BMJ Open

Incidence of household transmission of acute gastroenteritis (AGE) in a primary care sentinel network (1992–2017): cross-sectional and retrospective cohort study protocol

Simon de Lusignan,[1,2] Emmanouela Konstantara,[1] Mark Joy,[1] Julian Sherlock,[1] Uy Hoang,[1] Rachel Coyle,[1] Filipa Ferreira,[1] Simon Jones,[1,3] Sarah J O'Brien[4]

Correspondence to
Professor Simon de Lusignan; s.lusignan@surrey.ac.uk

ABSTRACT

Introduction Acute gastroenteritis (AGE) is a highly transmissible condition. Determining characteristics of household transmission will facilitate development of prevention strategies and reduce the burden of this disease. We are carrying out this study to describe household transmission of medically attended AGE, and explore whether there is an increased incidence in households with young children.

Methods and analysis This study used the Royal College of General Practitioners (RCGP) Research and Surveillance Centre (RSC) primary care sentinel network, comprising data from 1 750 167 registered patients (August 2017 database). We conducted a novel analysis using a 'household key', to identify patients within the same household (n=811 027, mean 2.16 people). A 25-year repeated cross-sectional study will explore the incidence of medically attended AGE overall and then a 5-year retrospective cohort study will describe household transmission of AGE. The cross-sectional study will include clinical data for a 25-year period—1 January 1992 until the 31 December 2017. We will describe the incidence of AGE by age-band and gender, and trends in incidence. The 5-year study will use Poisson and quasi-Poisson regression to identify characteristics of individuals and households to predict medically attended AGE transmitted in the household. This will include whether the household contained a child under 5 years and the age category of the first index case (whether adult or child under 5 years). If there is overdispersion and zero-inflation we will compare results with negative binomial to handle these issues.

Ethics and dissemination All RCGP RSC data are pseudonymised at the point of data extraction. No personally identifiable data are required for this investigation. The protocol follows STrengthening the Reporting of OBservational studies in Epidemiology guidelines (STROBE). The study results will be published in a peer-review journal, the dataset will be available to other researchers.

INTRODUCTION

Acute gastroenteritis (AGE) contributes significantly to the burden of infectious

---

Strengths and limitations of this study

► UK general practice lends itself to this type of study because it is a registration base system (one patient registered with one general practitioner (GP)), and practices have been computerised since the 1990s.

► The Royal College of General Practitioners (RCGP) Research and Surveillance Centre (RSC) is one of the oldest sentinel networks in Europe, it recently celebrated its 50th Anniversary of collecting data about influenza and other infections in primary care, including acute gastroenteritis.

► RCGP RSC practices have had feedback about data quality including the importance of flagging first or new (incident) or review attendances.

► RCGP RSC data include a household key that enables the pseudonymised records of individuals who live in the same household to be identified.

► Our data have limitations; we will underestimate household size if some residents in the same household are not registered with an RCGP RSC practice.

---

diseases as well as having wider societal impact.[1–9] It is estimated that around 25% of the population suffers from an AGE episode per year. However, in common with many other conditions general practitioners only see the tip of the epidemiological iceberg,[10 11] with only 2% attending primary healthcare.[12] AGE has direct healthcare costs from the disease itself as well as its complications; and indirect costs from the loss of days of work and other disruption caused by cases, clusters and outbreaks.[12–15]

The population groups most vulnerable to AGE are children under 5 years old[16] and individuals with immunodeficiency[17–20] and immunosuppression-related comorbidities.[21] WHO recommended rotavirus vaccination for infants,[22–25] which was introduced in the UK in 2013[26–28] and has resulted in a significant

decrease of AGE presentation in the target group as well as older individuals,[29–31] suggesting herd immunity.[32–35]

AGE is a readily transmissible condition spreading rapidly between individuals and within institutions.[12 13 36] Like other infectious diseases, AGE requires a susceptible host and a favourable environment to spread. This can happen via direct person-to-person contact[37 38] or an indirect route (ie, contact with contaminated surfaces[39 40]). Young children (aged under 5 years) may be an important vector of this condition,[36 41–43] as they have a twofold to eightfold greater risk than adults to acquire AGE[44] and are more likely to spread it to older children and adults.[45–47] Determining the characteristics of household transmission will allow the development of appropriately targeted prevention strategies[48] and ultimately reduce the burden on healthcare and society.

The mechanisms whereby household transmission may take place can be classified as transmission by food, water, animal or person-to-person. For example, *Escherichia coli* O157 has been transmitted from inadequately cooked beef, and in milk.[49] *Giardia intestinalis* is a good example of protozoa principally acquired from contaminated water, but then passed by person-to person transmission.[50] There is a reservoir of *Salmonella* in farm animals transmitted to humans via a range of foods, after which human-to-human transmission is important.[51] Most viral gastroenteritis is transmitted by person-to-person spread or inhalation of droplets; this will be the the most common form of AGE, and most often lasts a few days and will not be reported to GPs.[52]

We are carrying out this study to describe the household transmission of medically attended AGE. Specifically, whether adults who live in houses with children aged under 2 and 5 years have a higher incidence of AGE.

Objectives
These are grouped by those that are derived from the repeated cross-sectional study and those from the retrospective cohort study.

Twenty-five-year repeated cross-sectional study:
► The rate of presentation with a first or new case of medically attended AGE per year.
► Individual characteristics which can predict presentation of AGE.
► The weekly incidence rates of medically attended AGE for children and adults.

Five-year retrospective cohort study:
► The primary outcome measure is the rate of presentation of two or more individuals with medically attended AGE from the same household, within 10 days. We will determine whether adults in households with young children (aged under 2 years and under 5 years) have a higher incidence of medically attended AGE than those that do not.
► Household characteristics which can predict presentation of medically attended AGE.
► The sequence of medically attended AGE presentation in the same household.

METHODS AND ANALYSIS
Study design
The study has two components, a repeated cross-sectional study and a retrospective cohort study. The repeated cross-section allows us to calculate the incidence of medically attended AGE and report trends in the population and incidence of medically attended AGE over time. The repeated cross-sectional analysis allows us to maximise the data available as the population registered at the start of our observation period will be different from that at the end. By way of comparison our retrospective cohort study will run for a shorter period and only include those registered with one of our network practices at the end of the study. The retrospective cohort study will explore household transmission rates.

We will conduct the 25-year repeated cross-sectional (1 January 1992 until 31 December 2017) and the 5-year retrospective cohort study (1 January 2012 until 31 December 2017) that will use the routinely collected primary care data within the RCGP RSC network database. The rationale for our different time periods for the repeated cross-sectional and retrospective cohort studies is that while denominator and AGE incidence data are reliable for this longer period, our household key is only reliable for the last 5 years.

The August 2017 RCGP RSC database contained just under 1.8 million patients' data, the 2002 database 1 million and the 1992 database 350 000 (table 1). Each progressive extraction of RCGP RSC is larger as more practices join the network; we anticipate a larger number to be involved in this study.

Study setting and population
We will carry out this study in the Royal College of General Practitioners (RCGP) Research and Surveillance Centre (RSC) primary care sentinel network. RCGP RSC lends itself to conducting this type of research.[53 54] English general practice is a registration-based system, where individuals register with a single general practice and have a unique patient identifier (National Health Service (NHS) number).[55] The network design allows people in the same household to be identified. Primary care records have been computerised since the 1990s. Key data are recorded using the Read terminology; this allows detailed coding of diagnosis, symptoms and other patient information.[56] Since 2004, most practices have been electronically linked to their local laboratory, ensuring all lab results are automatically posted into the practice computerised medical record (CMR) system; nearly all prescriptions are issued by GPs.

The RCGP RSC network database is a growing and nationally representative sentinel network. Based on the practices who were members in August 2017, it includes data from 175 general practices, 1 750 167 registered patients, amounting to approximately 3.1% of the National population.[53 57] It is likely that a larger population will be available for this study.

Table 1 RCGP RSC database size by year of database assembly

Extraction date	Number of registered patients
31/08/2017	1 769 004
31/12/2016	1 750 168
31/12/2015	1 722 608
31/12/2014	1 689 412
31/12/2013	1 639 250
31/12/2012	1 588 022
31/12/2011	1 536 902
31/12/2010	1 493 611
31/12/2009	1 452 718
31/12/2008	1 414 621
31/12/2007	1 376 848
31/12/2006	1 337 105
31/12/2005	1 285 278
31/12/2004	1 237 254
31/12/2003	1 179 378
31/12/2002	1 034 446
31/12/2001	922 267
31/12/2000	820 163
31/12/1999	727 957
31/12/1998	648 774
31/12/1997	585 291
31/12/1996	523 976
31/12/1995	477 484
31/12/1994	431 264
31/12/1993	390 437
31/12/1992	343 484

RCGP RSC is one of the oldest surveillance networks; it has been producing a weekly return on infectious disease, including gastroenteritis, for over 50 years.[54] Over that period, practices have received feedback about their data quality, initially through practice visits, but more recently through online training and a practice dashboard.

The RCGP RSC is broadly representative of the national population in terms of: (1) age-sex distribution; (2) socioeconomic status, measured using the Index of Multiple Deprivation (IMD); (3) ethnicity[58 59] and (4) urban-rural distribution, using the Office of National Statistics (ONS) classification.[60] The population is slightly younger, more distributed towards London and urban locations, more ethnically mixed and less deprived—although these differences are small.[53] The RCGP RSC actively recruits in areas where it is under-represented.

PATIENT AND PUBLIC INVOLVEMENT
Patients were not involved in the development of this protocol.

Household key
The RCGP RSC network database has a 'household key,' which can flag patients who live in the same household. This is assigned by flagging groups of individual with an identical first line or address and post code (this is done programmatically within the GP CMR system—these personal data are not extracted). From the August 2017 database, we identified 811 027 households with a mean size of 2.16 people. This compares with the 2011 national census of 56 075 900 in 23 366 000 households and a mean size of 2.36[57]; 2.9% of the patients within the RSC did not have a household key assigned; this is largely because they do not have a properly formed post code.

The household key is defined as an identical first line of address and identical full post code. This matching is done programmatically, at the point of data extraction from the GP system so RCGP RSC staff do not have access to these data. We anticipate it is reliable because when a patient registers, the GP CMR system requires entry of the address once, and then assigns it to successive family members. However, it may underestimate true household size where one or more members of a household are registered with a different practice, which is not part of the RCGP RSC network. We class households of 12 or more as a 'communal establishment'—and these groups are outside the scope of this study, in line with handling of data from ONS.[61] Around a quarter of people live in two-person households (24.32), with a fifth each of living in single (22.34), three-person (19.08) or four-person (19.04) households (table 2).

AGE case definition
We will use the current RCGP RSC, ontologically developed case definition; which will be applied to all years of data analysed. An ontological approach to case definitions formally describes the concepts used to define the disease.[62] A disease such as AGE can be defined by a combination of one or more concepts such as: diagnosis, clinical features, lab results and any treatment given.[63] Historically, RSC defined AGE using the diagnostic Read codes that mapped to equivalent infectious intestinal diseases codes within the International Classification of Disease (ICD).[56] Our ontological approach was developed from the AGE definition used in the Second Study of Infectious Intestinal Disease in the Community,[1] with the Read code mappings updated to match ICD-10.[64] Our ontology can be readily mapped to other coding systems. For example, it will facilitate consistent case definition within the Systematised Nomenclature of Medicine—Clinical Terms (SNOMED CT) when it is rolled out across the NHS in 2018.[65 66]

AGE incidence
Using RCGP RSC data (August 2017 extract), we have looked at the incidence of AGE over our 25-year study period. We have looked at the incidence by age band. We have used two modes, one separating off children under

Table 2 Household size in RCGP RSC

Household size	Female	Male	Population	Number of households	Proportion of population	Proportion of households
1	181 422	189 754	371 176	371 176	22.34	46.25
2	212 256	191 772	404 028	202 014	24.32	25.17
3	163 992	152 916	316 908	105 636	19.08	13.16
4	158 282	158 010	316 292	79 073	19.04	9.85
5	69 385	71 170	140 555	28 111	8.46	3.50
>6	55 760	56 623	112 383	16 613	6.76	2.07

These data are from the August 2017 RCGP RSC extract and exclude households (communal establishments>12 people) and those without a valid post code, who do not have a household key.

RCGP RSC, Royal College of General Practitioners Research and Surveillance Centre.

2 years (figure 1) and another separating off children under 5 years (figure 2).

The number of incident cases of AGE in the RCGP RSC in 2016 was 16 325; over the 15-year period of the retrospective cohort there are 297 691 cases, and over the last 25 years 364 529 cases.

Sample size calculation

We based our sample size on a proportion test on the two groups: households with an under 5 and households without an under 5.[67] Since we are in the low prevalence rate regime, sample size of around 2650 enables us to detect an OR of around 2. Since we intend to extract a sample of a size much larger than 2650, we will be able to detect an OR of ≥2.1 with 80% power at the 5% level of significance.[67]

Outcome measures

Twenty-five-year repeated cross-sectional study:
► The rate of presentation with a first or new case of medically attended AGE per year.

► Individual characteristics that can predict presentation of medically attended AGE (box 1).
► The weekly incidence rates of medically attended AGE for children and adults.

Five-year retrospective study:
► The primary outcome measure is the rate of presentation of two or more individuals with medically attended AGE from the same household, within 10 days. We will determine whether adults in households with young children (aged under 2 years and under 5 years) have a higher incidence of medically attended AGE than those that do not.
► Household characteristics that can predict presentation of medically attended AGE within households (box 2).
► The sequence of medically attended AGE presentation in the same household.
► Individual characteristics that can also predict presentation of medically attended AGE within households (box 1).
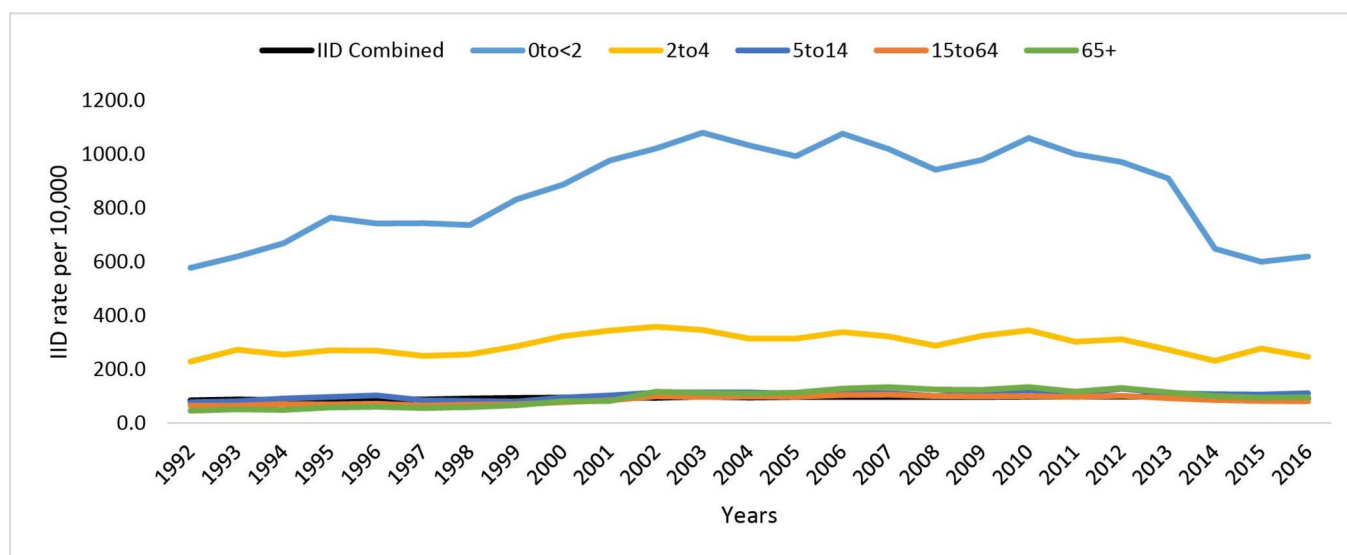

Figure 1 Incidence of gastroenteritis in Royal College of General Practitioners Research and Surveillance Centre 1992–2017, with children under 2 years in a separate age-band. IID, Infectious Intestinal Disease.
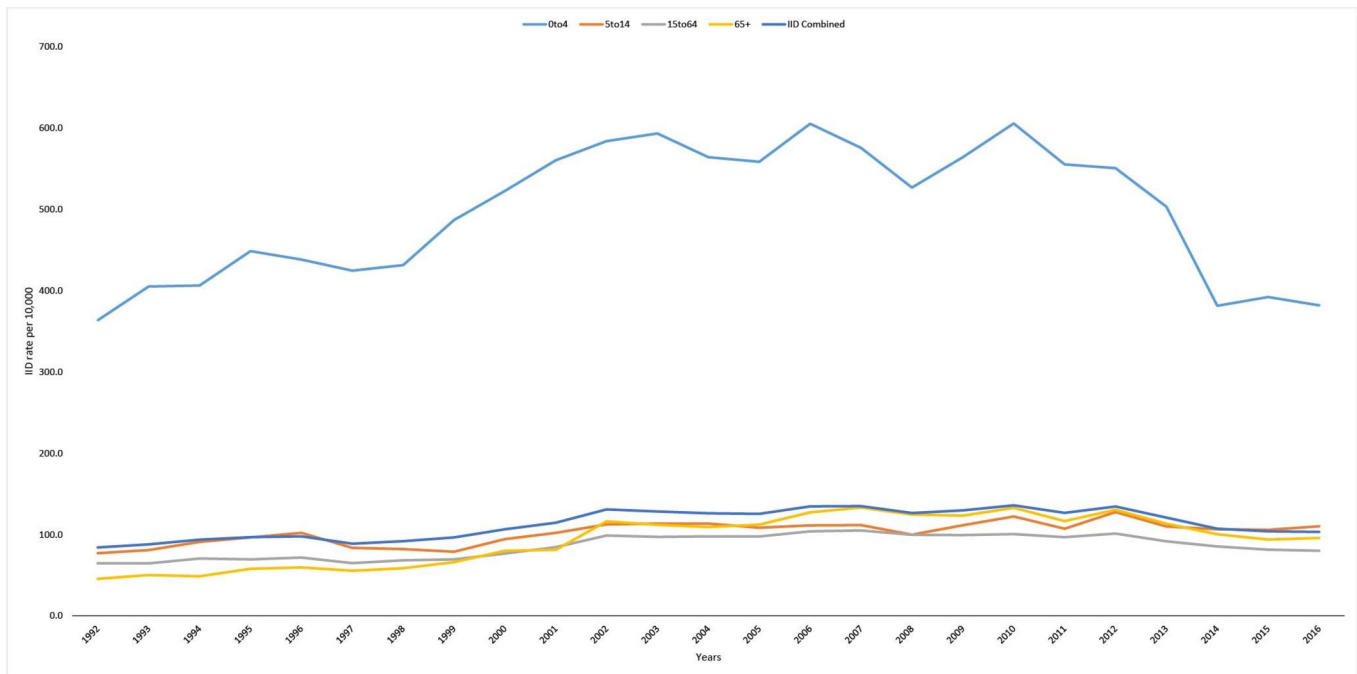
Figure 2 Incidence of gastroenteritis in Royal College of General Practitioners Research and Surveillance Centre 1992–2017, with children under 5 years in a separate age-band. IID, Infectious Intestinal Disease.

Study variables and available data

For the cross-sectional study, we will extract age-band, gender and medically attended AGE data.

The retrospective cohort study will use the more extensive demographic and clinical data available in the database, as described above. We will additionally include the chronic diseases:

► Cardiovascular and cerebrovascular (heart disease, stroke, chronic kidney disease and hypertensions);
► Respiratory (asthma and chronic obstructive pulmonary disease);
► Common mental health problems (anxiety and depression).

Box 1   Individual characteristics

Age
Gender
Ethnicity (2001 census)
► White
► Asian
► Black
► Mixed
► None
► Other
Socioeconomic status (2001 census)
► IMD quintile 1—most deprived
► IMD quintile 2
► IMD quintile 3
► IMD quintile 4
► IMD quintile 5—least deprived
Body mass index kg/m^2 (2001 census)
► Underweight (<18.5)
► Normal (18.5–24.9)
► Overweight (25.0–29.9)
► Obese class 1 (30.0–34.9)
► Obese class 2 (35.0–39.9)
► Obese class 3 (≥40.0)
► None (information missing)
History of rotavirus vaccine

IMD, Index of Multiple Deprivation.

Box 2   Household characteristics

Composition
Adults only (aged 18 years or more)
► 1–2 adults
► 3+ adults
Children aged under 5 years
► 1 child–1 adult
► 1 child–2+ adults
► 2+ children–1 adult
► 2+ children–2+ adults
Children aged over 5 years
► 1 child–1 adult
► 1 child–2+ adults
► 2+ children–1 adult
► 2+ children–2+ adults
Children of mixed ages and one adult
► 1 child <5 and 1+ >5–1 adult
► 2+ children <5 and 1+ >5–1 adult
► 1 child >5 and 1+ <5–2+ adults
► 2+ children >5 and 1+ <5–2+ adults
Size
Mean (for census comparison) and median age

Statistical analysis

The cross-sectional study will use descriptive statistics to describe any trend in incidence over the 25-year observation period. We will also use descriptive statistics to determine both the standardised and crude rate for each year and to observe any individual characteristics in cases of medically attended AGE, such as ethnicity and body mass index.

The 5-year retrospective study will show incidence of two or more individuals from the same household presenting with medically attended AGE within 10 days of each other. Households will be grouped by whether they contain a child aged under years or not. We will observe the intervals between presentations, as well as the sequence of occurrence for different age groups. We will explore Poisson and quasi-Poisson regression and we will address issues with overdispersion and zero-inflation by comparing results with negative binomial.

In order to determine representativeness of RCGP RSC household data, we will compare the mean size of households in our dataset with the 2011 census data mean from ONS[48] using a one-sample t-test.

We do not plan a sensitivity analysis. We do not expect the low percentage of individuals with no valid household key recorded to have an effect on the study (n=1 750 167, out of which 50 979 patients without a household key recorded; 2.91%).

All statistical analysis will be performed using the statistical package, RStudio, V.3.3.1.[49] The R scripts will be available to readers on request.

USE OF GUIDELINES

This protocol was produced following the STrengthening the Reporting of OBservational studies in Epidemiology (STROBE) checklist for cohort studies (see online supplementary file 1).

ETHICAL CONSIDERATIONS

The study does not require formal ethics committee approval. All data to be used have been anonymised at the point of data extraction. The study has been reviewed by the University of Surrey Research Integrity and Governance Office, tested against the Health Research Authority (HRA)/Medical Research Council 'is this research' tool (http://www.hra-decisiontools.org.uk/research/), and is considered to be an audit of current practice. No clinically identifiable information will be made available to researchers or in any publications.

DISSEMINATION

The final agreed protocol and the outputs of this study will be published in a peer-reviewed open access journal within the domains of primary care, epidemiology, surveillance, vaccines and infectious diseases. The research team will seek to present findings at relevant seminars and conferences.

A report with key findings, implication for practice and call for further research will be submitted to the funder at the end of the study. The data used for this research can be made available to other researchers on application to the corresponding author.

DISCUSSION

This is novel use of a household key within a primary care database to report if there is evidence of household transmission. This approach may be replicated to look for evidence of household transmission of other infections.

The RCGP RSC network database is appropriate for this research. RCGP RSC has records of over a quarter of a million cases of AGE, and good recording of its 'household' key. The limitation of these data are that we may under-record household size where one or more individuals in a household are registered with a different practice. Additionally, we know that not all AGE is medically attended; it is possible that other cases of AGE may not be reported to a patient's GP.

Limitations of the study

Medically attended AGE

As described in the introduction, we are measuring trends in medially attended AGE and recognise this is a small proportion of the total incidence of this condition.[10 11] Household characteristics associated with AGE and the sequence of presentation in the household may reflect healthcare seeking behaviour, perception of risk as well as the dynamics of transmission.

Presentation with AGE at the time of presentation with other conditions

Some people attended about their comorbidity at the same time that they presented with AGE. We do not have data to know whether that management of their comorbidity was discussed at a time their presentation with AGE, or AGE mentioned at the time of presentation for their chronic disease. The proportion was greatest in those aged over 65 years, where 1.2%–5.6% presented on the same day with a comorbidity.

The household key underestimates the household size

If some residents in a household are registered with a practice outside the sentinel network, they would not be included in the household. We think this will happen less with families, and more with younger people and in conurbations where there is more choice of general practices.

Microbiological diagnoses are not commonly made in AGE in primary care

For a variety of reasons including that AGE is often self-limiting and there may be local limitations on testing, we anticipate finding few microbiologically proven cases.

Where they are recorded in the primary care computerised medical record system, we will have access.

Some cases of AGE may go to the emergency department or hospital without attending primary care

We may not capture all more serious events, if they go straight to hospital. However, most practices record such data retrospectively into their primary care databases.

Notwithstanding these limitations, we anticipate reporting whether there is evidence of household spread of AGE using routine primary care data within the RCGP RSC database.

Author affiliations
[1]Department of Clinical and Experimental Medicine, University of Surrey, Guildford, UK
[2]Royal College of General Practitioners (RCGP) Research and Surveillance Centre (RSC), London, UK
[3]Center for Healthcare Innovation and Delivery Science, Department of Population Health, NYU School of Medicine, New York City, New York, USA
[4]NIHR Health Protection Research Unit in Gastrointestinal Infections, Liverpool, UK

Disclaimer The funders are updated periodically with the research; the funders have not had a role in the development of this protocol, and no access to the study dataset.

Competing interests None declared.

Patient consent Not required.

Ethics approval Approval for this work has been granted by the RCGP RSC study approval committee. Although in accordance with their policy for studies that were not a result of competitive peer-reviewed award, they require an open access peer-review protocol to be published.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement RCGP RSC data are available to analyse. Application form online at www.rcgp.org.uk/rsc. The authors will share their R scripts and the code lists used for this analysis.

REFERENCES

1. Tam CC, Viviani L, Rodrigues LC, et al. The second study of infectious intestinal disease (IID2): increased rates of recurrent diarrhoea in individuals aged 65 years and above. BMC Public Health 2013;13:739.
2. Tam CC, Rodrigues LC, Viviani L, et al. Longitudinal study of infectious intestinal disease in the UK (IID2 study): incidence in the community and presenting to general practice. Gut 2012;61:69–77.
3. Dalby-Payne J, Elliott E. Gastroenteritis in children. BMJ Clin Evid 2004;12:443–54.
4. Black RE, Cousens S, Johnson HL, et al. Global, regional, and national causes of child mortality in 2008: a systematic analysis. Lancet 2010;375:1969–87.
5. Wardlaw T, Salama P, Brocklehurst C, et al. Diarrhoea: why children are still dying and what can be done. Lancet 2010;375:870–2.
6. Mokomane M, Kasvosve I, de Melo E, et al. The global problem of childhood diarrhoeal diseases: emerging strategies in prevention and management. Ther Adv Infect Dis 2018;5:29–43.
7. Walker CLF, Rudan I, Liu L, et al. Global burden of childhood pneumonia and diarrhoea. Lancet 2013;381:1405–16.
8. Fischer Walker CL, Perin J, Aryee MJ, et al. Diarrhea incidence in low- and middle-income countries in 1990 and 2010: a systematic review. BMC Public Health 2012;12:220.
9. Liu L, Johnson HL, Cousens S, et al. Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. Lancet 2012;379:2151–61.
10. Last JM, Adelaide DP. The iceberg: 'completing the clinical picture' in general practice. 1963. Int J Epidemiol 2013;42:1608–13.
11. Last JM. Commentary: the iceberg revisited. Int J Epidemiol 2013;42:1613–5.
12. Sandmann FG, Jit M, Robotham JV, et al. Burden, duration and costs of hospital bed closures due to acute gastroenteritis in England per winter, 2010/11-2015/16. J Hosp Infect 2017;97:79–85.
13. Lopman BA, Reacher MH, Vipond IB, et al. Epidemiology and cost of nosocomial gastroenteritis, Avon, England, 2002-2003. Emerg Infect Dis 2004;10:1827–34.
14. Kirk MD, Pires SM, Black RE, et al. World health organization estimates of the global and regional disease burden of 22 foodborne bacterial, protozoal, and viral diseases, 2010: a data synthesis. PLoS Med 2015;12:e1001921.
15. Kosek M, Bern C, Guerrant RL. The global burden of diarrhoeal disease, as estimated from studies published between 1992 and 2000. Bull World Health Organ 2003;81:197–204.
16. Giaquinto C, van Damme P. REVEAL Study Group. Age distribution of paediatric rotavirus gastroenteritis cases in Europe: the REVEAL study. Scand J Infect Dis 2010;42:142–7.
17. Bok K, Green KY. Norovirus gastroenteritis in immunocompromised patients. N Engl J Med 2012;367:2126–32.
18. Yin Y, Metselaar HJ, Sprengers D, et al. Rotavirus in organ transplantation: drug-virus-host interactions. Am J Transplant 2015;15:585–93.
19. Raini SK, Nyangao J, Kombich J, et al. Human rotavirus group a serotypes causing gastroenteritis in children less than 5 years and HIV-infected adults in Viwandani slum, Nairobi. Ethiop J Health Sci 2015;25:39–46.
20. Schwartz S, Vergoulidou M, Schreier E, et al. Norovirus gastroenteritis causes severe and lethal complications after chemotherapy and hematopoietic stem cell transplantation. Blood 2011;117:5850–6.
21. Vehreschild MJ, Vehreschild JJ, Hübel K, et al. Diagnosis and management of gastrointestinal complications in adult cancer patients: evidence-based guidelines of the Infectious Diseases Working Party (AGIHO) of the German Society of Hematology and Oncology (DGHO). Ann Oncol 2013;24:1189–202.
22. Rotavirus vaccines. WHO position paper – January 2013. Wkly Epidemiol Rec 2013;88:49–64.
23. Karafillakis E, Hassounah S, Atchison C. Effectiveness and impact of rotavirus vaccines in Europe, 2006-2014. Vaccine 2015;33:2097–107.
24. Hungerford D, Read JM, Cooke RP, et al. Early impact of rotavirus vaccination in a large paediatric hospital in the UK. J Hosp Infect 2016;93:117–20.
25. World Health Organization. Meeting of the immunization strategic advisory group of experts, April 2009--conclusions and recommendations. Wkly Epidemiol Rec 2009;84:220–36.
26. Public Health England. Rotavirus vaccination programme for infants. 2013 https://www.gov.uk/government/collections/rotavirus-vaccination-progarmme-for-infants
27. Soares-Weiser K, Maclehose H, Bergman H, et al. Vaccines for preventing rotavirus diarrhoea: vaccines in use. Cochrane Database Syst Rev 2012;11:Cd008521.
28. Iturriza-Gómara M, Cunliffe N. Rotavirus vaccine: a welcome addition to the immunisation schedule in the UK. BMJ 2013;346:f2347.
29. Marlow R, Muir P, Vipond B, et al. Assessing the impacts of the first year of rotavirus vaccination in the United Kingdom. Euro Surveill 2015;20:30077.

30. Atchison CJ, Stowe J, Andrews N, *et al*. Rapid declines in age group-specific rotavirus infection and acute gastroenteritis among vaccinated and unvaccinated individuals within 1 year of rotavirus vaccine introduction in England and Wales. *J Infect Dis* 2016;213:243–9.
31. Thomas SL, Walker JL, Fenty J, *et al*. Impact of the national rotavirus vaccination programme on acute gastroenteritis in England and associated costs averted. *Vaccine* 2017;35:680–6.
32. Nakagomi O, Iturriza-Gomara M, Nakagomi T, *et al*. Incorporation of a rotavirus vaccine into the national immunisation schedule in the United Kingdom: a review. *Expert Opin Biol Ther* 2013;13:1613–21.
33. Mughini-Gras L, Pijnacker R, Heusinkveld M, *et al*. Societal burden and correlates of acute gastroenteritis in families with preschool children. *Sci Rep* 2016;6:22144.
34. Mast TC, Wang FT, Su S, *et al*. Evidence of herd immunity and sustained impact of rotavirus vaccination on the reduction of rotavirus-related medical encounters among infants from 2006 through 2011 in the United States. *Pediatr Infect Dis J* 2015;34:615–20.
35. Pollard SL, Malpica-Llanos T, Friberg IK, *et al*. Estimating the herd immunity effect of rotavirus vaccine. *Vaccine* 2015;33:3795–800.
36. Heusinkveld M, Mughini-Gras L, Pijnacker R, *et al*. Potential causative agents of acute gastroenteritis in households with preschool children: prevalence, risk factors, clinical relevance and household transmission. *Eur J Clin Microbiol Infect Dis* 2016;35:1691–700.
37. Karsten C, Baumgarte S, Friedrich AW, *et al*. Incidence and risk factors for community-acquired acute gastroenteritis in north-west Germany in 2004. *Eur J Clin Microbiol Infect Dis* 2009;28:935–43.
38. Conrad D, Dee K, Keenan A, *et al*. The role of household transmission in an outbreak of viral gastroenteritis in a primary school in Liverpool, England. *Public Health* 2013;127:882–4.
39. Repp KK, Hostetler TP, Keene WE. A norovirus outbreak related to contaminated surfaces. *J Infect Dis* 2013;208:295–8.
40. Wikswo ME, Kambhampati A, Shioda K, *et al*. Outbreaks of acute gastroenteritis transmitted by person-to-person contact, environmental contamination, and unknown modes of transmission--United States, 2009-2013. *MMWR Surveill Summ* 2015;64:1–16.
41. Enserink R, Ypma R, Donker GA, *et al*. Infectious disease burden related to child day care in the Netherlands. *Pediatr Infect Dis J* 2013;32:1–40.
42. Phillips G, Tam CC, Rodrigues LC, *et al*. Risk factors for symptomatic and asymptomatic norovirus infection in the community. *Epidemiol Infect* 2011;139:1676–86.
43. Götz H, Ekdahl K, Lindbäck J, *et al*. Clinical spectrum and transmission characteristics of infection with Norwalk-like virus: findings from a large community outbreak in Sweden. *Clin Infect Dis* 2001;33:622–8.
44. Perry S, de la Luz Sanchez M, Hurst PK, *et al*. Household transmission of gastroenteritis. *Emerg Infect Dis* 2005;11:1093–6.
45. Chapin AR, Carpenter CM, Dudley WC, *et al*. Prevalence of norovirus among visitors from the United States to Mexico and Guatemala who experience traveler's diarrhea. *J Clin Microbiol* 2005;43:1112–7.
46. Simmons K, Gambhir M, Leon J, *et al*. Duration of immunity to norovirus gastroenteritis. *Emerg Infect Dis* 2013;19:1260–7.
47. Sacri AS, De Serres G, Quach C, *et al*. Transmission of acute gastroenteritis and respiratory illness from children to parents. *Pediatr Infect Dis J* 2014;33:583–8.
48. Lopman B. *Global burden of norovirus and prospects for vaccine development*: CDC Found, 2015.
49. Adams NL, Byrne L, Smith GA, *et al*. Shiga toxin-producing escherichia coli O157, England and Wales, 1983-2012. *Emerg Infect Dis* 2016;22:590–7.
50. Casemore DP. Foodborne protozoal infection. *Lancet* 1990;336:1427–32.
51. World Health Organisation. Salmonella (non-typhoidal). 2018 http://www.who.int/news-room/fact-sheets/detail/salmonella-(non-typhoidal)
52. Seymour IJ, Appleton H. Foodborne viruses and fresh produce. *J Appl Microbiol* 2001;91:759–73.
53. Correa A, Hinton W, McGovern A, *et al*. Royal College of General Practitioners Research and Surveillance Centre (RCGP RSC) sentinel network: a cohort profile. *BMJ Open* 2016;6:e011092.
54. de Lusignan S, Correa A, Smith GE, *et al*. RCGP research and surveillance centre: 50 years' surveillance of influenza, infections, and respiratory conditions. *Br J Gen Pract* 2017;67:440–1.
55. NHS Digital. NHS number. https://digital.nhs.uk/services/nhs-number
56. de Lusignan S. Codes, classifications, terminologies and nomenclatures: definition, development and application in practice. *Inform Prim Care* 2005;13:65–9.
57. Office for National Statistics. 2011 census: population and household estimates for the United Kingdom, March 2011. 2011 https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/2011censuspopulationandhouseholdestimatesforenglandandwales/2012-07-16
58. Kumarapeli P, Stepaniuk R, de Lusignan S, *et al*. Ethnicity recording in general practice computer systems. *J Public Health* 2006;28:283–7.
59. Tippu Z, Correa A, Liyanage H, *et al*. Ethnicity recording in primary care computerised medical record systems: an ontological approach. *J Innov Health Inform* 2017;23:799.
60. Office for National Statistics. 2011 rural/urban classification. 2011 https://www.ons.gov.uk/methodology/geography/geographicalproducts/ruralurbanclassifications/2011ruralurbanclassification
61. Office for National Statistics. Occupied address (household) estimates from Administrative Data: 2011 and 2015. 2015 https://www.ons.gov.uk/census/censustransformationprogramme/administrativedatacensusproject/administrativedatacensusresearchoutputs/householdsandfamilies/occupiedaddresshouseholdestimatesfrommadministrativedata2011and2015
62. de Lusignan S. In this issue: Ontologies a key concept in informatics and key for open definitions of cases, exposures, and outcome measures. *J Innov Health Inform* 2015;22:170.
63. Liu L, Özsu MT. *Encyclopedia of database systems*. Heidelberg, Germany: Springer Berlin, 2009.
64. de Lusignan S, Shinneman S, Yonova I, *et al*. An ontology to improve transparency in case definition and increase case finding of infectious intestinal disease: database study in english general practice. *JMIR Med Inform* 2017;5:e34.
65. de Lusignan S, Chan T, Jones S. Large complex terminologies: more coding choice, but harder to find data--reflections on introduction of SNOMED CT (Systematized Nomenclature of Medicine--Clinical Terms) as an NHS standard. *Inform Prim Care* 2011;19:3–5.
66. NHS Digital. SNOMED CT implementation in primary care. https://digital.nhs.uk/services/terminology-and-classifications/snomed-ct/snomed-ct-implementation-in-primary-care
67. Hsieh FY, Bloch DA, Larsen MD. A simple method of sample size calculation for linear and logistic regression. *Stat Med* 1998;17:1623–34.