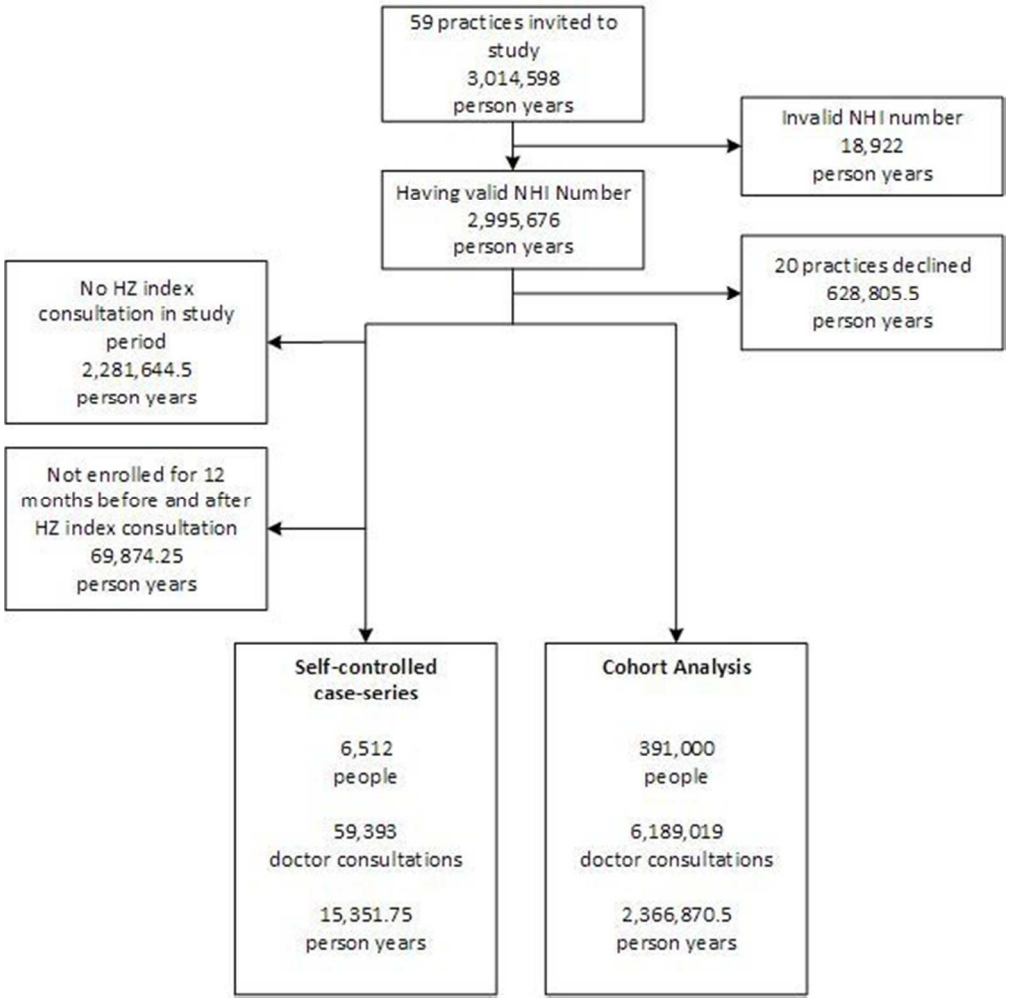

Figure 1: Recruitment Process

48x48mm (300 x 300 DPI)

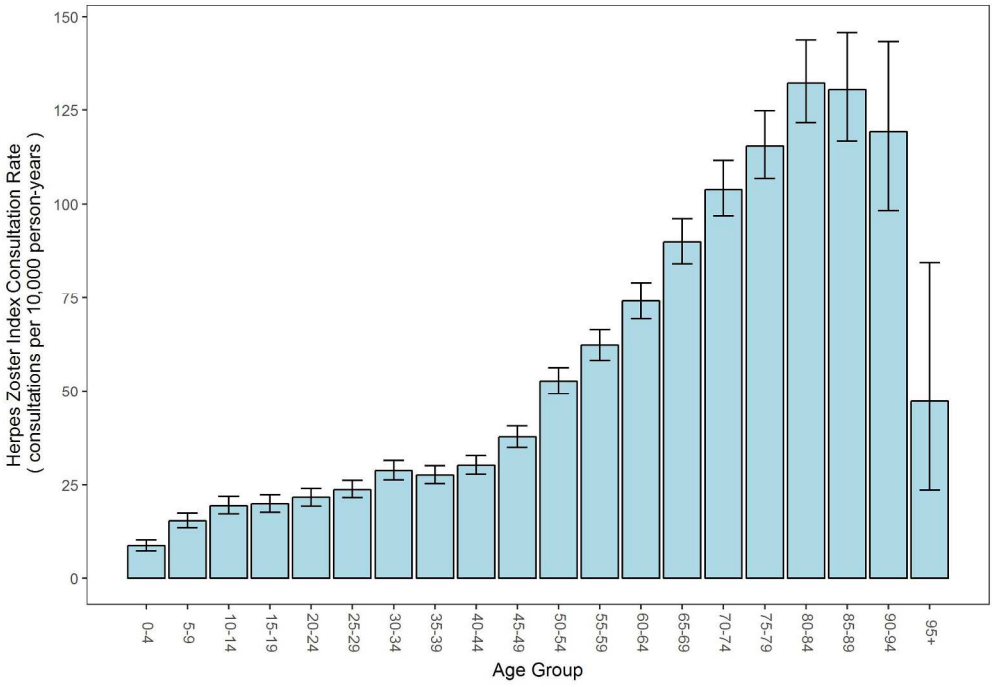

Figure 2: Herpes zoster index consultation rate by age group (bars represent 95% CIs)

210x145mm (300 x 300 DPI)

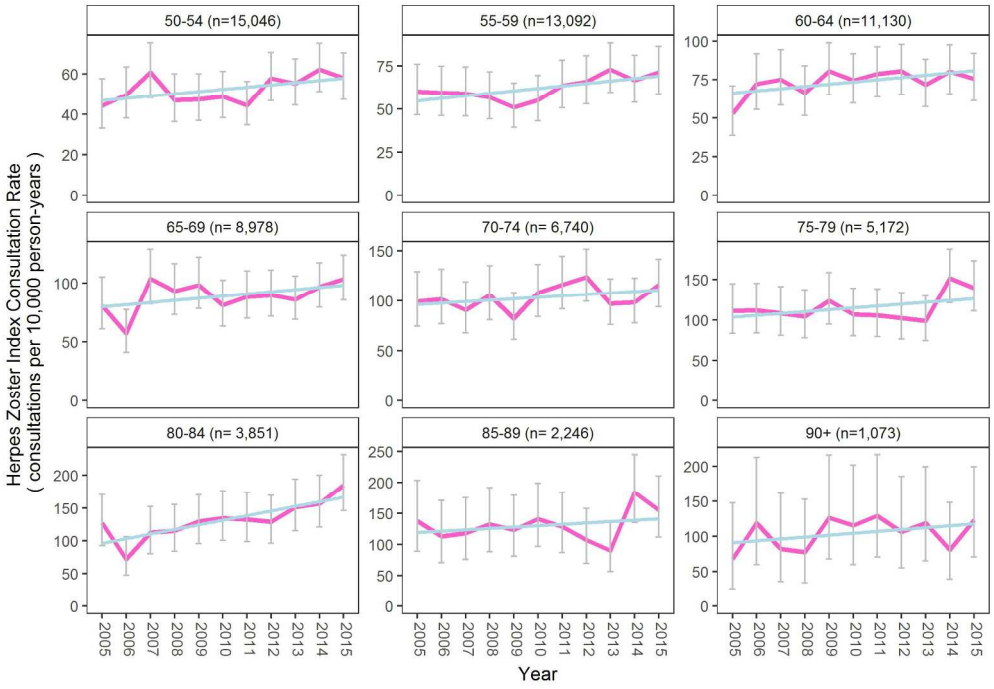

Figure 3: Herpes zoster index consultation rate over time from age of 50 years, by five year age groups (n is mean patient years per year, with 95% confidence intervals and linear regression shown in blue).

210x145mm (300 x 300 DPI)

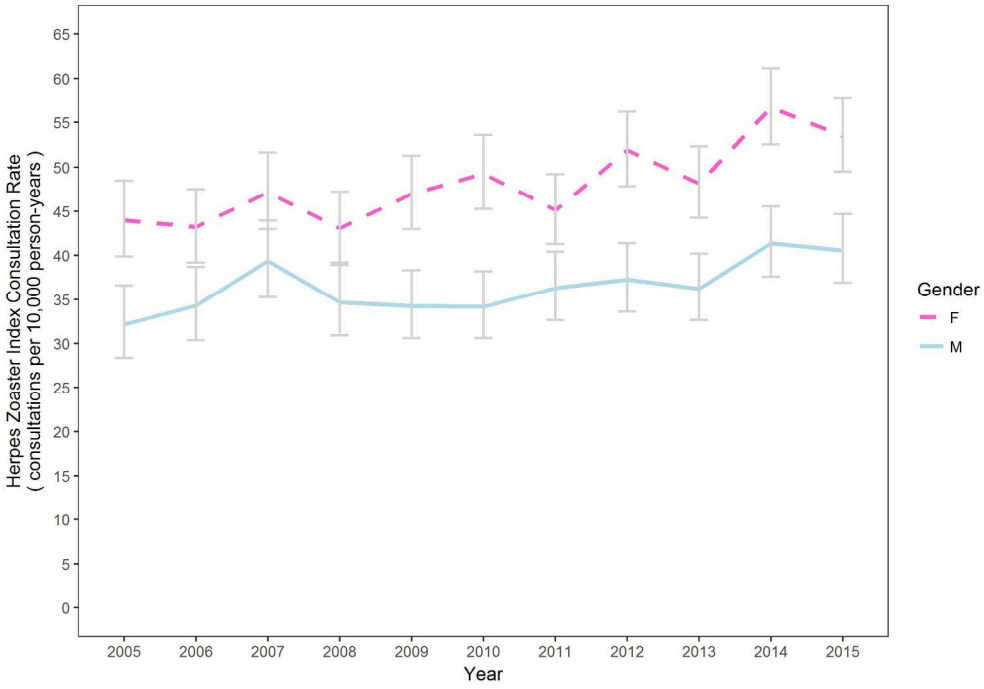

Figure 4: Herpes zoster consultation rate of index case over time by gender, age adjusted (95% CIs)

210x145mm (300 x 300 DPI)

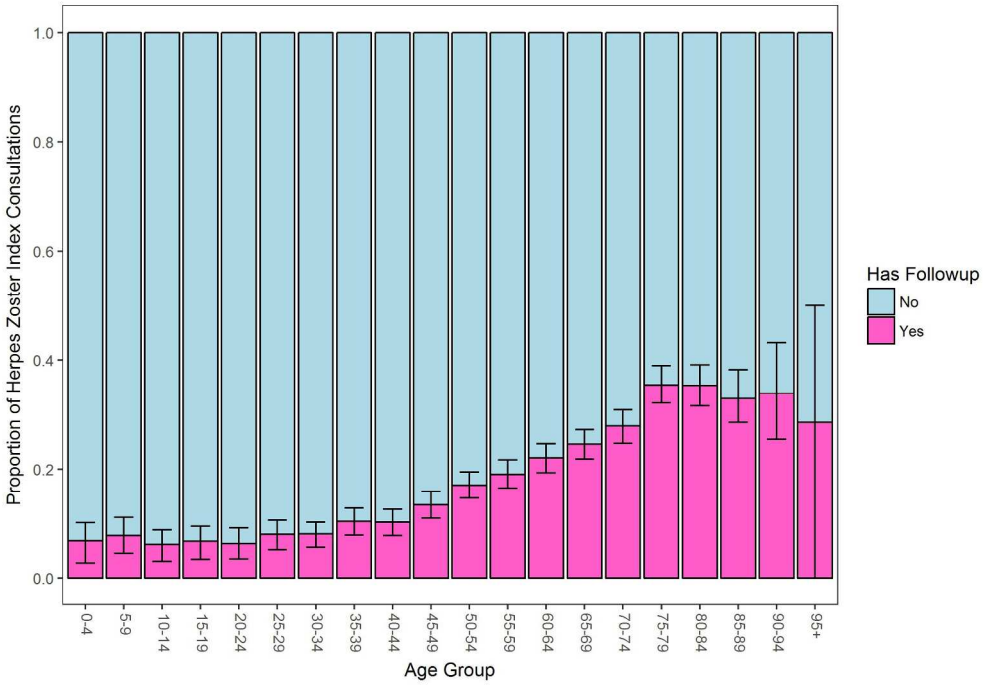

Figure 5: Proportion of herpes zoster consultations with or without follow-up visits by age group showing 95% CIs.

210x145mm (300 x 300 DPI)

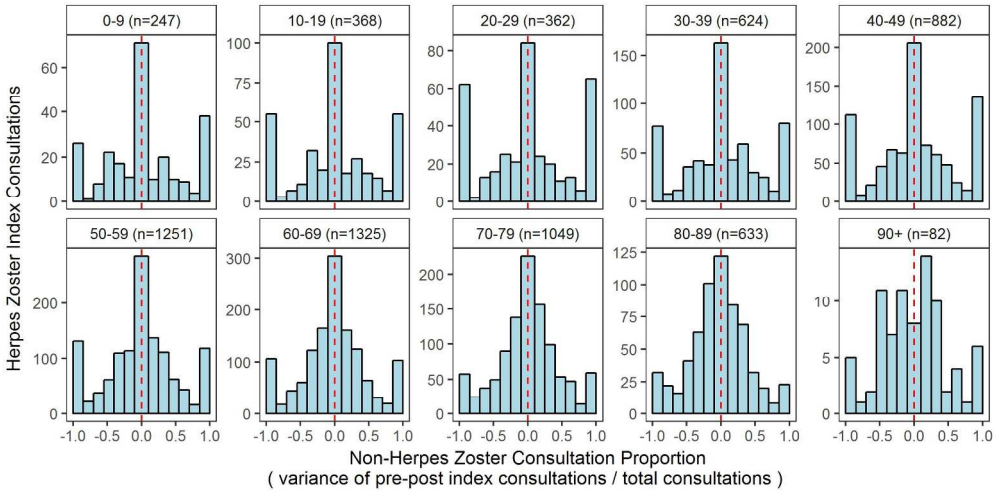

Figure 6 Distribution of overall consultation visits occurring prior to and after a herpes zoster index case by age group

211x105mm (300 x 300 DPI)

ORD statement – checklist of items, extended from the STROBE statement, that should be reported in observational stud collected health data.

	Item No.	STROBE items	Location in manuscript where items are reported	RECORD items	Location manuscri where ite reported
bstract					
	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found	Title - line 6 -7 Abstract line 10	RECORD 1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included. RECORD 1.2: If applicable, the geographic region and timeframe within which the study took place should be reported in the title or abstract. RECORD 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.	Title - lin Abstract - Title: line Abstract - 14, line 1 Abstract - 14
on					
d	2	Explain the scientific background and rationale for the investigation being reported	page 4 paragraph 1 and 2		
	3	State specific objectives, including any prespecified hypotheses	page 4 paragraph 3 - line 36		
gn	4	Present key elements of study design early in the paper	page 4 paragraph 1 line 47-48		
	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	page 5 line 53-		
s	6	*(a) Cohort study* - Give the	page 5 line 11	RECORD 6.1: The methods of study	Flow char

		eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study* - Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study* - Give the eligibility criteria, and the sources and methods of selection of participants *(b) Cohort study* - For matched studies, give matching criteria and number of exposed and unexposed *Case-control study* - For matched studies, give matching criteria and the number of controls per case		population selection (such as codes or algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided. RECORD 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided. RECORD 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data linkage process, including the number of individuals with linked data at each stage.	figure 1 Page 5 lin Reference Page 5 lin Figure 1 - chart Page 5, li line 49
	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.		RECORD 7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided.	Process, p line 42 an reference
es/ nt	8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	flow chart Figure 1 analysis page 4 line 12, page 5 lines 50 and 54 Not applicable		
	9	Describe any efforts to address potential sources of bias	page 5 line 34		

	10	Explain how the study size was arrived at	settings and participants from page 4 line 56 to page 5 line 9 Figure 1 - flow chart		
e	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why			Analysis line 49 – line 7
	12	(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) *Cohort study* - If applicable, explain how loss to follow-up was addressed *Case-control study* - If applicable, explain how matching of cases and controls was addressed *Cross-sectional study* - If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses	Page 6 line 6 Page 5 line 22 page 6 paragraph 1 page 5 line 38 page 6 line 38		
s and ethods		..	analysis page 6 line 8	RECORD 12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population.	Settings a participan page 5, li

				RECORD 12.2: Authors should provide information on the data cleaning methods used in the study.	analysis P line 8
		..		RECORD 12.3: State whether the study included person-level, institutional-level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.	Settings a participan 5 line 18
s	13	(a) Report the numbers of individuals at each stage of the study (*e.g.*, numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed) (b) Give reasons for non-participation at each stage. (c) Consider use of a flow diagram		RECORD 13.1: Describe in detail the selection of the persons included in the study (*i.e.,* study population selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in the text and/or by means of the study flow diagram.	figure 1 - chart
e data	14	(a) Give characteristics of study participants (*e.g.*, demographic, clinical, social) and information on exposures and potential confounders (b) Indicate the number of participants with missing data for each variable of interest (c) *Cohort study* - summarise follow-up time (*e.g.*, average and total amount)			Study coh demograp page 6 lin Accuracy identifica page 6 lin
ata	15	*Cohort study* - Report numbers of outcome events or summary measures over time *Case-control study* - Report			herpes zo consultati page 6 lin

		numbers in each exposure category, or summary measures of exposure *Cross-sectional study* - Report numbers of outcome events or summary measures			healthcare utilisation from line Table 1
ts	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period			Confidenc intervals f statistical analyses a presented figures. Methods, page 6 lin
yses	17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses			Accuracy herpes zo identifica page 6 lin
	18	Summarise key results with reference to study objectives			paragraph
s	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias		RECORD 19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.	study limi page 8 lin page 8 lin
on	20	Give a cautious overall interpretation of results			page 9 lin

		considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence			
bility	21	Discuss the generalisability (external validity) of the study results			conclusio 10 line 47 line 53
rmation					
	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based			funding, p line 21
ty of aw ng		..		RECORD 22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code.	not applic

: Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM, the RECORD
. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Medi*

BMJ Open

Quantifying the incidence and burden of herpes zoster in New Zealand general practice: a retrospective cohort study utilising a natural language processing software inference algorithm.

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-021241.R2
Article Type:	Research
Date Submitted by the Author:	12-Apr-2018
Complete List of Authors:	Turner, Nikki; University of Auckland, General Practice and Primary Health Care MacRae, Jayden; Datacraft Analytics Nowlan, Mary; University of Auckland Tamaki Campus, General Practice and Primary Health Care McBain, Lynn; University of Otago, Wellington, Stubbe, Maria; University of Otago, Primary Health Care and General Practice Dowell, Anthony; Wellington School of Medicine and Health Sciences,
Primary Subject Heading:	General practice / Family practice
Secondary Subject Heading:	Health informatics, Health services research, Infectious diseases, Research methods
Keywords:	herpes zoster, PRIMARY CARE, VIROLOGY, Health informatics < BIOTECHNOLOGY & BIOINFORMATICS

Quantifying the incidence and burden of herpes zoster in New Zealand general practice: a retrospective cohort study utilising a natural language processing software inference algorithm.

N M Turner, Associate Professor[1],

J MacRae, Chief Scientist[2]

M L Nowlan, Project Manager[1],

L McBain, Senior Lecturer[3],

M Stubbe, Senior Lecturer[3]

A Dowell, Professor[3]

1. Department of General Practice and Primary Care, University of Auckland, Private Bag 92019, Victoria Street West, Auckland 1142, New Zealand .
2. DataCraft Analytics Limited, PO Box 17271, Karori, Wellington 6147
3. Department of Primary Care and General Practice, University of Otago Wellington, PO Box 7343 Newtown Wellington 6242 New Zealand

Correspondence to Dr. N Turner; n.turner@auckland.ac.nz

Abstract

Objective: To investigate the incidence of primary care presentations for herpes zoster (zoster) in a representative New Zealand population and to evaluate the utilisation of primary health care services following zoster diagnosis.

Design: A cross-sectional retrospective cohort study used a natural language processing software inference algorithm to identify general practice consultations for zoster by interrogating 22 million electronic medical record (EMR) transactions routinely recorded from January 2005 to December 2015. Data-linking enabled analysis of the demographics of each case. The frequency of doctor visits were assessed prior to and after the first consultation diagnosing zoster to determine health service utilisation.

Setting: General practice, using EMRs from two primary health organisations located in the lower North Island, New Zealand.

Participants: Thirty-nine general practices consented interrogation of their EMRs to access de-identified records for all enrolled patients. Out-of-hours and practice nurse consultations were excluded.

Main outcome measures: The incidence of first and repeated zoster-related visits to the doctor across all age groups and associated patient demographics. To determine whether zoster affects workload in general practice.

Results: Overall, for 6,189,019 doctor consultations, the incidence of zoster was 48.6 per 10,000 patient years (95% CI 47.6 to 49.6). Incidence increased from the age of 50 years to a peak rate of 128 per 10,000 in the 80-90 years age group and was significantly higher in females than males ($p < 0.001$). Over this 11 year period, incidence increased gradually, notably in those aged 80-85 years. Only 19% of patients had one or more follow-up zoster consultations within 12 months of a zoster index consultation. The frequency of consultations, for any reason, did not change between periods before and after the diagnosis.

Conclusions: Zoster consultations in general practice are rare and the burden of these cases on overall general practice caseload is low.

Article summary

Strengths and limitations of the study

- This study used a novel and validated natural language processing software inference algorithm to identify herpes zoster presentation rates and service utilisation using primary care electronic medical records over an eleven year period.

- Despite a low frequency of zoster cases, the large data set enabled analysis of rates of zoster incidence by age bands and different demographics across the whole time period.

- The algorithm was designed to maximise specificity and accuracy, thereby generating a conservative estimate of the burden of zoster presentations in primary care by keeping false positives to a minimum.

- The gold standard for this study was based on doctor decision making, and the algorithm is limited by the quantity and detail of the recorded information in each consultation. Details of the reason for each visit, such as specific zoster complications, were not determined, just that the visit was identified as zoster related.

- 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 rates.

Introduction¶

Infection with varicella-zoster virus (VZV) establishes life-long persistence in sensory nerve ganglia. When the immune system is unable to maintain the suppression of the virus, it manifests as a clinical syndrome known as herpes zoster (zoster; shingles).[1] About one third of the population experiences zoster, with greater incidence in older age groups.[2] The major risk factor for zoster is a decline in VZV immunity as cell-mediated immunity wanes with age.[3,4] Cohort studies, from at least 22 countries, give incidence of zoster in the general population from 3 to 5 per 1000 person-years [5–10], and a lifetime incidence of 20 - 30%.[2] Much higher incidence is reported in older adults, increasing from 50 – 60 years of age, and ranging from 8 – 12 per 1000 person-years.[2,7] Around half of the zoster cases over 70 years of age will develop postherpetic neuralgia (PHN).[11] The main rationale for the use of a zoster vaccine is to prevent the long-lasting effects of PHN. A live attenuated zoster vaccine is internationally available for adults over the age of 50 years, and is on the national schedule in some countries, notably, the United Kingdom and Australia at the age of 70 years, and recommended in the USA from age 60 years.[1]

Several studies have reported increasing incidence rates over time across all age groups [6], in countries both with [12–14] and without childhood varicella immunisation programmes.[2,15,16] The World Health Organization recommends, when considering the optimal age and dosing schedule for zoster vaccines, that countries should take into consideration the age-dependent burden of disease, vaccine effectiveness, duration of protection and cost effectiveness of the vaccine.[17] While there is a body of research around zoster and the burden of zoster disease, this is almost exclusively based on observational studies from administrative databases or health records, which will underestimate cases not coming to medical attention.[6] Decisions around vaccination strategies for countries remain hindered by a more complete understanding of the burden of the disease for a population.

The objective of this study was to interrogate data from general practice electronic medical records (EMR) to identify zoster-related primary care presentations and service utilisation associated with consultations for zoster-related conditions. Previous studies have shown the potential of utilising a novel software algorithm to enable the exploration of consultation notes in EMR systems used commonly for OECD primary care records.[18] To date this software has shown the ability to analyse service utilisation for H1N1 influenza and childhood respiratory diseases while eliminating reliance on clinical coding.[19][20]

Methods¶

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

Setting and participants

The development of the algorithm methodology has been earlier described.[18] New Zealand has universal enrolment of the population with a primary care (general) practice, and universal computerised recording of general practice doctor (GP) consultations. The study was conducted

in the lower North Island in a mixed urban and rural setting. It consisted of 39 consenting general practices from two primary health organisations (PHOs) giving a total unique enrolled population of 391,000 over the study period between 1 January 2005 to 31 December 2015. This included individuals who both joined and left the cohort during this period. The cohort totalled 2,366,870.5 person-years and contained over 22 million medical record transactions representing 6,189,019 doctor consultations.

Figure 1 shows the selection process, and steps taken through the development and analysis of the data set. Data was extracted directly from the EMR using software that automates the collection and secure transmission of large datasets. The complete dataset was filtered to identify doctor consultations generated during standard office hours. Practice nurse and out-of-hours consultations were excluded. Each consultation record was linked to the individual's unique National Health Index number. This individual unique identifier is assigned to every person who uses health services in New Zealand to enable records to be matched between datasets. All data were analysed on the premises of the PHO utilising rigorous protocols to ensure patient confidentiality, and no member of the research team accessed identifiable data. Rates were age-standardised, where appropriate, using the direct method and exact confidence intervals (CI) were calculated.[21]

Patient and public involvement

This retrospective observational study on general practice doctor notes did not directly involve patients.

Process

The software algorithm was designed to replicate the diagnosis and assessment made by clinical experts.[18] It was trained, validated and tested using three independent data sets of 800 doctor consultation records. Each data set was stratified to contain 600 randomly chosen records and a further 200 random records from all that contained a simple keyword related to zoster. Clinical records from any single practice or provider could only exist in a single data set so as to avoid any training bias during validation or test procedures. Each record in all data sets was independently classified by two general practice clinical experts (LMB and AD). At the completion of coding all records, the clinical experts reviewed, discussed and reached consensus on any records where they showed a discrepancy in coding. The focus of algorithm development was to maximise specificity and positive predictive values to reduce false positives because of the expected relative rarity of zoster consultations.

A zoster index consultation was defined as the initial zoster consultation occurring for an individual within the past 12 months. A follow-up visit was any visit identified as related to zoster occurring within 12 months of any prior zoster consultation.

Analysis

The demographic characteristics of the study cohort was measured by age, gender, patient-identified ethnicity, and socioeconomic deprivation quintile via the New Zealand Deprivation Index (a small census-area measure of socioeconomic deprivation).[22,23] These were compared with those of all patients enrolled with the PHO (N=3,014,598 person-years) and the 2013 New Zealand Census data (N=4,353,198 people). Patients were observed for the period they were enrolled in a participating practice over the study period. In order to maintain a consistency of analysis across the study period, both deprivation and ethnicity were determined from the most recently recorded information available for each patient. Where appropriate, consultation rates were adjusted for algorithm sensitivity, specificity and age-adjusted for the 2013 New Zealand Census population using direct-standardisation. All data aggregation, transformation, cleaning and storage were done in Microsoft SQL Server, and statistical analysis was undertaken in R. Confidence intervals for crude rates were calculated using a bootstrap over 1000 replicate rounds with resampling. Confidence intervals on age-standardised rates were made using the method described by Fay and Feuer [21]. Hypothesis tests were conducted using the Fischer exact test for cohort analysis and the Wilcoxon signed rank test for self-controlled case series.

Results

Study cohort demographics

The age and gender of the study cohort approximated the national census profile. There was a larger proportion of 18-23 year olds in the local enrolled population than in the study cohort, which is likely to be due to self-exclusion of practices providing student health services to the universities within the catchment area. While the demographic characteristics of the study cohort closely matched those of the enrolled population for deprivation, due to the geocoding system used for deprivation, 6% of both the enrolled and study cohort populations with incorrectly documented address information were not allocated deprivation scores. The study cohort had a higher proportion of people in the least deprived socioeconomic quintile with a correspondingly lower proportion in the two most deprived. As compared with the enrolled population, the study cohort had fewer with Pacific Island ethnicity (3% versus 5%) and more 'Other' ethnicities, which included New Zealand European and Asian (88% versus 86%).

Accuracy of herpes zoster identification

The natural language algorithm had a positive-predictive value (PPV) of 0.82 (95% CI 0.72 to 0.92), specificity of 0.9998 (95% CI 0.9997 to 0.9999) and sensitivity of 0.84 (95% CI 0.74 to 0.92). This was more accurate than using keywords only (PPV 0.66, specificity 0.9994, and sensitivity 1.0), or using a single clinical expert (PPV 0.53, specificity 0.9991, and sensitivity 0.93).

Herpes zoster consultations

The overall age-adjusted apparent rate of zoster index consultations was 42.7 per 10,000 person-years observed (95% CI 41.9 to 43.5), with an estimated true rate of 48.6 (95% CI 47.6 to 49.6). There were 10,316 index consultations for zoster and 3,060 zoster-related follow-up consultations. The apparent rate for zoster index consultations was 16.7 per 10,000 doctor consultations (95% CI 16.3 to 17.0) with an estimated true rate of 17.5 (95% CI 17.1 to 17.9). This was the equivalent to one in 571 doctor consultations.

The rate of consultations were much higher in older age groups, as shown in Figure 2, with the highest rate in the 80 – 90 year age group at 128 consultations per 10,000 person-years.

There was a significant increase in rate of zoster consultations over time (odds ratio 0.86 [95% CI 0.78 to 0.94]; p = 0.0015). The rate of increase over time was particularly noticeable in the older age groups. This trend is shown in Figure 3.

Gender

When comparing age-standardised data by gender, females experienced a statistically significant higher rate of zoster over the study period compared to males (F=48.3, M=36.6, p< 0.001) as shown in Figure 4.

While the age-standardised rate of zoster increased over time generally, there was no statistically significant difference in the rate of increase between males or females.

Ethnicity and socioeconomic status

There were fewer index zoster consultations of those with Pacific Island ethnicity (age-adjusted rate 29.1 per 10,000 patient-years [95% CI 25.6 to 33.1]; p < 0.01) and Māori (New Zealand indigenous; rate 38.9 [36.3 to 41.6]; p = 0.019) than those from other ethnicities (rate 42.3 [41.4 to 43.2]). There was no significant difference in the rate of zoster index consultations across the different socioeconomic quintiles.

Health service utilisation

Herpes zoster-related consultations

Of the 10,316 zoster index consultations identified by the cohort analysis, most had no zoster follow-up consultations. Of these, 19% had a zoster follow-up consultation and only 5.8% consulted their doctor more than once in relation to zoster (Table 1).

Table 1: Distribution of follow-up consultations following zoster index cases

Follow-ups	0	1	2	3	4	5	6	7	8	9	10	11	12
Total number	8354	1365	372	100	73	16	14	8	2	6	2	3	1

Percentage	80.98	13.23	3.61	0.97	0.71	0.16	0.14	0.08	0.02	0.06	0.02	0.03	0.01

With increasing age, particularly from 45 years onwards, there was an increasing likelihood of follow-up consultations per episode as shown in Figure 5.

Distribution of non-herpes zoster consultations around index cases

To assess any correlation between zoster consultations and any changes in overall (non-zoster) consultations to general practice, we undertook a self-controlled case series analysis, consisting of 6,823 of the zoster index cases, and measured the variance in the number of all non-zoster consultations for 12 months prior to and 12 months after each zoster-index consultation occurred (a total of 27 months observation for each index consultation). This is expressed as the variance between the number of non-zoster consultations before and after the index consultation as a proportion of their sum. Positive proportions represented those with more consultations after the index consultation, while negative proportions represented those with more consultations before. Those with no consultations before or after were considered to have a proportion of zero as in Figure 6. There were no statistically significant differences in the distributions at a threshold of 0.001.

Data sharing

No additional data are available.

Discussion

Consistent with previously reported rates of incidence of zoster in primary care presentations, across many countries [5–8,10], this study found that the overall rate of incidence of zoster is 48.6 cases per 10,000 patient-years, with higher rates in the elderly [7] and a 32% higher rate in females than males [9]. The peak age of incidence seen in this study was in the 80 to 84 years age group, which is older than previously reported. An Australian study using Medicare Benefits Schedule items reported a peak age of 60-69 years.[24] The importance of showing a peak at an older age is that it will affect modelling for decision-making around the ideal age for zoster vaccine introduction.

Similar to other international findings, the overall incidence of zoster has increased from 2005 to 2016 across all age groups [6]. New Zealand did not have a childhood varicella vaccination programme over this period, supporting previous commentary that this increase is unlikely to be related to a decline in circulating varicella virus.[12]

Previous literature has reported differences between ethnic groups, most notably, a reduced self-reported occurrence was seen in US Blacks [25]. Our New Zealand -based study reported a lower age-adjusted incidence for those of Pacific Island ethnicity at 29.1 per 10,000 patient-years (95% CI 25.6 to 33.1) and a rate of 38.9 per 10,000 patient-years (95% CI 36.3 to 41.6]) for New Zealand indigenous Māori when compared to other ethnicities. This is in contrast to almost all other important health statistics for older Pacific Island and Māori populations in New

Zealand, for which there is a consistent equity gap, and poorer age-adjusted health outcomes are associated with these groups.[26] There does not appear to be any socioeconomic link as there was no significant difference between different levels of socioeconomic deprivation. A different burden of childhood varicella seems unlikely, because Māori are not a migrant group. This raises the question as to whether there are significant differences in rates across other ethnic groups internationally.

An important original finding from this study was the lack of evidence for increased burden of utilisation of health services at the primary care level. Utilising a large primary care data-set over an 11 year period, we have demonstrated an equivalent of one zoster-related consultation in every 571 general practice consultations. Furthermore, the burden of subsequent consulting was very low with 80% of zoster-related presentations requiring no follow-up and 13% requiring only a single follow-up consultation. While there was an increasing likelihood of follow-up consultations following a zoster episode with increasing age, particularly from 45 years onwards, the burden on general practice consultation rates was not significantly affected, overall. An episode of zoster is reported to frequently reduce overall quality of health particularly in older age groups, most likely related to the prolonged effects of post herpetic neuralgia.[27,28] International literature reports hospitalisation in adults aged over 50 years at a yearly rate of 28/100,000 zoster related hospitalisations as primary diagnosis (ranging from 6.1 in the 50-54 year age group to 95.8/100,000 persons in the over 80 years age group).[29] In New Zealand, during 2015 there were 361 hospitalisations with a primary diagnosis of zoster across all ages (Ministry of Health data). The burden of zoster admissions to hospital is severe when length of stay, cost and mortality are considered.[30] However, our study has demonstrated that these complications are not translating through to increasing utilisation of general practice services. This indicates that overall burden on the primary care services is less than has previously been suggested, and that zoster contributes very little to the overall utilisation burden in general practice.

Study Limitations

This study is specifically focused on the burden of zoster from the general practice perspective. As such, we are unable to comment on other aspects of the burden of zoster, such as patient quality of life measures, or the burden on the health service beyond the general practice, which includes the frequency of referrals to secondary healthcare services, hospitalisation and prescriptions.

The gold standard for this study was based on doctor decision making, and the algorithm is limited by the quantity and detail of the recorded information in each consultation. In particular, repeat consultations may underreport ongoing zoster-related symptoms when the primary reason for a visit is in relation to other comorbidities. We did not examine in detail the reason for the zoster-related visit to assess the incidence of zoster complications, such as post herpetic neuralgia. However this may not be a major concern as, for medico-legal reasons, significant clinical observations about ongoing zoster complications or progress are likely to be recorded. While recognising this, the overall incidence found in this study matches other international data, suggesting good concordance.

The number of less deprived individuals in the population studied was higher than that of the regional and national populations, but was otherwise well matched. Due to the self-exclusion of student health centres, the rate of zoster incidence in younger age-groups may be underreported, although, it is unlikely to affect overall incidence rates since cases are predominantly in those over 50 year olds. This study did not include out-of-hours presentations and patients not enrolled with a general practice. However, the main purpose of this study was to report on service utilization burden in routine general practice, not periodic acute presentations in after-hours clinics. New Zealand has very high registration in general practice, particularly of the elderly population.[31]

Study Strengths

In this study, a very large data set of doctor consultations were examined by the way of a software inference algorithm. This methodology interrogated the free-text electronic medical records of over 6 million unique doctor consultations over an 11 year period. The large data set enabled analysis of rates of zoster incidence by age bands and different demographic measures, across the whole time period, despite the low frequency of zoster cases.

The strength of the NLP algorithm-based method used in our study is improved accuracy, when compared with the use of clinical codes or a simple keyword search, or review by a single clinical expert.[19,32] This methodology, as opposed to a simple keyword search, is able to identify the context in which pertinent terms are being used in clinical narrative. For example, a keyword may be used by a clinician to express either the presence or absence of the disease, which impacts the specificity and positive predictive value of that approach and ultimately over-estimating disease. A single clinical expert is prone to make errors, and has previously been shown to perform worse than a simple keyword search or NLP algorithm. In our study, two independent clinical coders reached concordance to provide a robust gold-standard comparison.

Using this NLP algorithm methodology, we have demonstrated its ability to review the burden of low frequency conditions, such as zoster, in primary care, and follow changes in patterns over times with large numbers.

Unanswered questions and future research

This study only focused on burden at the primary care level. Future research is needed on the comparative burden of disease across the full spectrum of health services: community, primary care and hospital inpatient care. Further important questions include disaggregation of the burden of disease by comorbidities, the effect of the use of antivirals and other treatment modalities and why the rate of zoster is continuing to increase over time. In addition, studies internationally have shown significant burden of disease on the quality of life of individuals, particularly those with severe or prolonged disease complications.[28] A qualitative study could provide insight into the effect shingles has on individuals and the potential value of a vaccine in reducing this burden. The low level of utilisation of general practice services suggests, however, that the burden of more severe disease falls on a small number of individuals and our results

may prompt further discussion around modelling for whom to introduce zoster vaccines to in a population.

Conclusions

Overall, the rate of general practice doctor consultations related to zoster showed that this condition is rare in primary care and, while repeat visits following a zoster episode became increasingly common with age, the disease does not represent a significant burden on overall general practice workload. The peak age for consultations was older than has previously been reported, and there appear to be significant differences between ethnic groups, unrelated to socioeconomic circumstances. These are important findings particularly when considering the introduction of zoster vaccines across national immunisation schedules.

Use of a novel software algorithm to enable the exploration of consultation notes in EMRs is an efficient and effective way of identifying conditions, patterns, changes over time and the burden of disease on primary care services. We have also shown the methodology has the ability to review the burden of low frequency conditions in primary care.

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

Ethics approval: This study was approved by the University of Auckland Human Participants Ethics Committee Ref. 017617 on 25 Jul 2016¶

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

Contributors: NT, MN, AD conceived the study. All authors contributed to the development of the overall study methodology. MS and MN managed the ethical approval and consent processes. AD, LMB and NT provided clinical input into the algorithm design. JMR designed and built the natural language processing tools, programmed and trained the algorithm, and conducted the data analyses. AD and LMB classified the consultation records in the gold standard sets. NT and MN 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 in the study and can take responsibility for the integrity of the data and accuracy of the analysis.

Competing Interests. All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare no other support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years' no other relationships or activities that could appear to have influenced the submitted work.

Transparency declaration The lead author (NT) affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Figure legends

Figure 1: Recruitment Process

Figure 2: Herpes zoster index consultation rate by age group (bars represent 95% CIs)

Figure 3: Herpes zoster index consultation rate over time from age of 50 years, by five year age groups (n is mean patient years per year, with 95% confidence intervals and linear regression shown in blue).

Figure 4: Herpes zoster consultation rate of index case over time by gender, age adjusted (95% CIs)

Figure 5: Proportion of herpes zoster consultations with or without follow-up visits by age group showing 95% CIs.

Figure 6 Distribution of overall consultation visits occurring prior to and after a herpes zoster index case by age group

Table 1: Distribution of follow-up consultations following zoster index cases

References¶

1 Eshleman E, Shahzad A, Cohrs RJ. Varicella zoster virus latency. *Future Virol* 2011;**6**:341–55.

2 Brisson M, Edmunds WJ, Law B, *et al.* Epidemiology of varicella zoster virus infection in Canada and the United Kingdom. *Epidemiol Infect* 2001;**127**:305–14.

3 Wehrhahn MC, Dwyer DE. Herpes zoster: epidemiology, clinical features, treatment and prevention. *Australian Prescriber* 2012;**35**:143–7.

4 Schmader K, Gnann JW, Jr., *et al.* The Epidemiological, Clinical, and Pathological Rationale for the Herpes Zoster Vaccine. *J Infect Dis* 2008;**197**:S207–15.

5 Gauthier A, Breuer J, Carrington D, *et al.* Epidemiology and cost of herpes zoster and post-herpetic neuralgia in the United Kingdom. *Epidemiol Infect* 2009;**137**:38–47.

6 Yawn BP, Gilden D. The global epidemiology of herpes zoster. *Neurology* 2013;**81**:928–30.

7 Kawai K, Gebremeskel BG, Acosta CJ. Systematic review of incidence and complications of herpes zoster: towards a global perspective. *BMJ Open* 2014;**4**:e004833.

8 Pinchinat S, Cebrián-Cuenca AM, Bricout H, *et al.* Similar herpes zoster incidence across Europe: results from a systematic literature review. *BMC Infect Dis* 2013;**13**:170.

9 Yawn BP, Saddier P, Wollan PC, Sauver JLS, Kurland MJ, Sy LS. A population-based study of the incidence and complication rates of herpes zoster before zoster vaccine introduction. In: *Mayo Clinic Proceedings*. Mayo Clinic 2007. 1341–9.

10 Reid JS, Ah Wong B. Herpes zoster (shingles) at a large New Zealand general practice: incidence over 5 years. *N Z Med J* 2014;**127**:56–60.

11 Johnson RW. Herpes zoster and postherpetic neuralgia. *Expert Rev Vaccines* 2010;**9**:21–6.

12 Hales CM, Harpaz R, Joesoef MR, *et al.* Examination of links between herpes zoster incidence and childhood varicella vaccination. *Ann Intern Med* 2013;**159**:739–45.

13 Jumaan AO, Yu O, Jackson LA, *et al.* Incidence of Herpes Zoster, Before and After Varicella-Vaccination–Associated Decreases in the Incidence of Varicella, 1992–2002. *J Infect Dis* 2005;**191**:2002–7.

14 Leung J, Harpaz R, Molinari N-A, *et al.* Herpes zoster incidence among insured persons in the United States, 1993-2006: evaluation of impact of varicella vaccination. *Clin Infect Dis* 2011;**52**:332–40.

15 Toyama N, Shiraki K, Society of the Miyazaki Prefecture Dermatologists. Epidemiology of herpes zoster and its relationship to varicella in Japan: A 10-year survey of 48,388 herpes zoster cases in Miyazaki prefecture. *J Med Virol* 2009;**81**:2053–8.

16 Russell ML, Schopflocher DP, Svenson L, *et al.* Secular trends in the epidemiology of shingles in Alberta. *Epidemiol Infect* 2007;**135**:908–13.

17 Varicella and herpes zoster vaccines: WHO position paper, June 2014. *Wkly Epidemiol Rec* 2014;**89**:265–87.

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**:78.

20 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.

21 Fay MP, Feuer EJ. Confidence intervals for directly standardized rates: a method based on the gamma distribution. *Stat Med* 1997;**16**:791–801.

22 Houghton F. Degrees of Deprivation in New Zealand — An Atlas of Socioeconomic Difference: By P. Crampton, C. Salmond, R. Kirkpatrick, R. Scarborough and C. Skelly. *N Z Geog* 2001;**57**:62–62.

23 Salmond C, Crampton P, Sutton F. NZDep91: A New Zealand index of deprivation. *Aust N Z J Public Health* 1998;**22**:835–7.

24 Karki S, Newall AT, Raina MacIntyre C, *et al.* Healthcare Resource Utilisation Associated with Herpes Zoster in a Prospective Cohort of Older Australian Adults. *PLoS One* 2016;**11**:e0160446.

25 Schmader K, George LK, Burchett BM, *et al.* Racial differences in the occurrence of herpes zoster. *J Infect Dis* 1995;**171**:701–4.

26 Anderson I, Crengle S, Kamaka ML, *et al.* Indigenous health in Australia, New Zealand, and the Pacific. *Lancet* 2006;**367**:1775–85.

27 Serpell M, Gater A, Carroll S, *et al.* Burden of post-herpetic neuralgia in a sample of UK residents aged 50 years or older: findings from the Zoster Quality of Life (ZQOL) study. *Health Qual Life Outcomes* 2014;**12**:92.

28 Gater A, Abetz-Webb L, Carroll S, *et al.* Burden of herpes zoster in the UK: findings from the zoster quality of life (ZQOL) study. *BMC Infect Dis* 2014;**14**:402.

29 Stein AN, Britt H, Harrison C, *et al.* Herpes zoster burden of illness and health care resource utilisation in the Australian population aged 50 years and older. *Vaccine* 2009;**27**:520–9.

30 Hobbelen PHF, Stowe J, Amirthalingam G, *et al.* The burden of hospitalisation for varicella and herpes zoster in England from 2004 to 2013. *J Infect* 2016;**73**:241–53.

31 Enrolment in a primary health organisation. Ministry of Health NZ. http://www.health.govt.nz/our-work/primary-health-care/about-primary-health-organisations/enrolment-primary-health-organisation (accessed 27 Oct 2017).

32 Wu M-F, Yang Y-W, Lin W-Y, *et al.* Varicella zoster virus infection among healthcare workers in Taiwan: seroprevalence and predictive value of history of varicella infection. *J Hosp Infect* 2012;**80**:162–7.

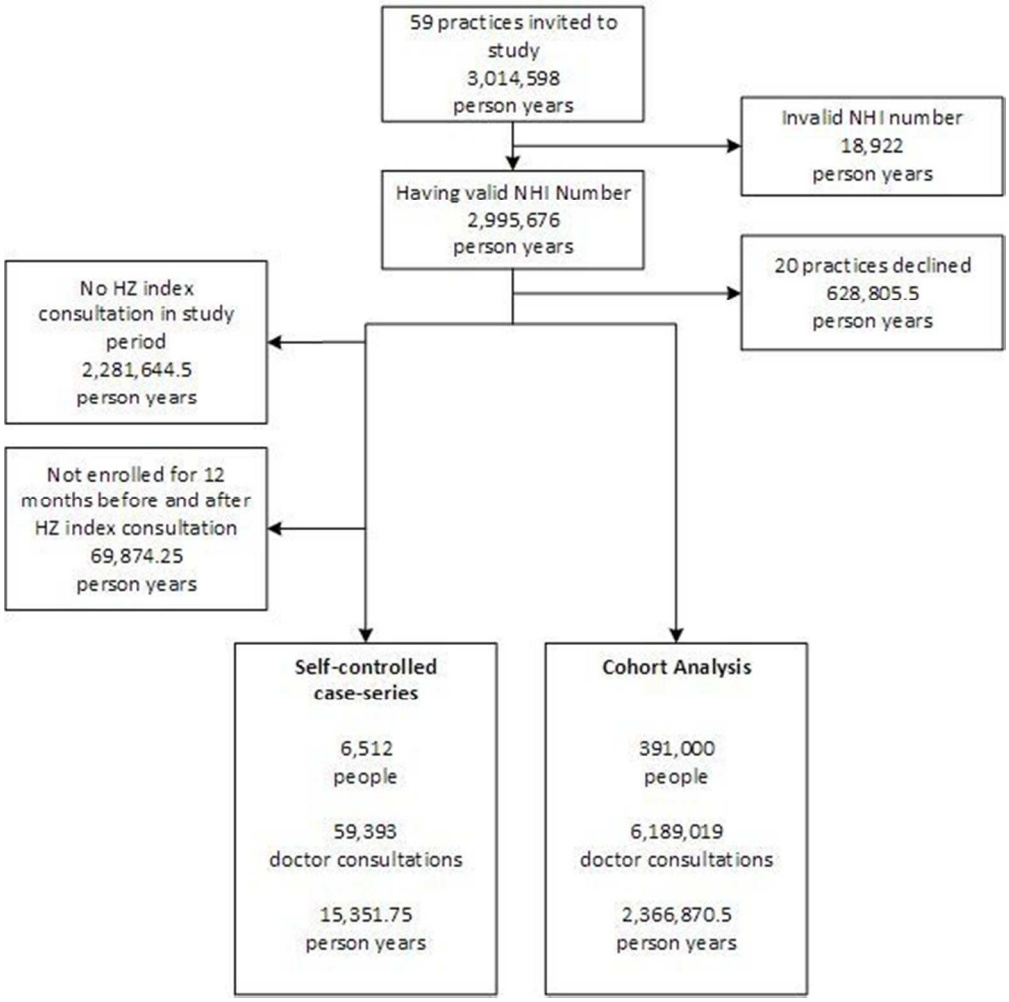

Figure 1: Recruitment Process

48x48mm (300 x 300 DPI)

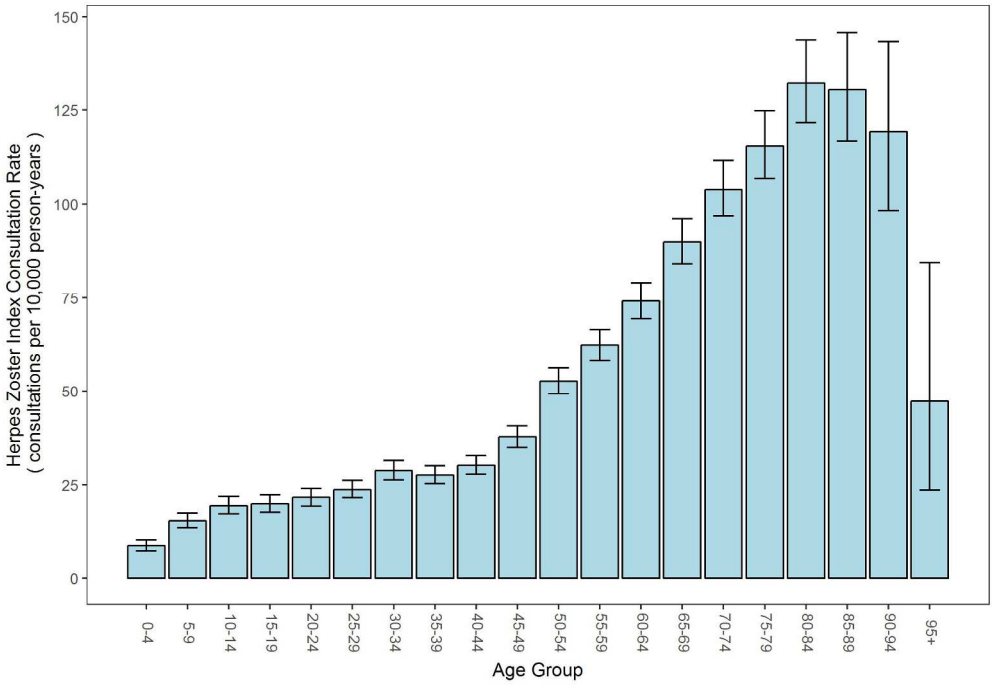

Figure 2: Herpes zoster index consultation rate by age group (bars represent 95% CIs)

210x145mm (300 x 300 DPI)

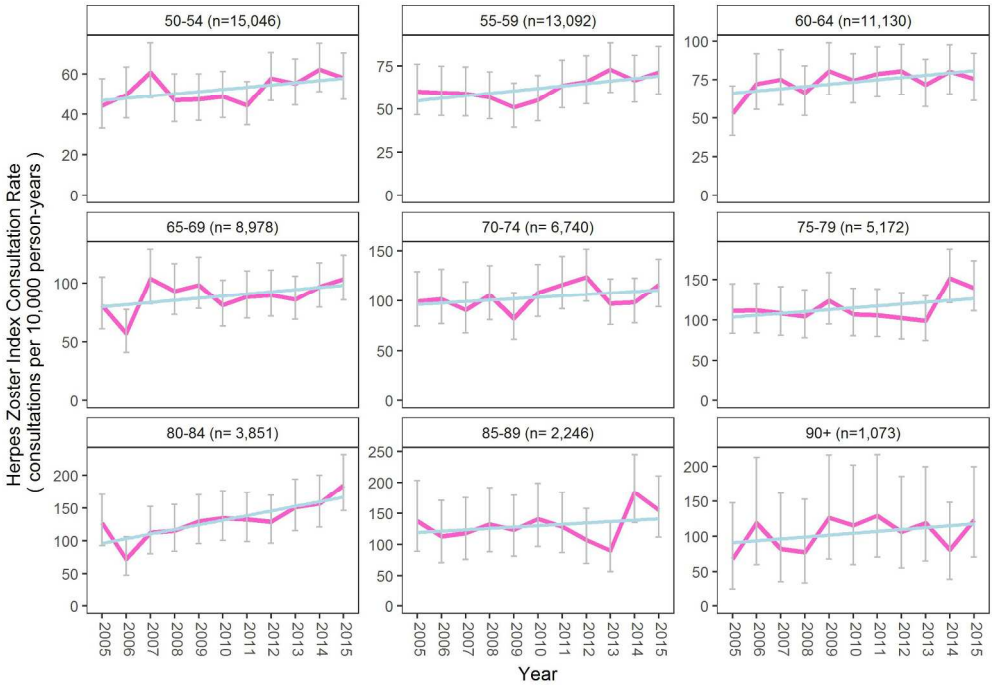

Figure 3: Herpes zoster index consultation rate over time from age of 50 years, by five year age groups (n is mean patient years per year, with 95% confidence intervals and linear regression shown in blue).

210x145mm (300 x 300 DPI)

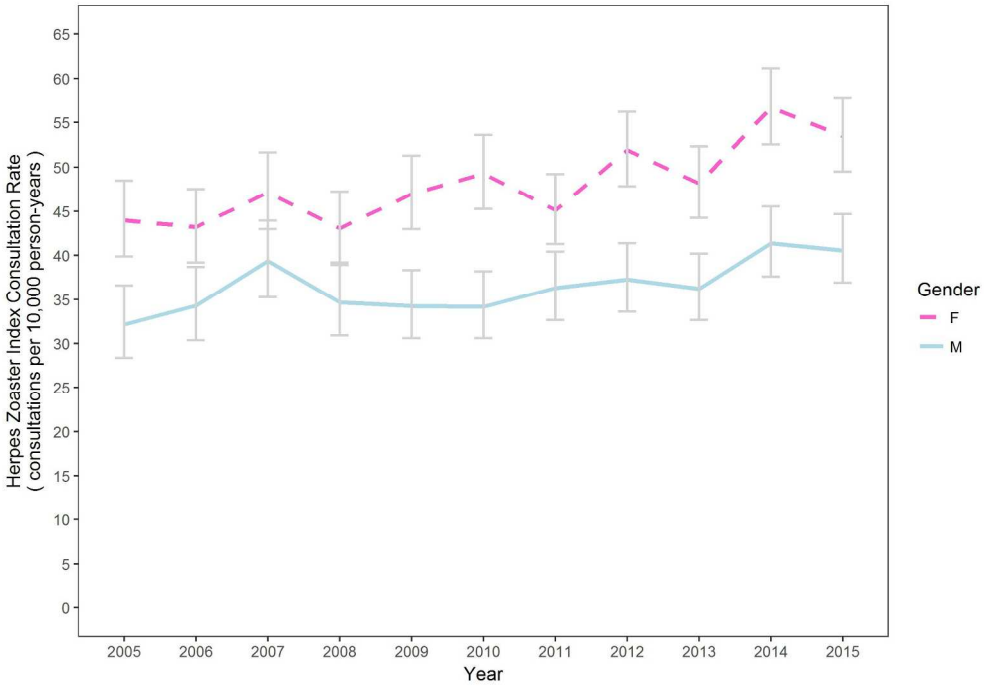

Figure 4: Herpes zoster consultation rate of index case over time by gender, age adjusted (95% CIs)

210x145mm (300 x 300 DPI)

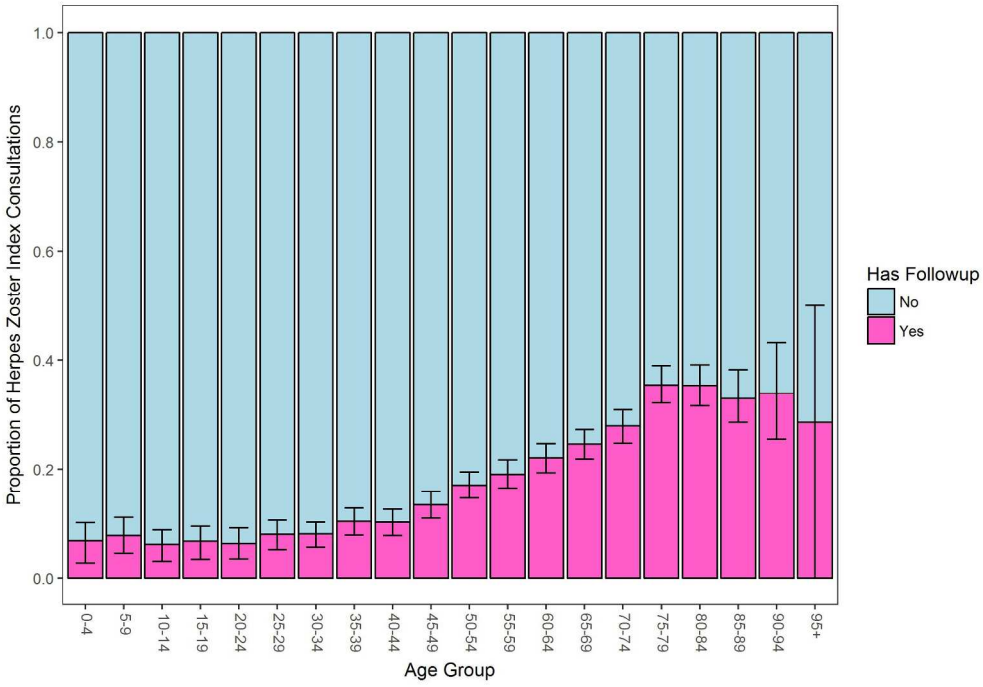

Figure 5: Proportion of herpes zoster consultations with or without follow-up visits by age group showing 95% CIs.

210x145mm (300 x 300 DPI)

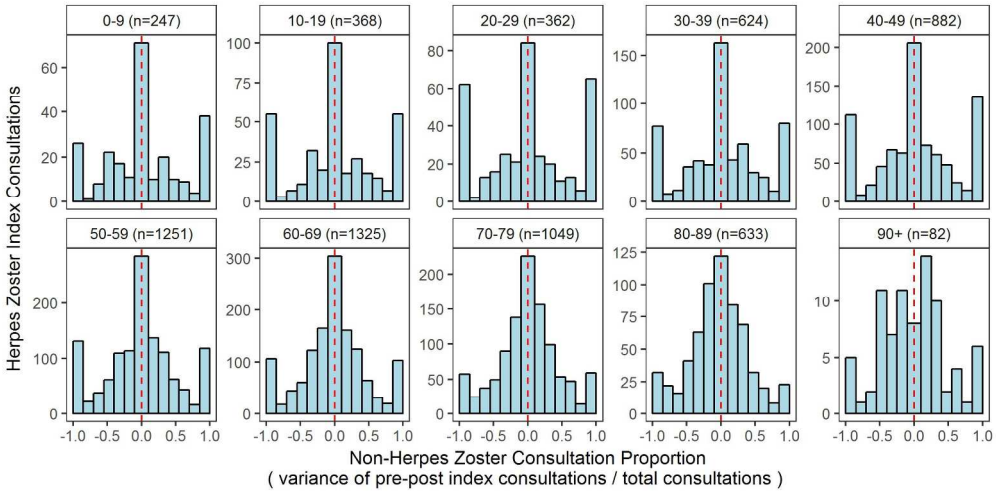

Figure 6 Distribution of overall consultation visits occurring prior to and after a herpes zoster index case by age group

211x105mm (300 x 300 DPI)

ORD statement – checklist of items, extended from the STROBE statement, that should be reported in observational stud collected health data.

	Item No.	STROBE items	Location in manuscript where items are reported	RECORD items	Location manuscri where ite reported
bstract					
	1	(a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found	Title - line 6 -7 Abstract line 10	RECORD 1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included. RECORD 1.2: If applicable, the geographic region and timeframe within which the study took place should be reported in the title or abstract. RECORD 1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.	Title - lin Abstract - Title: line Abstract - 14, line 1 Abstract - 14
on					
d	2	Explain the scientific background and rationale for the investigation being reported	page 4 paragraph 1 and 2		
	3	State specific objectives, including any prespecified hypotheses	page 4 paragraph 3 - line 36		
gn	4	Present key elements of study design early in the paper	page 4 paragraph 1 line 47-48		
	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	page 5 line 53-		
s	6	*(a) Cohort study* - Give the	page 5 line 11	RECORD 6.1: The methods of study	Flow char

		eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study* - Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study* - Give the eligibility criteria, and the sources and methods of selection of participants *(b) Cohort study* - For matched studies, give matching criteria and number of exposed and unexposed *Case-control study* - For matched studies, give matching criteria and the number of controls per case		population selection (such as codes or algorithms used to identify subjects) should be listed in detail. If this is not possible, an explanation should be provided. RECORD 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided. RECORD 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data linkage process, including the number of individuals with linked data at each stage.	figure 1 Page 5 lin Reference Page 5 lin Figure 1 - chart Page 5, li line 49
	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.		RECORD 7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided.	Process, p line 42 an reference
es/ nt	8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	flow chart Figure 1 analysis page 4 line 12, page 5 lines 50 and 54 Not applicable		
	9	Describe any efforts to address potential sources of bias	page 5 line 34		

	10	Explain how the study size was arrived at	settings and participants from page 4 line 56 to page 5 line 9 Figure 1 - flow chart		
e	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why			Analysis line 49 – line 7
	12	(a) Describe all statistical methods, including those used to control for confounding (b) Describe any methods used to examine subgroups and interactions (c) Explain how missing data were addressed (d) *Cohort study* - If applicable, explain how loss to follow-up was addressed *Case-control study* - If applicable, explain how matching of cases and controls was addressed *Cross-sectional study* - If applicable, describe analytical methods taking account of sampling strategy (e) Describe any sensitivity analyses	Page 6 line 6 Page 5 line 22 page 6 paragraph 1 page 5 line 38 page 6 line 38		
s and ethods		..	analysis page 6 line 8	RECORD 12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population.	Settings a participan page 5, li

				RECORD 12.2: Authors should provide information on the data cleaning methods used in the study.	analysis P line 8
		..		RECORD 12.3: State whether the study included person-level, institutional-level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.	Settings a participan 5 line 18
s	13	(a) Report the numbers of individuals at each stage of the study (*e.g.*, numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed) (b) Give reasons for non-participation at each stage. (c) Consider use of a flow diagram		RECORD 13.1: Describe in detail the selection of the persons included in the study (*i.e.,* study population selection) including filtering based on data quality, data availability and linkage. The selection of included persons can be described in the text and/or by means of the study flow diagram.	figure 1 - chart
e data	14	(a) Give characteristics of study participants (*e.g.*, demographic, clinical, social) and information on exposures and potential confounders (b) Indicate the number of participants with missing data for each variable of interest (c) *Cohort study* - summarise follow-up time (*e.g.*, average and total amount)			Study coh demograp page 6 lin Accuracy identifica page 6 lin
ata	15	*Cohort study* - Report numbers of outcome events or summary measures over time *Case-control study* - Report			herpes zo consultati page 6 lin

		numbers in each exposure category, or summary measures of exposure *Cross-sectional study* - Report numbers of outcome events or summary measures			healthcare utilisation from line Table 1
ts	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (e.g., 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period			Confidenc intervals f statistical analyses a presented figures. Methods, page 6 lin
yses	17	Report other analyses done—e.g., analyses of subgroups and interactions, and sensitivity analyses			Accuracy herpes zo identifica page 6 lin
	18	Summarise key results with reference to study objectives			paragraph
s	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias		RECORD 19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.	study limi page 8 lin page 8 lin
on	20	Give a cautious overall interpretation of results			page 9 lin

		considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence			
bility	21	Discuss the generalisability (external validity) of the study results			conclusio 10 line 47 line 53
rmation					
	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based			funding, p line 21
ty of aw ng		..		RECORD 22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code.	not applic

: Benchimol EI, Smeeth L, Guttmann A, Harron K, Moher D, Petersen I, Sørensen HT, von Elm E, Langan SM, the RECORD
. The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Medi*