**Retrospective cohort study to estimate the burden of complications from common gastrointestinal infections and remediable factors that contribute to this burden**

**Study Overview**

The aim of our study is to measure health problems that can follow infectious gastroenteritis (diarrhoea) due to a range of bacterial infections. In particular we are looking at infections with bacteria called *Campylobacter*, *Salmonella* and verocytotoxin producing *Escherichia coli* (VTEC).

When a person has one of these infections, laboratories must report the results to Public Health England (PHE), the national public health authority that monitors these diseases. We will work with PHE to identify those with a confirmed laboratory report of *Campylobacter*, *Salmonella* and VTEC from 1st January 2009 to 31st December 2014. These records will then be linked to primary care (general practice) data using the Clinical Practice Research Datalink (CPRD); linked to secondary care data using the Hospital Episode Statistics (HES) data and linked to the Office of National Statistics (ONS) mortality file on deaths registration and linked to the English Index of Multiple Deprivation (IMD). These linkages will allow us to quantify the common complications seen in primary care, the more severe complications seen in secondary care (hospitals) and registered deaths, as well as estimate the healthcare costs associated with such complications.

Specifically, linking these datasets will help us address the following questions:

1. What are the complications that occur following infection with *Campylobacter*, *Salmonella* or VTEC?
2. Is there a link between the use of antibiotics and acid reducing drugs in the development of complications?
3. What is the cost to the NHS of these complications from gastrointestinal infections?

The pattern of these complications across socioeconomic groups can be identified by the proposed linkage to IMD data. Because some of the complications are uncommon, we need to follow up a large group of people with infection to achieve this. The only practical way to do this is by the proposed record linkage, using past cases of infection that occurred during 2009-2014. To protect patient confidentiality, the linkage will be undertaken by designated specialist bodies, and records are anonymised so that researchers working on the data cannot identify any individual. However, if anyone does not want their data included in a study like this and informs us, we will take steps to exclude their records.

**FAQs**

**What is data linkage?**

A data linkage is when one source of information (data) is merged with another source to provide additional information about individuals. This can only be done if you have the same identifiers (such as name and NHS number) in both sources of information.

**How will the linkage be completed?**

This study will involve a data linkage using personal identifiers (name, date of birth and NHS number) of records of infection with one of these three types of bacteria to hospital records and GP records (for those who have consented to this at your GP practice). The data linkage will be conducted in a secure manner by third parties not involved in the study. To do this, we will request for laboratory records from Public Health England. GP records will be provided by the CPRD, and hospital records will be provided by the Health and Social Care Information Centre (HSCIC). The HSCIC is a national government approved agency which will perform the linkage of the laboratory, GP and hospital records using name, date of birth and NHS number. They will send the data back to us with all personal identifiers (name, date of birth and NHS number) removed and after this point we will not be able to identify individuals.All data will be handled in accordance to the Data Protection Act 1998. **Your records may be included in this data linkage study unless you inform us not to access the relevant records that are routinely held by the organisations named above.** Further information on compliance of this study with the Data Protection Act 1998 is outlined in the table at the end of this section. For details about the full process of this data linkage, please see the diagram which shows what will happen to your data at the end of this document.

**How many people will you follow-up?**

We estimate we will be able to follow-up a total of 302,000 people. This is based on typical annual estimates; we expect to obtain a laboratory-confirmed diagnosis from 235,000 individuals over the five-year study period (approximately 200,000 *Campylobacter*, 30,000 *Salmonella* and 5,000 VTEC).

The estimated coverage rate of CPRD is 8.8% of the UK population, with 65-70% of English contributing practices consenting to record linkage. We therefore expect that approximately 13,400 laboratory-confirmed cases can be linked to CPRD. Five controls are required for each case, i.e. an additional 67,000 individuals. The total estimated sample size is therefore 302,000. The sample size is dictated by the total number of laboratory-diagnosed cases of the infections of interest within the study period.

**How can I take part?**

We would like to carry out a data linkage with the medical records of all those with a laboratory confirmed infection of *Campylobacter*, *Salmonella* or VTEC between 1st January 2009 and 31st December 2014. As the study uses only routinely collected data, we will not be approaching individuals in person to participate in the study.

**Why did you not request consent to access my medical records?**

The estimated 67,000 patients from CPRD will be patients from GP Practices who have consented to have their medical records linked. These patients no longer need to consent again to this study. The remaining 235,000 patients from PHE data have not been contacted to seek their consent. This is because the study is using retrospective data routinely collected by PHE without consent. As such, it was deemed impractical to contact all 235,000 patients for consent to use their data in this study while explaining to patients why their data is routinely held by PHE. An exemption under Section 251 of the NHS Act 2006 (previously Section 60 of the Health and Social Care Act 2001) allows PHE to receive patient-identifiable data from other organisations without patient consent in order to monitor infectious disease. HSCIC are responsible for hospital data and are able to link datasets using identifiers through a trusted data linkage service. This project has been discussed with three Patient and Public Involvement in Research representatives and they believe this is in the interest of the public.

**How can I opt out?**

If you, your child or a dependent have had a laboratory diagnosis for *Campylobacter*, *Salmonella* or *Escherichia coli* O157 between 1st January 2009 and 31st December 2014 and **you do not want your or their records to be linked, you can contact the Principal Investigator for this study** **oluwaseun.esan@phc.ox.ac.uk** **and we will inform the third party performing the linkage on our behalf**.

**When can I opt out?**

After data linkage has been performed, we will be unable to identify specific individuals in our data. This is because information such as name and NHS number will have been removed. For this reason we ask that, if you would like to opt out, you contact us **before 30th April 2016.**

**How will data be stored?**

Data will be stored on a secure server at the University of Oxford that will be available only to members of the team carrying out this research in accordance with the Data Protection Act 1998.

**Who is funding this study?**

This study is funded by the National Institute of Health Research: Health Protection Research Unit in Gastrointestinal Infections.

**Has this study been reviewed?**

The study has received ethical approval from the Health Research Authority – NRES Committee Yorkshire & The Humber, South Yorkshire Research Ethics Committee (Ref: 15-YH-0395), and received conditional approval from the Health Research Authority - Confidential Advisory Group (Ref: 16/CAG/0027).

**Who can I contact with questions or comments?**

You can call us on 01865 289300, email us at oluwaseun.esan@phc.ox.ac.uk or write to us at

Oluwaseun Esan

Nuffield Department of Primary Care Health Sciences

Gibson Building

Radcliffe Observatory Quarter

Woodstock Road

Oxford

OX2 6GG

**Table 1. Compliance with the principles outlined in the Data Protection Act 1998**

|  |  |  |
| --- | --- | --- |
| 1 | Fair processing | Only the identifiers required for linkage are requested. Linkage will be performed by third parties not involved in the research |
| 2 | Used for the specified purpose | The data will be used only for the analysis outlined in the approved study protocol, |
| 3 | Minimum necessary for the purpose | Only the minimum identifiers required for linkage are requested. The research dataset will contain no patient identifiers. |
| 4 | Accuracy | Only cases with a confirmed case of infection with Campylobacter, Non-typhoidal Salmonella and VTEC will be selected for linkage. |
| 5 | Kept for the minimum time necessary | The identifiers will be destroyed once linkage is complete and validated. The linked dataset containing no patient identifiable information data will be stored for five years in line with the NHS Records Retention policy. |
| 6 | In accordance to the rights of data subject | PHE is the national health authority responsible for monitoring infectious diseases. Patients can opt out from this study if they do not want their data to be linked. |
| 7 | Security and confidentiality protection | Only named researchers will have access to the research dataset. Only the pseudonymised research dataset will be encrypted on transfer to the research team. The research dataset will be stored on a secure server at the University of Oxford. Access to the research data set will require a University of Oxford secure log-in. Remote access will not be permitted. |
| 8 | Not disclosed outside the EU | Research data will only be accessed by named researchers at the University of Oxford. It will not be transferred to any other individual. It will not be transferred to any other data storage device. |

 **Data linkage diagram**

Data linkage of PHE data using personal identifiers to HES, CPRD, IMD-10 and ONS mortality file

PHE laboratory data – collation of NTS, Campylobacter and VTEC

Data sent securely to HSCIC

Data stored on University of Oxford secure server for 5 years

Data sent securely to researchers at Uni. Oxford

Pseudoanonymised research data set including dummy unique ID

Personal identifiers removed from PHE, CPRD, HES, ONS and IMD-10

Linked PHE, CPRD, HES, ONS and IMD -10 with personal identifiers