Epidemiology and Infection

cambridge.org/hyg

A novel approach to estimate the local population denominator to calculate disease incidence for hospital-based health events in England

James Campling[1], Elizabeth Begier[2], Andrew Vyse[1], Catherine Hyams[3,4,5], Dave Heaton[6], Jo Southern[2], Adam Finn[3], Harish Madhava[2], Bradford D. Gessner[2] and Gillian Ellsbury[1]

[1]Pfizer Limited, Walton Oaks, Dorking Road, Tadworth Surrey, KT20 7NS, UK; [2]Vaccines Medical Development, Scientific and Clinical Affairs, Pfizer Inc, Collegeville, PA, USA; [3]Bristol Vaccine Centre, Schools of Cellular and Molecular Medicine and of Population Health Sciences, University of Bristol, Bristol, BS2 8AE, UK; [4]Academic Respiratory Unit, Learning and Research Building, Southmead Hospital, Bristol, BS10 5NB, UK; [5]Acute Medical Unit, Southmead Hospital, Bristol, BS10 5NB, UK and [6]Harvey Walsh, Open Health Group, Marlow, Buckinghamshire, SL7 2FF, UK

Original Paper

Cite this article: Campling J *et al* (2022). A novel approach to estimate the local population denominator to calculate disease incidence for hospital-based health events in England. *Epidemiology and Infection* **150**, e150, 1–14. https://doi.org/10.1017/S0950268822000917

Received: 18 January 2022
Revised: 17 April 2022
Accepted: 9 May 2022

Key words:
Community-acquired pneumonia; epidemiology; incidence; pneumonia

Author for correspondence:
James Campling,
E-mail: james.campling@pfizer.com

Abstract

While incidence studies based on hospitalisation counts are commonly used for public health decision-making, no standard methodology to define hospitals' catchment population exists. We conducted a review of all published community-acquired pneumonia studies in England indexed in PubMed and assessed methods for determining denominators when calculating incidence in hospital-based surveillance studies. Denominators primarily were derived from census-based population estimates of local geographic boundaries and none attempted to determine denominators based on actual hospital access patterns in the community. We describe a new approach to accurately define population denominators based on historical patient healthcare utilisation data. This offers benefits over the more established methodologies which are dependent on assumptions regarding healthcare-seeking behaviour. Our new approach may be applicable to a wide range of health conditions and provides a framework to more accurately determine hospital catchment. This should increase the accuracy of disease incidence estimates based on hospitalised events, improving information available for public health decision making and service delivery planning.

Introduction

When considering the introduction of an immunisation programme, it is paramount that the incidence of the diseases of interest is estimated as accurately as possible. Calculating annual incidence rates (expressed as the number of cases per 100 000 population) depends on the accurate estimation of two parameters: (1) the number of people diagnosed with the disease during a specified time interval, (2) the size of the population from which the cases originated at the start of the time interval of interest. Measuring each parameter has its own challenges, but here we focus on challenges associated with estimating the size of local populations within England, hereafter referred to as the denominator. For national datasets where the catchment area is determined based on clear geographic boundaries, the denominator can be estimated using census data which are maintained through annually adjusted estimates. However, many surveillance studies use health centres such as clinics and hospitals, and in these cases, the denominator population usually is not clearly defined.

To estimate healthcare facility catchment populations, a few map-based approaches have previously been proposed (e.g. defined urban conurbation area, crow-fly distance, road distance and road time access) [1–5], all of which rely on census data to provide population estimates based on where the boundary is drawn on the map from the given approach. However, in England, and for several reasons, geographically defined denominators may provide a poor estimate of the population accessing care at a particular health centre. The National Health Service (NHS) provides healthcare free of charge for all residents in England and allows patients to choose where they receive medical care, which is an important principle of the English healthcare system. Although geography plays an important role in influencing this choice, other factors may be important including public transport, parking, waiting times, traffic considerations both for patients and visiting family members, experience with a particular hospital, GP recommendation, ambulance preference, hospital capacity, specialist services and hospital reputation [6]. Moreover, while it might be expected that those who live close to a

hospital would preferentially choose that location, many people live equidistant to more than one hospital (both in terms of distance and travel time). In summary, no standardised methodology exists to estimate incidence based on the person seeking healthcare at a given facility.

In this report, we describe a novel methodology to estimate local population denominators for the Bristol AvonCAP study – a study set up with the specific aim of measuring the burden of hospitalised respiratory disease in England, to provide evidence for informed decision making for public health interventions including vaccines, that have the potential to alleviate some of this burden. The study was designed to measure the incidence of hospitalised community-acquired pneumonia (CAP) and other acute lower respiratory tract diseases (aLRTD) in two large secondary care hospitals located in Bristol. We think this methodology could be replicated for other health outcomes and other regions in England (or elsewhere if a high level of formal primary care practice registration exists), which could substantially improve disease incidence estimates and thus accurate public health decision-making.

Methods

Methodology overview

The conceptual distinction between previously proposed approaches to determine population denominators and our methodology is that the former are based on assumptions about which hospitals patients are expected to use. Our new methodology attempts to minimise the use of assumptions by utilising multiple data sources to assess which hospitals these populations have used in the past.

The NHS in England allocates an annual budget to local geographically defined clinical commissioning groups (CCGs) broadly based on population numbers and utilisation in prior years. In April 2021, there were 106 CCGs across England and their boundaries were drawn to complement local healthcare resources [7]. See the Method step 1 section for an important organisational change for the NHS.

Robust systems are used by CCGs to reimburse hospital care, therefore we hypothesised that CCG geographical regions may be helpful in determining hospital catchment areas and local populations. To test our hypothesis, we utilised Hospital Episode Statistics (HES) data which were re-used with the permission of NHS Digital via Harvey Walsh Limited. aLRTD admissions at the study hospitals between April 2017–March 2020 were linked to aggregated general practitioner (GP) data to understand from which CCG the hospitals' patients came (Methods Part 1). Then, we estimated the proportion of patients hospitalised at the study hospitals among all patients hospitalised with LRTD for each practice and multiplied that by count of patients registered at that GP practice to calculate the Bristol hospital catchment population (Methods Part 2).

In England, all hospitalisations in NHS hospitals are captured in HES and all acute care is provided by NHS hospitals. HES contains information on bed days, length of admission, outpatient appointments, attendances at Accident and Emergency Departments at NHS hospitals in England, discharge diagnoses and hospital death [8]. The primary diagnosis and other clinical conditions are specified using the tenth revision of the International Classification of Diseases version 10 (ICD-10) [9]. Furthermore, in England a high proportion of the population are registered with General Practice where it is not possible to be registered at two practices concurrently [10, 11].

Method step 1 – defining GP practices associated with patients treated at study hospitals

To understand from where patients treated at the study hospitals originated (i.e. to which CCG the patients' GP practices belong), HES data were extracted for all adult patients coded for aLRTD between April 2017–March 2020 and filtered to include only patients treated at the study hospitals: North Bristol NHS Trust (NBT), and University Hospitals Bristol NHS Foundation Trust & Weston NHS Foundation Trust (UHBW). Finally, data were analysed to determine in which CCG area the patients lived based on their GP registration. There are 6 CCG regions in the South West of England within a 1-hour drive of the study hospitals, as illustrated in Figure 1.

Fig. 1 shows a map of the CCGs described in the results pie chart (Fig. 2) along with the location of relevant hospitals. In July 2022 NHS England establised 42 integrated care systems (ICS) and as a consequence CCGs were closed down and new statutory organisations called integrated care boards (ICB) were introduced. The remit of an ICB includes managing the NHS budget and arranging for the provision of health services in the ICS area. The boundaries of the new ICSs in the south-west of England remain unchanged from the previous CCG boundaries and therefore this change does not impact this analysis (https://www.england.nhs.uk/integratedcare/).

Method step 2 – defining the catchment population of study hospitals

As patients registered in the CCG might seek care at a different hospital for a variety of reasons, we could not assume every patient registered with a GP in the Bristol, North Somerset and South Gloucestershire (BNSSG) CCG used the study hospitals. Therefore, we estimated the proportion of patients from each GP practice treated at the study hospitals among all BNSSG CCG patients, stratified by age group. This proportion was used to calculate the study hospitals' catchment population. All aLRTD hospitalisations (based on ICD-10 codes; Appendix 1) occurring between April 2017 – March 2020 among patients registered in the BNSSG CCG were analysed by GP practice. For each GP practice, the per cent of hospitalisations occurring at study hospitals was calculated within each age-group (18–34, 35–49, 50–64, 65–74, 75–84 and ⩾85 years). The percentage of hospitalisations occurring at study hospitals was the number of patients at each GP practice who were admitted for aLRTD at study hospitals (study hospital aLRTD patients) divided by the total number of patients at that GP practice who were hospitalised for aLRTD at any English hospital in the time period (overall aLRTD inpatients). This proportion (i.e. per cent of aLRTD inpatients using study hospitals) was multiplied by the practice population for each GP practice by age strata to provide an expected Bristol hospital catchment population contribution for each GP practice (once all age groups summed). GP populations were obtained from NHS Digital 'Patients Registered at a GP Practice' data for October 2019. Finally, the catchment population contribution for each GP practice in the BNSSG CCG was combined to provide an expected total Bristol hospital catchment population. In summary, if:

- E = Calculated catchment population
- SHP = Number of patients at a GP practice hospitalised at a study hospital with aLRTD during 2017–2019
- OL = Overall number of patients at a GP practice hospitalised in England with aLRTD during 2017–2019
- POP = Local GP population
- i = Each individual practice

https://doi.org/10.1017/S0950268822000917 Published online by Cambridge University Press

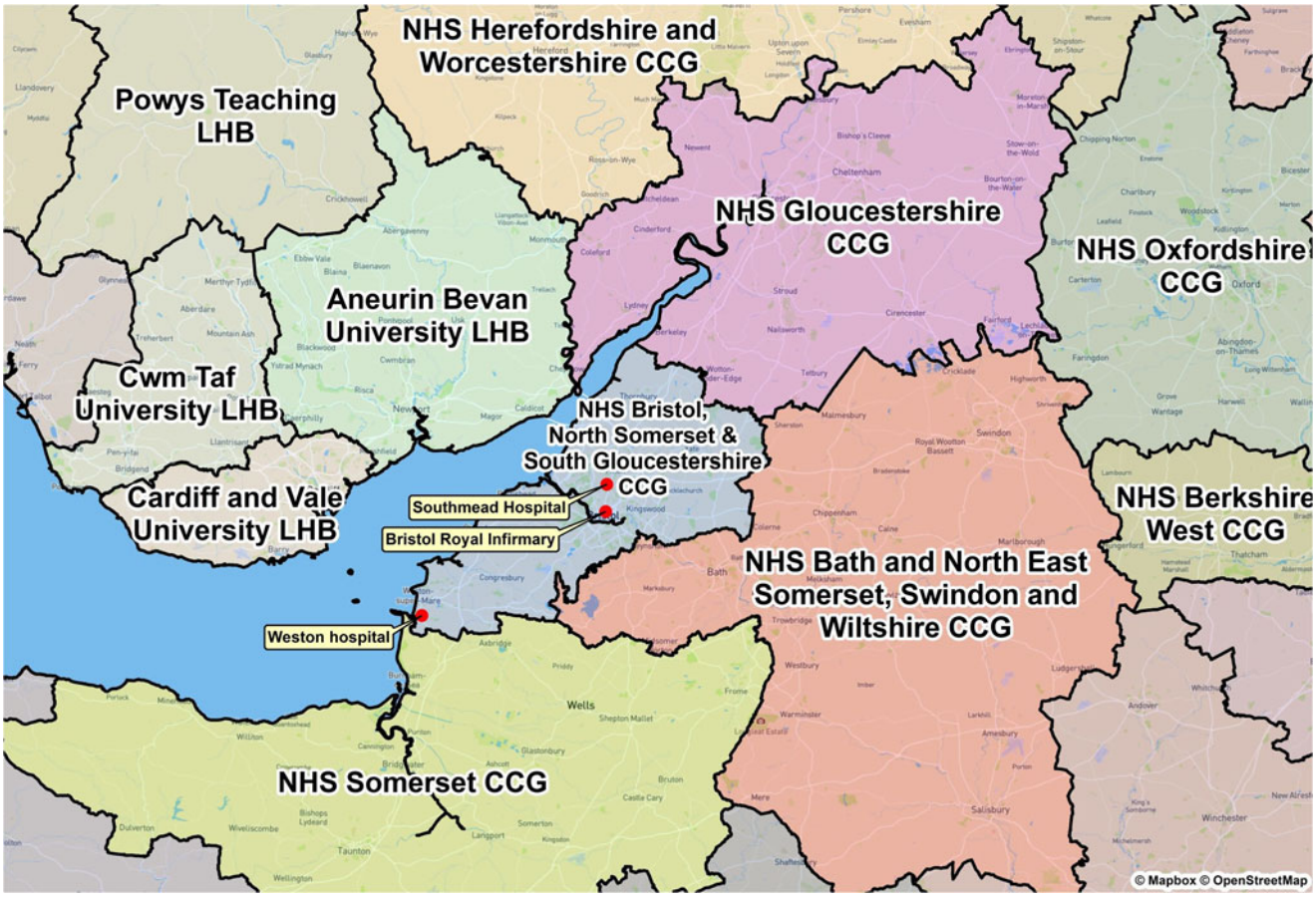

Fig. 1. South West England clinical commissioning groups map.

Then:

$$E = \sum \left(\left(\frac{\mathrm{SHP}_i}{\mathrm{OL}_i} \right) \mathrm{POP}_i \right)$$

Drive-time methodology

The BNSSG CCG used a 20-minute drive-time for their healthcare utilisation mapping purposes [12]. We have included this alternative methodological approach to allow comparison between our methodology and other methodologies in current use. We obtained data from the BNSSG CCG which divides the CCG region into small geographical areas used by the UK census known as lower layer super output areas (LSOA). LSOAs have a population of between 1000–3000 people or 400–1200 households [13]. Data were filtered according to estimated drive-time from each LSOA to the study hospitals according to the Automobile Association (AA) route planner, (AA, Hampshire, UK) [14]. UK population data by LSOA for all ages (0 – ⩾90 years) were downloaded from the UK Office of National Statistics census website. Population estimates were derived for the following drive-times from the study hospitals 20, 25, 30, 40 and 60 minutes by matching the LSOA population data with the drive-time data.

Results

In 2019, there were 82 GP practices in the BNSSG CCG. Figure 2 shows the proportion of patients that attended the study hospitals in 2019 that were registered at GP practices in both the BNSSG CCG as well as six other CCGs that, combined, represented where >99% of patients hospitalised at study hospitals were registered. The majority of hospitalised patients (96%) were registered at BNSSG CCG GP practices, with most of the remaining 4% based in the surrounding CCGs.

Substantial variability existed by GP practice in the per cent of all persons hospitalised for aLRTD who were hospitalised at a study hospital with much less variability by age (Fig. 3) (based on a representative sample of 10 anonymised GP practices within the BNSSG CCG). Lower proportions were reported for GP practices that were located either close to the CCG boundary or close to Weston hospital (a non-study hospital situated in the BNSSG CCG). Full tables reporting these data for all GP practices located in the BNSSG CCG for 2017, 2018, 2019 and the combined data can be found in Appendix 2.

The degree to which the estimates from our methodology compared to estimates produced by other methods varied, including within specific age groups (Table 1 and Fig. 4). The total CCG population (the sum of the population of all GP practices in the CCG) overestimated the catchment population compared to our estimates by 15% to 24%. By contrast, the population living within a 20 minute drive of the study hospitals underestimated the catchment population by 10% to 29%. As drive-time increased linearly, the estimated population increased non-linearly such that the population based on a 60 minute drive-time overestimated the catchment population by 276% to 428%. The degree of underestimation or overestimation from other methods did not vary substantially by age group.

The map in Fig. 5 shows the location of the study hospitals and Weston General Hospital. The BNSSG CCG boundary is shown in

https://doi.org/10.1017/S0950268822000917 Published online by Cambridge University Press

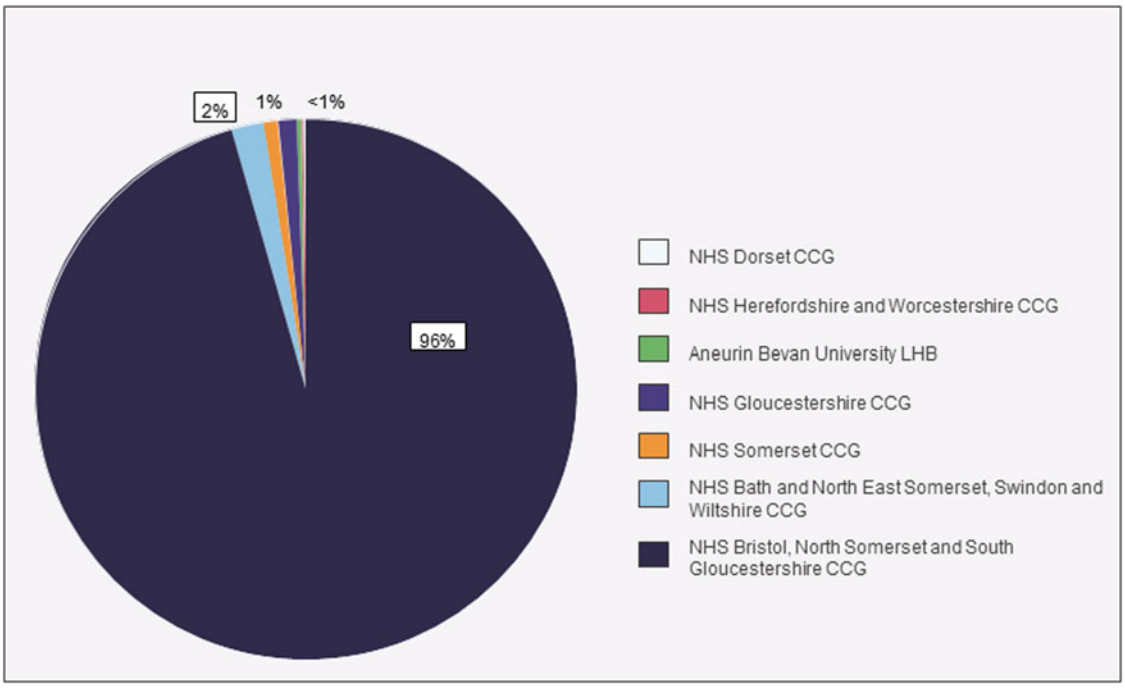

Fig. 2. 2017–2019 study hospital admissions by clinical commissioning group of the patients' GP practices.

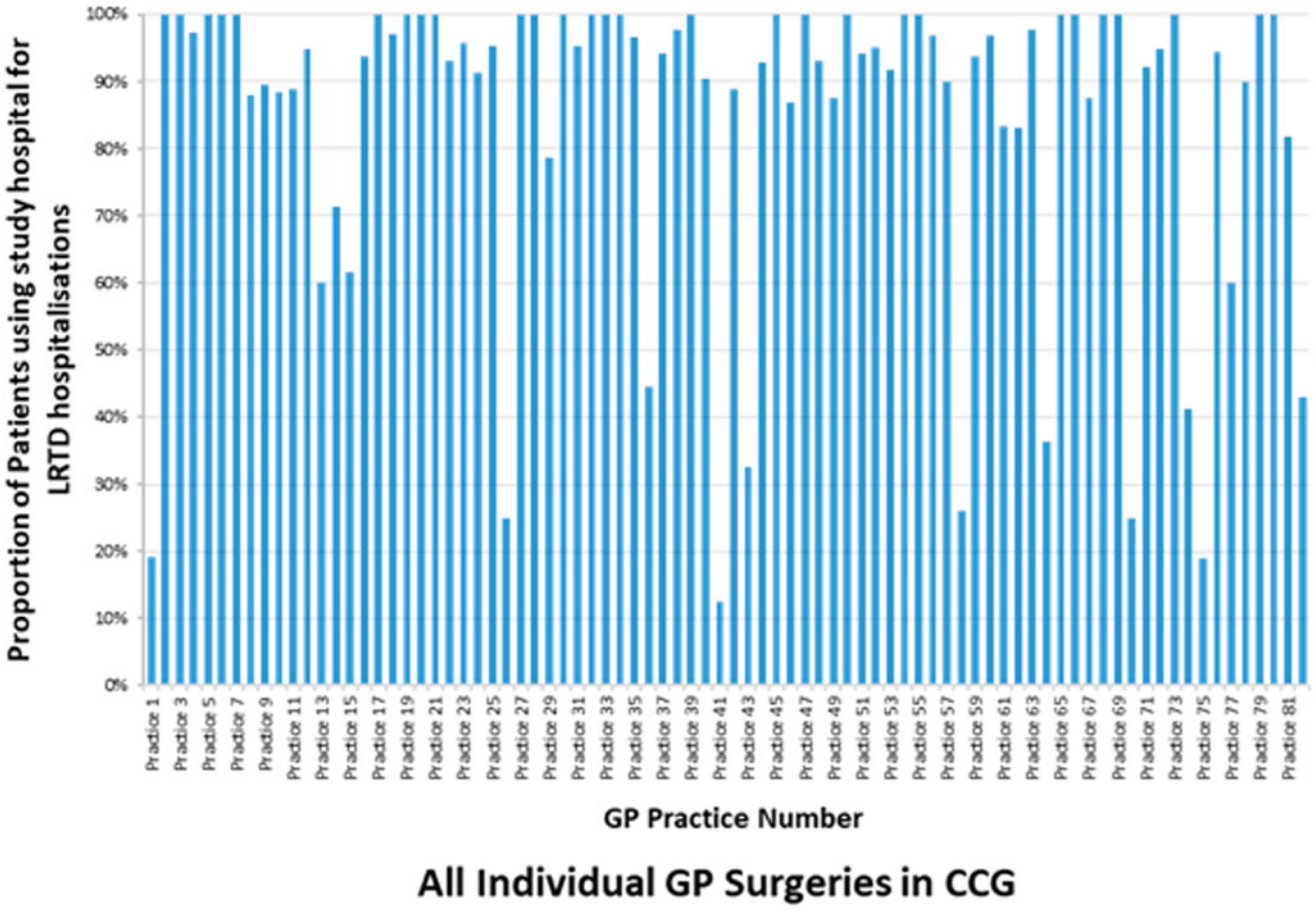

Fig. 3. A bar chart showing the proportion of persons hospitalised for acute lower respiratory tract disease who were hospitalised at a study hospital, stratified by individual anonymised general practice and patient age group.

https://doi.org/10.1017/S0950268822000917 Published online by Cambridge University Press

Table 1. Comparison of study hospital catchment population estimates based on different approaches

Age group	Estimated catchment (Study method)	Total CCG catchment	Estimated based on ⩽20 min drive-time	Estimated based on ⩽25 min drive-time	Estimated based on ⩽30 min drive-time	Estimated based on <40 min drive-time	Estimated based on <60 min drive-time
Five adult age groupings							
18–34	231 342	268 093 (↑16%)	208 924 (↓10%)	238 301 (↑3%)	295 130 (↑28%)	442 590 (↑91%)	870 841 (↑276%)
35–49	184 269	211 568 (↑15%)	130 881 (↓29%)	162 469 (↓12%)	211 452 (↑15%)	337 781 (↑83%)	714 415 (↑288%)
50–64	152 380	178 970 (↑17%)	108 404 (↓29%)	143 508 (↓6%)	196 307 (↑29%)	331 795 (↑118%)	732 702 (↑381%)
65–74	74 245	89 015 (↑20%)	52 954 (↓29%)	73 368 (↓1%)	102 148 (↑38%)	175 757 (↑137%)	391 718 (↑428%)
75–84	45 989	55 720 (↑21%)	33 712 (↓27%)	46 919 (↑2%)	65 244 (↑42%)	111 109 (↑142%)	239 310 (↑420%)
85+	19 229	23 938 (↑24%)	15 280 (↓21%)	20 400 (↑6%)	28 261 (↑47%)	47 108 (↑145%)	99 865 (↑419%)
Two adult age groupings							
18–64	567 991	658 631 (↑16%)	448 209 (↓21%)	544 278 (↓4%)	702 889 (↑24%)	1 112 166 (↑96%)	2 317 958 (↑308%)
⩾65	139 463	168 673 (↑21%)	101 946 (↓27%)	140 687 (↑1%)	195 653 (↑40%)	333 974 (↑139%)	730 893 (↑424%)
Total	707 454	827 304 (↑17%)	550 155 (↓22%)	684 965 (↓3%)	898 542 (↑27%)	1 446 140 (↑104%)	3 048 851 (↑331%)

black and travel time boundaries are identified by colour to the study hospitals based on the shortest travel time to either study hospital.

Discussion

Incidence studies based on counts of hospitalisations from one or a few study hospitals are common, but there is no standard methodology to define a health centre's catchment population for the purpose of accurately estimating incidence denominators. Traditional geography-based approaches (such as defining a population with a certain drive-time to a study health centre) that rely on census data do not account for the nuanced ways in which populations access healthcare and therefore are prone to error. We devised a novel approach for establishing local population estimates in England to support disease incidence studies conducted at single or multiple hospital sites. This approach was made possible because nearly everyone in England is registered with a GP and because of the comprehensive healthcare data captured by NHS Digital [15]. Moreover, a strength of our approach is that it is uses healthcare utilisation data to calculate specific study hospital usage by GP centre and age group and makes no assumptions about which health centres are used by a population within a particular census area.

Depending on the precise method, the geography-based approaches assessed in our study would have overestimated or underestimated the true catchment population and thus either underestimated or overestimated aLRTD incidence. At the extreme, defining the catchment population as those people living within a 60 minute drive from a study hospital would have overestimated the catchment population by 4-fold to 5-fold and thus underestimated incidence to the same degree. At the other extreme, a drive-time of 20 minutes would have underestimated denominators by 20–25% and thus overestimated incidence. Alternatively, the use of the entire CCG population would have overestimated denominators by 15%. The differences between geographically estimated denominators and our method are likely to vary by location and thus, the specific results from our study are illustrative of the principle and cannot be used to make conclusions about the relative accuracy of using an entire CCG population or drive-time for other areas. For example, higher density areas with a larger number of hospitals would decrease the accuracy of drive-time or CCG for defining the catchment area of any particular hospital. This was illustrated in our study by demonstrating that for some practices and age groups, less than 20% of the practice population with an aLRTD hospitalisation presented to a study hospital. Since the only way to document the distortion in catchment population estimate for any particular health centres inherent in traditional estimates would be to first employ the methods described here, we suggest a better approach is simply to use our methods, or some similar approach, to define incidence denominators.

Other issues must be considered when using our approach. For example, the percentage of people with aLRTD hospitalisation who were hospitalised in a study hospital was relatively stable for older age groups and larger practices but varied substantially for younger populations and smaller practices, predominantly because of small absolute case counts for the latter groups. We largely overcame this issue by combining data for multiple years and creating larger age bands for younger populations. This issue will be more problematic for rarer diseases, which may require even larger age bands, greater numbers of study years, or aggregating individual ICD-10 codes into a common outcome.

https://doi.org/10.1017/S0950268822000917 Published online by Cambridge University Press

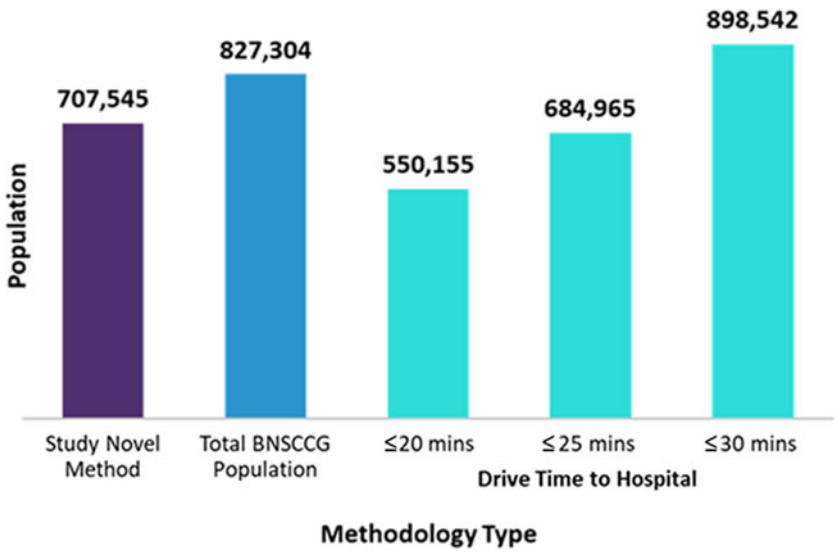

Fig. 4. Comparison of study hospital population size (⩾18yrs) by methodology.

The AvonCAP study was designed primarily to inform decisions on respiratory vaccine use among older adults, including vaccines to prevent the pneumococcal, respiratory syncytial virus, and SARS-CoV-2 infection. Policymakers, including vaccine technical committees, have consistently indicated that disease burden is the number one factor in setting priorities for vaccines [16, 17]. Disease incidence, and usually severe disease incidence using hospitalisation as a proxy, is the cornerstone of disease burden and usually is the key outcome driving cost-effectiveness models. Cost-effectiveness values in turn are often used for policy and pricing decisions. For example in England, a vaccine must be below a threshold of £ 30 000 per Quality Adjusted Life Year

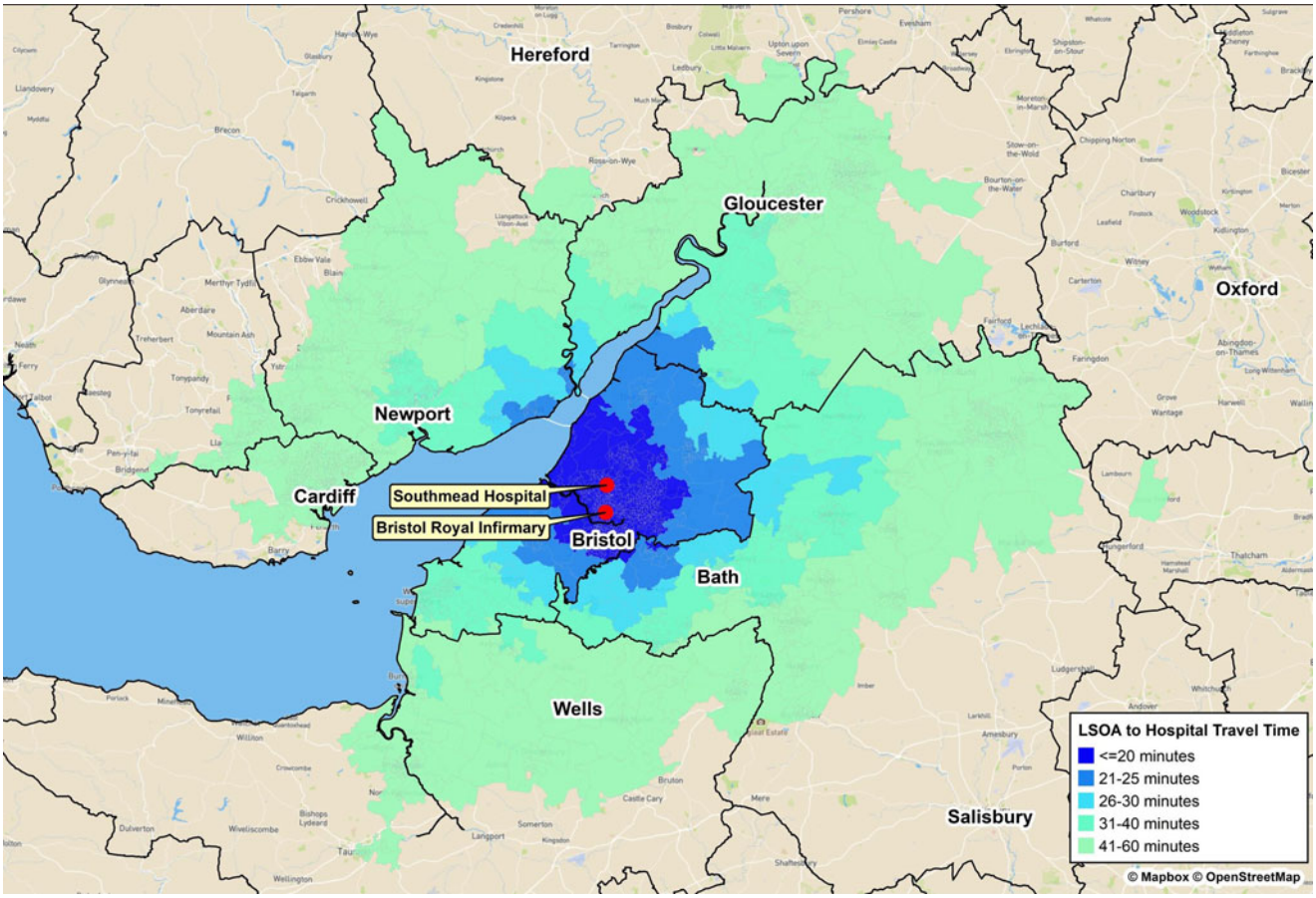

Fig. 5. Map showing travel time by car to study hospitals

https://doi.org/10.1017/S0950268822000917 Published online by Cambridge University Press

(QALY) saved to meet the criteria to be recommended for a national immunisation programme [8]. Since disease incidence underlies all these downstream measures, its accurate determination is critical for policy decisions. This requires a focus not just on the accurate determination of case counts (that is, numerators) but also the catchment population for the surveillance system (that is, denominators).

Our approach has a few limitations. We could not account for people who were not registered with a GP; although, nearly all English residents are registered [10]. Our methodology also did not include the 4% of people that use the study hospitals but are registered with a GP practice outside of the CCG. However, this will be largely addressed in Avon-CAP by excluding from incidence calculations patients with a study outcome living outside the CCG. Our approach requires a new estimate to be calculated for each disease of interest because some conditions will be disproportionately observed in some hospitals due to therapy area specialism. As discussed above, our approach may not be suitable for rare diseases or surveillance systems with small populations. Lastly, our methodology is appropriate for the particular circumstances of England and remains so with the recent transition to the ICS structure. The extent to which this approach can be generalised to other countries will need to be evaluated on a case-by-case basis, but other areas where nearly all persons are formally registered with a primary care provider could consider its use.

We will use the described methodology to define denominators for incidence calculations within the AvonCAP study, which in turn should contribute to providing better data for informing decisions related to adult respiratory vaccine use. A similar approach could be used to refine previous estimates where these are being used to inform respiratory disease vaccine decision making. A historical study reporting disease incidence of hospitalised pneumonia in England was conducted in Hull and the East Riding of Yorkshire [5]. This study included 8 hospitals in the region and a geography-based approach was used to define the denominator. Whilst an effort was made to specifically exclude defined postcode areas reflecting a geographic region unlikely to use the study hospitals the accuracy of the denominator used in this study remains uncertain. A more recent study published hospitalised CAP incidence estimates from Nottingham, England and used a denominator based on the entire population of the Greater Nottingham area, but the market share of the two study hospitals used was not formally defined [3, 18]. Since the Greater Nottingham area is surrounded by other urban areas with hospitals that also treat CAP, it is unclear how well Greater Nottingham census data matches the hospital catchment population, and this could be formally evaluated by replicating our methodology. More generally, the method we describe may be used for other disease incidence calculations and for relatively common diseases could be extended to focus on specific groups such as those with underlying comorbidities. While the approach we describe takes considerably more human and financial resources than using census data (through commissioning a specialist vendor that holds an appropriate license to analyse the data), this cost is negligible compared to the inefficiencies introduced when inaccurate disease incidence estimates are used as a core basis for public health decision making.

Conclusion

Use of the entire CCG or drive-times does not account for the nuanced ways that populations access healthcare and may overestimate or underestimate denominators and distort incidence estimates. Our data-driven method provides more accurate incidence estimates and thus can improve public health decision-making. Denominators for hospital-based incidence studies should be based on healthcare usage rather than geographical boundaries.

Author contributions. JC, EB, AV, DH & GE contributed to the initial design of the methodology. All authors contributed to the analysis, interpretation, and discussion of the results. We would like to acknowledge the assistance of Qi Yan, PhD (Pfizer, Inc.) who provided indispensable medical writing and literature review support for this manuscript and Harvey Walsh, Open Health Group who performed the denominator calculation. HES Data were re-used with the permission of NHS Digital via Harvey Walsh, Open Health Group.

Conflict of interest. JC, EB, AV, JS, HM, BG & GE are employees of Pfizer Vaccines and hold stock or stock options. DH is an employee of Harvey Walsh Ltd. CH is the Principal Investigator of the Avon CAP study which is an investigator-led University of Bristol study funded by Pfizer and has previously received support from the NIHR in an Academic Clinical Fellowship. AF is a member of the Joint Committee on Vaccination and Immunization (JCVI) and chair of the World Health Organization European Technical Advisory Group of Experts on Immunization (ETAGE) committee. In addition to receiving funding from Pfizer as Chief Investigator of the Avon CAP study, he leads another project investigating the transmission of respiratory bacteria in families jointly funded by Pfizer and the Gates Foundation.

Data availability statement. The data that support the findings of this study are available from Harvey Walsh, Open Group. Restrictions apply to the availability of these data, which were used under licence for this study. Data are available from the authors with the permission of Harvey Walsh, Open Group.

Disclosure. This study was conducted as a collaboration between the University of Bristol, Pfizer and Open Health Group. Pfizer is the study sponsor.

References

1. **Schuurman N *et al.*** (2006) Defining rational hospital catchments for non-urban areas based on travel-time. *International Journal of Health Geographics* **5**, 43.
2. **Roderick P *et al.*** (1999) What determines geographical variation in rates of acceptance onto renal replacement therapy in England? *Journal of Health Services Research & Policy* **4**, 139–146.
3. **Pick H *et al.*** (2020) Pneumococcal serotype trends, surveillance and risk factors in UK adult pneumonia, 2013–18. *Thorax* **75**, 38–49.
4. **Clarke JM, Barahona M and Darzi AW** (2019) Defining hospital catchment areas using multiscale community detection: a case study for planned orthopaedic care in England. *bioRxiv*, 619692. https://doi.org/10.1101/619692.
5. **Elston JW *et al.*** (2012) Increasing incidence of invasive pneumococcal disease and pneumonia despite improved vaccination uptake: surveillance in Hull and East Yorkshire, UK, 2002–2009. *Epidemiology & Infection* **140**, 1252–1266.
6. **Dealey C** (2005) The factors that influence patients' choice of hospital and treatment. *British Journal of Nursing* **14**, 576–579.
7. **Anon** (2021) NHS Clinical Commissioning Group (CCG) reorganisation as of 1 April 2021. NHS Business Services Authority. In.
8. **Joint Committee on Vaccination and Immunisation** (2013) Code of practice June 2013. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/224864/JCVI_Code_of_Practice_revision_2013_-_final.pdf. In.
9. **World Health Organization** (2004) *ICD-10: International Statistical Classification of Diseases and Related Health Problems: Tenth Revision in*, 2nd Edn. Geneva: World Health Organization. https://apps.who.int/iris/handle/10665/42980 (Accessed September 2021).
10. **Baker C** (2016) Population estimates & GP registers: why the difference? https://commonslibrary.parliament.uk/population-estimates-gp-registers-why-the-difference/#_ftn1. In.

11. **UK Government** (2019) Petition; Allow students to register at two GPs. https://petition.parliament.uk/archived/petitions/229573#:~:text=It%20is%20not%20possible%20for,risks%20in%20maintaining%20duplicate%20records. In.
12. **Personal Communication with BNSSG CCG** (2020) Drive-time used for healthcare utilisation planning. In.
13. **The National Archives; Office for National Statistics** (2016) Home > Guidance and Methodology > Geography > A beginner's guide to UK geography > Census Geography > Super Output Area (SOA). https://webarchive.nationalarchives.gov.uk/ukgwa/20160106001702/http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/census/super-output-areas--soas-/index.html. In.
14. **The AA Route Planner**. https://www.theaa.com/route-planner/traffic-news/index.jsp. In.
15. **NHS Digital** (2021) https://digital.nhs.uk/data. In.
16. **Munira SL and Fritzen SA** (2007) What influences government adoption of vaccines in developing countries? A policy process analysis. *Social Science & Medicine* **65**, 1751–1764.
17. **Bryson M *et al.*** (2010) A global look at national immunization technical advisory groups. *Vaccine* **28**, A13–A17.
18. **Rodrigo C *et al.*** (2015) Impact of infant 13-valent pneumococcal conjugate vaccine on serotypes in adult pneumonia. *European Respiratory Journal* **45**, 1632–1641.

Appendix 1: ICD-10 codes used for the analysis

Appendix 1

ICD-10 Code	ICD-10 Description
I110	Hypertensive heart disease with (congestive) heart failure
I130	Hypertensive heart and renal disease with (congestive) heart failure
I132	Hypertensive heart and renal disease with both (congestive) heart failure and renal failure
I50	Heart failure
I500	Congestive heart failure
I501	Left ventricular failure
I509	Heart failure, unspecified
J09	Influenza due to identified avian influenza virus
J09X	Influenza due to identified zoonotic or pandemic influenza virus
J10	Influenza due to identified seasonal influenza virus
J100	Influenza with pneumonia, seasonal influenza virus identified
J101	Influenza with other respiratory manifestations, seasonal influenza virus identified
J108	Influenza with other manifestations, seasonal influenza virus identified
J11	Influenza, virus not identified
J110	Influenza with pneumonia, virus not identified
J111	Influenza with other respiratory manifestations, virus not identified
J118	Influenza with other manifestations, virus not identified
J12	Viral pneumonia, not elsewhere classified
J120	Adenoviral pneumonia
J121	Respiratory syncytial virus pneumonia
J122	Parainfluenza virus pneumonia
J123	Human metapneumovirus pneumonia
J128	Other viral pneumonia
J129	Viral pneumonia, unspecified
J13	Pneumonia due to Streptococcus pneumoniae
J13X	Pneumonia due to Streptococcus pneumoniae
J14	Pneumonia due to Haemophilus influenzae
J14X	Pneumonia due to Haemophilus influenzae
J15	Bacterial pneumonia, not elsewhere classified
J150	Pneumonia due to Klebsiella pneumoniae
J151	Pneumonia due to Pseudomonas

(*Continued*)

https://doi.org/10.1017/S0950268822000917 Published online by Cambridge University Press

Appendix 1 (*Continued.*)

ICD-10 Code	ICD-10 Description
J152	Pneumonia due to staphylococcus
J153	Pneumonia due to streptococcus, group B
J154	Pneumonia due to other streptococci
J155	Pneumonia due to *Escherichia coli*
J156	Pneumonia due to other Gram-negative bacteria
J157	Pneumonia due to Mycoplasma pneumoniae
J158	Other bacterial pneumonia
J159	Bacterial pneumonia, unspecified
J16	Pneumonia due to other infectious organisms, not elsewhere classified
J160	Chlamydial pneumonia
J168	Pneumonia due to other specified infectious organisms
J17	Pneumonia in diseases classified elsewhere
J170	Pneumonia in bacterial diseases classified elsewhere
J171	Pneumonia in viral diseases classified elsewhere
J172	Pneumonia in mycoses
J173	Pneumonia in parasitic diseases
J178	Pneumonia in other diseases classified elsewhere
J18	Pneumonia, organism unspecified
J180	Bronchopneumonia, unspecified
J181	Lobar pneumonia, unspecified
J182	Hypostatic pneumonia, unspecified
J188	Other pneumonia, organism unspecified
J189	Pneumonia, unspecified
J20	Acute bronchitis
J200	Acute bronchitis due to Mycoplasma pneumoniae
J201	Acute bronchitis due to Haemophilus influenzae
J202	Acute bronchitis due to streptococcus
J203	Acute bronchitis due to coxsackievirus
J204	Acute bronchitis due to parainfluenza virus
J205	Acute bronchitis due to respiratory syncytial virus
J206	Acute bronchitis due to rhinovirus
J207	Acute bronchitis due to echovirus
J208	Acute bronchitis due to other specified organisms
J209	Acute bronchitis, unspecified
J21	Acute bronchiolitis
J210	Acute bronchiolitis due to respiratory syncytial virus
J211	Acute bronchiolitis due to human metapneumovirus
J218	Acute bronchiolitis due to other specified organisms
J219	Acute bronchiolitis, unspecified
J22	Unspecified acute lower respiratory infection
J22X	Unspecified acute lower respiratory infection
J40	Bronchitis, not specified as acute or chronic
J40X	Bronchitis, not specified as acute or chronic

(*Continued*)

https://doi.org/10.1017/S0950268822000917 Published online by Cambridge University Press

Appendix 1 (*Continued.*)

ICD-10 Code	ICD-10 Description
J41	Simple and mucopurulent chronic bronchitis
J410	Simple chronic bronchitis
J411	Mucopurulent chronic bronchitis
J418	Mixed simple and mucopurulent chronic bronchitis
J42	Unspecified chronic bronchitis
J42X	Unspecified chronic bronchitis
J43	Emphysema
J430	MacLeod syndrome
J431	Panlobular emphysema
J432	Centrilobular emphysema
J438	Other emphysema
J439	Emphysema, unspecified
J44	Other chronic obstructive pulmonary disease
J440	Chronic obstructive pulmonary disease with acute lower respiratory infection
J441	Chronic obstructive pulmonary disease with acute exacerbation, unspecified
J448	Other specified chronic obstructive pulmonary disease
J449	Chronic obstructive pulmonary disease, unspecified
J45	Asthma
J450	Predominantly allergic asthma
J451	Nonallergic asthma
J458	Mixed asthma
J459	Asthma, unspecified
J46	Status asthmaticus
J46X	Status asthmaticus
J47	Bronchiectasis
J47X	Bronchiectasis
J85	Abscess of lung and mediastinum
J850	Gangrene and necrosis of lung
J851	Abscess of lung with pneumonia
J852	Abscess of lung without pneumonia
J853	Abscess of mediastinum
J86	Pyothorax
J860	Pyothorax with fistula
J869	Pyothorax without fistula
J90	Pleural effusion, not elsewhere classified
J90X	Pleural effusion, not elsewhere classified
J91	Pleural effusion in conditions classified elsewhere
J91X	Pleural effusion in conditions classified elsewhere
J95	Postprocedural respiratory disorders, not elsewhere classified
J950	Tracheostomy malfunction
J951	Acute pulmonary insufficiency following thoracic surgery
J952	Acute pulmonary insufficiency following nonthoracic surgery
J953	Chronic pulmonary insufficiency following surgery

(*Continued*)

https://doi.org/10.1017/S0950268822000917 Published online by Cambridge University Press

Appendix 1 (*Continued.*)

ICD-10 Code	ICD-10 Description
J954	Mendelson syndrome
J955	Postprocedural subglottic stenosis
J958	Other postprocedural respiratory disorders
J959	Postprocedural respiratory disorder, unspecified
J96	Respiratory failure, not elsewhere classified
J960	Acute respiratory failure
J9600	Acute respiratory failure, Type I [hypoxic]
J9601	Acute respiratory failure, Type II [hypercapnic]
J9609	Acute respiratory failure, Type unspecified
J961	Chronic respiratory failure
J9610	Chronic respiratory failure, Type I [hypoxic]
J9611	Chronic respiratory failure, Type II [hypercapnic]
J9619	Chronic respiratory failure, Type unspecified
J969	Respiratory failure, unspecified
J9690	Respiratory failure, unspecified, Type I [hypoxic]
J9691	Respiratory failure, unspecified, Type II [hypercapnic]
J9699	Respiratory failure, unspecified, Type unspecified
J98	Other respiratory disorders
J980	Diseases of bronchus, not elsewhere classified
J981	Pulmonary collapse
J982	Interstitial emphysema
J983	Compensatory emphysema
J984	Other disorders of lung
J985	Diseases of mediastinum, not elsewhere classified
J986	Disorders of diaphragm
J988	Other specified respiratory disorders
J989	Respiratory disorder, unspecified
J99	Respiratory disorders in diseases classified elsewhere
J990	Rheumatoid lung disease
J991	Respiratory disorders in other diffuse connective tissue disorders
J998	Respiratory disorders in other diseases classified elsewhere

https://doi.org/10.1017/S0950268822000917 Published online by Cambridge University Press

Appendix 2: Anonymised GP Practice Data

GP Practice names are anonymised and presented as Practice 1, Practice 2 etc…

Appendix 2

Total practice population by age							18 to 34		35–49		50–64		65–74		75–84		⩾85	
18–34	35–49	50–64	65–74	75–84	85+	Practice name	Proportion	Population	Proportion	Population	Proportion	Population	Proportion	Population	Proportion	Population	Proportion	Population
2638	2216	2498	1497	997	481	Practice 1	19%	507	10%	216	15%	375	18%	276	17%	169	9%	46
3254	3063	2342	1043	773	315	Practice 2	100%	3254	93%	2839	100%	2342	100%	1043	100%	773	100%	315
979	1176	1132	482	362	161	Practice 3	100%	979	100%	1176	100%	1132	100%	482	100%	362	100%	161
3899	2965	1874	728	377	134	Practice 4	97%	3796	98%	2891	97%	1825	98%	717	100%	377	100%	134
2714	2507	2030	894	469	234	Practice 5	100%	2714	94%	2364	100%	2030	100%	894	100%	469	99%	231
2383	1914	1262	558	302	133	Practice 6	100%	2383	100%	1914	100%	1262	100%	558	97%	294	100%	133
1488	743	267	38	8	5	Practice 7	100%	1488	89%	660	89%	239	100%	38	100%	8	100%	5
4487	4894	2750	721	348	109	Practice 8	88%	3949	97%	4750	100%	2750	100%	721	96%	335	100%	109
11 794	9416	5370	2301	1451	615	Practice 9	90%	10 565	100%	9416	100%	5370	100%	2301	98%	1429	100%	615
7402	2155	597	142	44	14	Practice 10	88%	6548	97%	2091	100%	597	100%	142	100%	44	100%	14
2731	2160	2474	1017	580	203	Practice 11	89%	2428	88%	1906	77%	1895	88%	894	80%	466	75%	152
4707	4042	2873	717	457	185	Practice 12	95%	4459	100%	4042	99%	2850	100%	717	100%	457	100%	185
767	716	945	452	307	199	Practice 13	60%	460	42%	298	28%	260	19%	87	11%	33	17%	33
2816	2960	3384	2044	1301	676	Practice 14	71%	2011	83%	2445	86%	2914	68%	1389	63%	826	59%	398
1459	1255	1682	833	428	142	Practice 15	62%	898	79%	986	83%	1395	67%	555	68%	292	62%	88
3400	3203	2811	1363	699	219	Practice 16	94%	3188	100%	3203	99%	2780	96%	1303	99%	692	98%	215
2462	2112	1838	774	609	244	Practice 17	100%	2462	100%	2112	100%	1838	100%	774	100%	609	99%	241
2967	2781	2872	1544	1024	371	Practice 18	97%	2882	100%	2781	99%	2832	99%	1528	99%	1018	100%	371
3148	2796	1415	395	253	96	Practice 19	100%	3148	97%	2718	100%	1415	100%	395	100%	253	100%	96
2578	2921	1939	666	356	139	Practice 20	100%	2578	87%	2532	91%	1768	100%	666	98%	349	98%	136
1611	2108	2100	1136	684	481	Practice 21	100%	1611	92%	1932	100%	2100	100%	1136	100%	684	100%	481
5815	5152	3864	1720	1009	372	Practice 22	93%	5409	100%	5152	95%	3682	100%	1720	99%	1002	100%	372
3097	2930	2396	1240	668	323	Practice 23	96%	2962	93%	2735	100%	2396	100%	1240	100%	668	100%	323
2522	2610	3103	1638	1323	497	Practice 24	91%	2303	100%	2610	99%	3070	98%	1604	100%	1323	99%	493
4476	4634	3087	1322	613	290	Practice 25	95%	4263	100%	4634	97%	3004	100%	1322	100%	613	99%	286
1887	1719	1796	973	615	368	Practice 26	25%	472	26%	442	18%	331	19%	188	15%	91	12%	45
2714	2193	1942	761	467	243	Practice 27	100%	2714	100%	2193	95%	1847	100%	761	100%	467	99%	240
2057	1907	1406	689	483	227	Practice 28	100%	2057	97%	1841	100%	1406	100%	689	100%	483	100%	227
4204	4278	4354	2423	1668	755	Practice 29	79%	3303	89%	3792	88%	3825	86%	2077	83%	1389	75%	564
1657	2538	1881	825	499	188	Practice 30	100%	1657	94%	2397	94%	1763	100%	825	103%	514	100%	188

1871	1348	1322	516	342	163	Practice 31	95%	1782	100%	1348	100%	1322	100%	516	99%	338	100%	163
838	1193	921	435	176	80	Practice 32	100%	838	100%	1193	100%	921	92%	402	100%	176	100%	80
1124	1212	1437	846	658	222	Practice 33	100%	1124	88%	1061	98%	1401	100%	846	100%	658	100%	222
1779	1156	1176	353	250	98	Practice 34	100%	1779	100%	1156	100%	1176	100%	353	100%	250	100%	98
5366	3798	2498	1175	667	249	Practice 35	97%	5187	100%	3798	100%	2498	100%	1175	99%	662	100%	249
1467	1112	854	354	172	41	Practice 36	44%	652	37%	412	26%	219	36%	126	22%	38	35%	14
2625	1978	2690	1164	602	200	Practice 37	94%	2475	94%	1868	96%	2590	100%	1164	95%	573	100%	200
3148	2360	2127	1150	790	413	Practice 38	98%	3075	91%	2145	100%	2127	100%	1150	100%	790	96%	397
2447	2243	1376	534	434	183	Practice 39	100%	2447	98%	2206	100%	1376	100%	534	100%	434	100%	183
1776	1927	1934	1163	795	374	Practice 40	90%	1607	100%	1927	100%	1934	100%	1163	99%	788	99%	371
1397	1527	1495	946	678	244	Practice 41	13%	175	17%	254	22%	332	21%	203	16%	106	13%	32
1516	1419	627	213	144	81	Practice 42	89%	1348	100%	1419	100%	627	100%	213	100%	144	100%	81
7943	8350	9129	5466	3173	1211	Practice 43	33%	2581	34%	2860	34%	3111	34%	1862	39%	1242	21%	259
2183	1527	1006	445	259	154	Practice 44	93%	2027	100%	1527	97%	974	100%	445	100%	259	100%	154
7909	5507	2884	848	329	110	Practice 45	100%	7909	94%	5163	99%	2852	100%	848	100%	329	100%	110
3642	3841	2733	1248	663	242	Practice 46	87%	3167	94%	3628	100%	2733	100%	1248	100%	663	100%	242
1116	490	171	23	7	4	Practice 47	100%	1116	94%	459	100%	171	100%	23	100%	7	100%	4
2591	2526	2253	1088	791	338	Practice 48	93%	2412	97%	2452	95%	2140	100%	1088	100%	791	100%	338
5580	3486	2631	1245	599	232	Practice 49	88%	4883	88%	3084	100%	2631	100%	1245	100%	599	100%	232
863	826	981	504	283	117	Practice 50	100%	863	88%	723	100%	981	100%	504	98%	277	100%	117
5097	4548	3372	1671	863	432	Practice 51	94%	4803	97%	4428	100%	3372	100%	1671	100%	863	99%	430
2945	3684	3650	2554	1681	661	Practice 52	95%	2798	100%	3684	90%	3285	99%	2537	97%	1636	98%	645
2287	2483	2120	1066	654	274	Practice 53	92%	2096	96%	2391	96%	2044	100%	1066	100%	654	100%	274
1396	1437	1239	750	518	311	Practice 54	100%	1396	100%	1437	100%	1239	96%	723	100%	518	100%	311
2637	2211	2004	1014	567	262	Practice 55	100%	2637	100%	2211	100%	2004	100%	1014	99%	562	100%	262
2719	2173	1740	892	583	303	Practice 56	97%	2634	98%	2126	99%	1719	100%	892	100%	583	98%	298
1453	1486	1608	871	724	298	Practice 57	90%	1308	92%	1362	82%	1319	100%	871	95%	690	100%	298
2425	2671	1992	905	473	181	Practice 58	26%	633	26%	683	22%	433	17%	156	13%	60	13%	24
2161	1848	1856	999	808	295	Practice 59	94%	2026	100%	1848	100%	1856	99%	986	100%	808	100%	295
3953	3462	2943	1224	772	387	Practice 60	97%	3829	98%	3393	100%	2943	100%	1224	100%	772	100%	387
1006	928	1086	675	396	182	Practice 61	83%	838	100%	928	100%	1086	100%	675	94%	374	100%	182
18 188	188	11	1	1	0	Practice 62	83%	15 099	100%	188	100%	11	0%	0	0%	0	0%	0
3551	3011	3333	1811	1129	459	Practice 63	98%	3468	100%	3011	100%	3333	99%	1798	100%	1129	99%	456
2837	2784	3126	1900	1241	552	Practice 64	36%	1032	19%	539	22%	687	21%	403	18%	222	10%	58
1642	1473	1051	397	204	69	Practice 65	100%	1642	100%	1473	100%	1051	100%	397	100%	204	100%	69
5289	5309	4780	2495	1795	915	Practice 66	100%	5289	98%	5225	99%	4739	100%	2495	100%	1795	100%	911

(*Continued*)

Appendix 2 (*Continued.*)

Total practice population by age							18 to 34		35–49		50–64		65–74		75–84		⩾85	
18–34	35–49	50–64	65–74	75–84	85+	Practice name	Proportion	Population	Proportion	Population	Proportion	Population	Proportion	Population	Proportion	Population	Proportion	Population
5525	3868	2392	1084	560	230	Practice 67	88%	4834	95%	3684	100%	2392	100%	1084	98%	549	100%	230
2016	1629	1585	775	515	271	Practice 68	100%	2016	95%	1548	93%	1482	100%	775	100%	515	100%	271
1626	1393	1124	424	243	83	Practice 69	100%	1626	98%	1359	100%	1124	100%	424	100%	243	100%	83
1580	1622	1934	1178	775	368	Practice 70	25%	395	17%	270	24%	470	24%	288	11%	85	11%	41
16 118	2141	1564	708	532	298	Practice 71	92%	14 846	100%	2141	97%	1513	100%	708	99%	527	100%	298
2859	2514	1149	423	245	93	Practice 72	95%	2709	100%	2514	100%	1149	100%	423	100%	245	100%	93
1346	1372	1749	1341	884	315	Practice 73	100%	1346	100%	1372	100%	1749	100%	1341	100%	884	99%	311
1650	1600	2227	1147	741	282	Practice 74	41%	679	29%	457	40%	901	27%	307	26%	195	32%	91
1927	1938	2202	1362	893	413	Practice 75	19%	367	4%	78	16%	351	22%	303	17%	153	16%	65
4840	6245	6856	4225	2815	1241	Practice 76	94%	4571	98%	6106	98%	6722	97%	4114	96%	2706	94%	1168
635	552	683	355	240	73	Practice 77	60%	381	100%	552	100%	683	100%	355	100%	240	100%	73
2522	2421	1404	533	308	105	Practice 78	90%	2270	93%	2254	97%	1360	98%	522	100%	308	100%	105
2697	2479	2657	1464	1251	352	Practice 79	100%	2697	100%	2479	100%	2657	100%	1464	100%	1251	99%	349
1579	2622	1866	1107	669	383	Practice 80	100%	1579	100%	2622	100%	1866	100%	1107	100%	669	99%	379
4892	3973	3126	1641	738	252	Practice 81	82%	4003	94%	3752	93%	2913	98%	1609	100%	738	100%	252
1401	1613	2012	1371	916	423	Practice 82	43%	600	25%	403	26%	519	26%	363	22%	200	18%	77
268 093	211 568	178 970	89 015	55 720	23 938		18–34	231 342	35–49	184 269	50–64	152 380	65–74	74 245	75–84	45 989	⩾85	19 229
																	⩾65	139 463