

Survival analysis of a cohort of *Clostridium difficile* infected and non- infected patients admitted to Addenbrooke's Hospital between 2005 and 2007

Short Title: *C. difficile* Life Table Study (CDLTS)

9 July 2009

Principal Investigator:

Dr Mark Reacher

Consultant Epidemiologist, Health Protection Agency, East of England

Visiting Specialist in Epidemiology, Cambridge University Hospitals NHS Foundation Trust

Affiliated Lecturer in Epidemiology and Public Health, Cambridge University School of Medicine

Honorary Senior Lecturer, London School of Hygiene and Tropical Medicine

Co-Investigators:

Dr Nicholas Brown

Consultant Microbiologist and Head of Infection Control Addenbrooke's Hospital, Cambridge University Hospitals NHS Foundation Trust

Dr Mark Farrington,

Consultant Microbiologist and Director Department of Clinical Microbiology, Addenbrooke's Hospital, Cambridge University Hospitals NHS Foundation Trust

Ms. Cheryl Trundle

Head Infection Control Nurse, Addenbrooke's Hospital, Cambridge University Hospitals NHS Foundation Trust

Dr Philip Jones

Former Director, Ipswich Hospital, Department of Medical Microbiology

Mr Neville Verlander,

Statistician, Statistics, Modelling and Bioinformatics Department, Health Protection Agency, Centre for Infections, Colindale Avenue, London

1. Purpose and summary

To determine if *Clostridium difficile* (*C. difficile*) infection is associated with excess all cause mortality.

Survival of individuals with an episode of *C. difficile* infection at Addenbrookes Hospital will be compared with a population of patients who had no episode of *C. difficile* infection admitted to the same medical specialities, using record linkage of clinical records and UK national death registration and the statistical methods of survival analysis. Cox Proportional Hazard Models will be used to adjust for co-morbidities, treatment, smoking, alcohol consumption and Index of Multiple Deprivation Score derived from post code of residence at index admission.

2. Background to *C. difficile* infection

C. difficile Associated Disease (CDAD) ranges from apparently uncomplicated diarrhoea to shock, pseudo membranous colitis, toxic megacolon and death. Although the less frequent and severe complications of CDAD - pseudo membranous colitis and toxic megacolon, are easily recognised, there remains concern that overall mortality from all levels of severity of CDAD may be increased, and that the seriousness of this infection may continue to be under-estimated.

Between 44,563 and 55,658 episodes of *C. difficile* infection were reported in each of the years 2004 to 2007 declining to 40,704 in 2008, by Hospitals in England in compliance with the government's mandatory reporting scheme. Over 8,000 death certificates mentioned *C. difficile* in 2007 in residents of England and Wales (Office for National Statistics 28 Aug 2008).

Although the NHS has tolerated a high incidence of *C. difficile* infection, *C. difficile* is probably largely preventable by high quality nursing, infection control and by avoidance of unnecessary broad spectrum antibiotics. There is concern that *C. difficile* infection may be associated with increased mortality, but currently published data on this are unreliable and patients may be continuing to be put at unacceptable risk.

Approximately 95% of *C. difficile* cases are aged 50 years and older. Transmission of *C. difficile* is by direct and indirect faeco-oral transmission from person to person of spores of toxin producing strains. Diarrhoea illness generally occurs in individuals rendered susceptible by exposure to broad spectrum antibiotics, which suppress the normal gut flora and permit proliferation of *C. difficile* with toxin production. The incubation period for *C. difficile* associated diarrhoea is generally greater than two days, so that individuals with onset of diarrhoea greater than two days following an admission to a health care facility are generally regarded as having acquired infection within that Health Care facility, and if less than two days after admission, to have acquired their infection elsewhere.

3. Why a study of mortality associated with *C. difficile* is needed

There is concern that *C. difficile* infection may be associated with decreased patient survival, but this remains uncertain because published research so far is deficient in one or more of the following ways:

- difficulty in assigning a clear causative role for *C. difficile* infection in the complex chain of events leading to individual deaths
- absence of suitable reference populations for determining relative risk of death in infected compared to non-infected individuals
- Failure to accurately measure and adjust for co-morbidities, risk factors and treatment.

Precise measurement of mortality associated with *C. difficile* infection is needed to inform priorities for prevention and control.

To address the deficiencies in current knowledge and to provide up to date data relevant to the NHS, we intend to conduct an historic cohort study in a representative sample of patients who had an episode of *C. difficile* infection compared to a representative sample of non- *C. difficile* infected patients admitted to the same clinical specialties at Addenbrooke's Hospital in 2005, 2006 and 2007. We will compare the relative probability of survival in each group by linkage of clinical records and national death registration data bases, calculation of exact time from index admission to death, and use of survival analysis, including Kaplan Meier plots and Cox proportional hazards models, to make full adjustment of all cause mortality for patients' risk factors, co-morbidities, treatments, smoking, alcohol

consumption and Multiple Deprivation Score derived from post code of address of residence at index admission.

Co-morbidity and treatment data in individual patients will be obtained by clinical notes review using a data abstraction questionnaire and analysis method recently developed for a study of the clinical severity of *C. difficile* PCR ribotype 027: a case-case study. (Morgan O. et al 2008¹).

This study will provide, for the first time, an accurate measure of all cause mortality associated with *C. difficile* infection and will be generalizable to the UK and many health systems internationally.

4. Study setting and Generalisability

Addenbrooke's Hospital is a favourable site to conduct this project because:

- It is a large institution with a sufficient number of cases of *C. difficile* infection for recruitment to this study in each of the years 2005, 2006 and 2007.
- It's Hospital Infection Control Department has conducted consistent active surveillance throughout the hospital of *C. difficile* infections with complete recording in the departmental case register for many years, including the years 2005, 2006 and 2007.
- The Microbiology Laboratory used the Vero cytotoxin test for detection of *C. difficile* infection, which is regarded as the Gold Standard test, during 2005, 2006 and 2007.
- The Microbiology Laboratory Computer data base has complete data on all *C. difficile* test positive results taken in Addenbrooke's in patients.
- The clinical specialities to be studied have a case mix typical of District General Hospitals in the UK and similar hospitals in other high income industrialised economies.

Addenbrooke's Hospital purchases its clinical microbiology service from the Health Protection Agency, which shares premises with the other laboratory services comprising the Department of Pathology.

5. Pilot study

A pilot study was undertaken by Dr Philip Jones Consultant Microbiologist at a Hospital in the East of England of patients admitted to the Hospital in the East of England in 2006. This pilot demonstrated the feasibility and utility of record linkage between microbiology records, clinical notes and national death registration data (accessible by the local Primary Care Trust for patients resident in that county). Follow up of deaths of all *C. difficile* infected and a sample of non- *C. difficile* infected patients following discharge from the Hospital in the East of England was achieved.

The pilot showed death at one year in 45 of 79 (57%) of *C. difficile* cases compared to 68 of 287 (24%) in the sample of non-cases drawn from similar clinical specialties (Relative Risk 2.4). Co-morbidity and treatment were not measured and therefore not adjusted for in this analysis. The pilot raises the important question of whether this greater than two fold increase risk of all cause mortality associated with *C. difficile* infection observed in this pilot study is real.

6. Sample size considerations

Sample size considerations for this study are based on the results of the Hospital in the East of England pilot study of the proportion of subjects who had died at one year. This is valid because the biology of the infection and case mix at the Hospital in the East of England and Addenbrooke's hospitals in the clinical specialities to be studied, are similar.

The pilot study showed all cause mortality for non-cases was 68/287 (0.24) and for cases 45/ 79 (0.57) of *C. difficile* (a difference in proportion of 0.33) at one year.

A statistical sample size and power calculation was conducted by Mr. Neville Verlander, Statistician , at the HPA Centre for Infections, informed by the result of the Hospital in the East of England pilot study (Table 1)

The table shows the minimum number of individuals required in **each of two study groups of the same size** to detect a significant difference in proportion of outcome between the two groups with type 1 error , Alpha = 0.05 two tailed, and Power = 80% (Type 2 error = 0.20)

Table 1: Sample size calculations

Difference P2 – P1	0.1	0.15	0.2
P1			
0.1	219	112	71
0.15	270	133	82
0.2	313	151	91
0.3	376	175	103

P1 is the proportion outcome in Group 1, here deaths in the non-case reference group.

P2 is the proportion of deaths in Group 2, here the *C. difficile* infected group of cases.

P2 – P1 is the difference in proportion of deaths between Group 2(cases) and Group 1(non-cases).

Table 2 compares the results from the pilot study with those from the power calculation table

Table 2 Comparison of results from the pilot study and the sample size calculation in table 1

	Proportion in non-cases P1	Proportion in cases	Difference in proportion P2-P1	Number non-cases: Number of cases
Sample size table	0.20	0.40*	0.20	91: 91
Pilot study **	0.24	0.57	0.33	287 : 79
Sample size table	0.30	0.50*	0.20	103 : 103

* These values are not given in table 1

** These values are from the pilot study

The sample size table indicates that compared to the pilot study, a further study with 103 non-cases and 103 cases will be able to distinguish a statistically significant difference in death outcome which is smaller (0.2 compared to 0.33) and with a higher proportion of deaths in the non-case group (0.3 compared to 0.24) compared to the pilot study. The same difference of proportion dead of 0.20, would require a smaller number of non-cases and cases (91:91) if the proportion dead in non-cases was less (0.20).

The sample size calculation is conventional and in this context conservative in three respects

- a) It uses comparison of proportions with no consideration of precise time lived, whereas in the proposed study, precise time from date of index admission to date of death will be calculated for each subject and integrated comparing the proportion surviving over time by use of Kaplan Meier survival plots, a form of life table analysis.
- b) Follow up from index admission to review of death registration data will exceed one year in the proposed study compared to one year follow up from the pilot study used in the power calculation.
- c) The sample size calculation uses a two sided (two tailed) probability distribution for type 1 or Alpha error, which admits the possibility that the difference in outcome (death) could be more as well as less in non-cases than in cases.

Based on the results of the pilot study and the sample size calculations, we will therefore recruit 105 cases and 105 non-cases, because this enrolment will be adequate to demonstrate a statistically significant difference in the proportion of deaths of a little as 0.2 (a 20% difference) at one year. This enrolment will also probably be able to detect a statistically significant difference of less than 20% using survival analysis of non-cases and cases, because this uses measurement of individual time lived.

7. Methods

7.1 Summary of methods

We will undertake an historic cohort study of a random sample of cases of *C. difficile* infection and a random sample of patients with no episodes of *C. difficile* infection (non-cases) admitted for more than 48 hours to the same clinical specialties in Addenbrooke's Hospital between 1 January 2005 and 31 December 2007. We will compare all cause mortality between patients who have had an episode of *C. difficile* infection (cases) and the reference group of patients who did not have such an episode of infection (non-cases), by record linkage between hospital records and national death registration. Only cases who are UK residents at the time of admission will be included because death registration is directed to the UK resident population. The statistical methods of survival analysis will be used comprising Kaplan-Meier survival plots and Cox proportional hazards models for multivariable adjustment for risk factors including time between admission and the taking of the earliest specimen testing positive for *C. difficile* (Earliest specimen date and time \leq 48 hours versus $>$ 48 hours after admission), co-morbidities, surgery, antibiotic exposures, age and sex. Individuals for whom no record of death is located from national death registration will be treated as censored within survival analysis.

7.2 The study population

The study population will comprise individuals admitted to Addenbrooke's Hospital for more than 48 hours between 1 January 2005 and 31 December 2007 within the clinical specialties of General Medicine, Care of the Elderly and Orthopaedics, and whose home residential address on admission was within the UK.

7.3 Obtaining the list of patient admission episodes defining the study population (List A)

The population at risk is held in Addenbrooke's Hospital Patient Administration System (PAS) which is an electronic data base containing details of all in patient admission episodes. It is run by Addenbrooke's Hospital Informatics Department.

The informatics staff will use the PAS to select the subset of admission episodes to Addenbrooke's Hospital meeting the following criteria:

- Admission date between 1 January 2005 and 31 December 2007 inclusive
- Clinical directorate of admission General Medicine or Care of the Elderly or orthopaedics.
- Home address at admission within the United Kingdom.
- Duration of admission greater than 48 hours (Calculated from recorded date and time of discharge, transfer to another hospital or death) minus admission date and time >48 hours.

The Informatics department will download from the PAS a password protected file in Microsoft (MS) Excel file of the line listing of each of these admission episodes limited to the following variables

- First name
- Family name
- Addenbrooke's Hospital Number
- NHS Number
- Admission date,
- Age on admission (calculated in years from date of birth minus date of admission).

The file will be saved to CD and given to the lead investigator, who will take the disc and its file to the Health Protection Agency East of England Regional Epidemiology Unit (EEREU) located within the Cambridge University Institute of Public Health on the Forvie Site adjacent to Addenbrooke's Hospital. The MS Excel file will be copied to a password protected directory on the EEREU local area network. When this is done, the CD will be destroyed.

7.4 Case ascertainment and diagnosis of *C. difficile* infection at Addenbrooke's Hospital

Throughout the study period, standard operating procedures and policies were in place at Addenbrooke's Hospital for the early recognition, microbiological sampling and testing and control of infectious diarrhoea by the nursing staff, infection control team and clinical microbiology laboratory.

The Hospital Nursing staff were trained to be alert for possible cases of infectious diarrhoea, to apply enteric precautions in managing cases and to promptly send faecal specimens for microbiological testing. Each suspected case was reviewed by the Addenbrooke's Hospital Infection control team.

Specimens of faeces were taken and submitted to the Addenbrooke's Clinical Microbiology Laboratory. Faeces were tested for *C. difficile* using the Verocytotoxin assay, culture for enteric pathogenic bacteria and microscopy for *Cryptosporidium* spp and *Giardia lamblia*.

The Addenbrooke's Hospital Control of Infection team was responsible for diagnosing all episodes of *C. difficile* infection in Addenbrooke's Hospital in the study interval based on the clinical and laboratory criteria below.

7.5 Case definition of *C. difficile* at Addenbrooke's Hospital 2005 to 2007

- Diarrhoea stool observed by nursing staff and / or infection control team as liquid or semi liquid
- Three or more loose stools observed and recorded by Nursing staff in one or more 24 hour period
- One or more stool specimens test positive for *C. difficile* by the Vero cell cytotoxin assay and negative for enteric bacterial pathogens by culture, and negative for *Cryptosporidium* spp oocysts and *Giardia lamblia* by microscopy

7.6. Addenbrooke's Hospital Infection control team data base of *C. difficile* infected cases

C. difficile cases diagnosed by the Addenbrooke's Infection Control team, relevant clinical and microbiological data were entered into the Addenbrooke's Hospital Infection control team data base.

7.7 Selection of cases and non-cases for recruitment to the study

7.7.1 Summary

A random sample of patients who had an episode of *C. difficile* infection and an equal number of reference patients who did not have an episode of *C. difficile* infection during an admission episode to Addenbrooke's will be recruited to the study cohort. See also graphic "*Clostridium difficile* Life Table Study Data Flows Addenbrooke's Hospital.

7.7.2 Creation of a list of hospital admission episodes associated with being a case of *C. difficile*.

The Addenbrooke's Hospital Infection Control team data base of case patients diagnosed as having *C. difficile* infection (Described at 7.6) will be created from the main infection control data base by downloading the subset of cases and admission episodes meeting the criteria for the population at risk:

- Diagnosed by the infection control team as being a case of *C. difficile* infection
- admission within the Clinical Directorates of Medicine, Care of the elderly and orthopaedics
- admission dates between 1 January 2005 and 31 December 2007

7.7.3 Creation of a list of hospital admission episodes associated with a first episode of being a case of *C. difficile*.

The infection control team data base download of *C. difficile* infections described at 7.7.2) will be matched using Personal identifiers to the list of total admission episodes between 1 January 2005 and 31 December 2007 in the specialties of General Medicine, Care of the Elderly and Orthopaedics downloaded from the Patient Administration System (PAS) (List A).

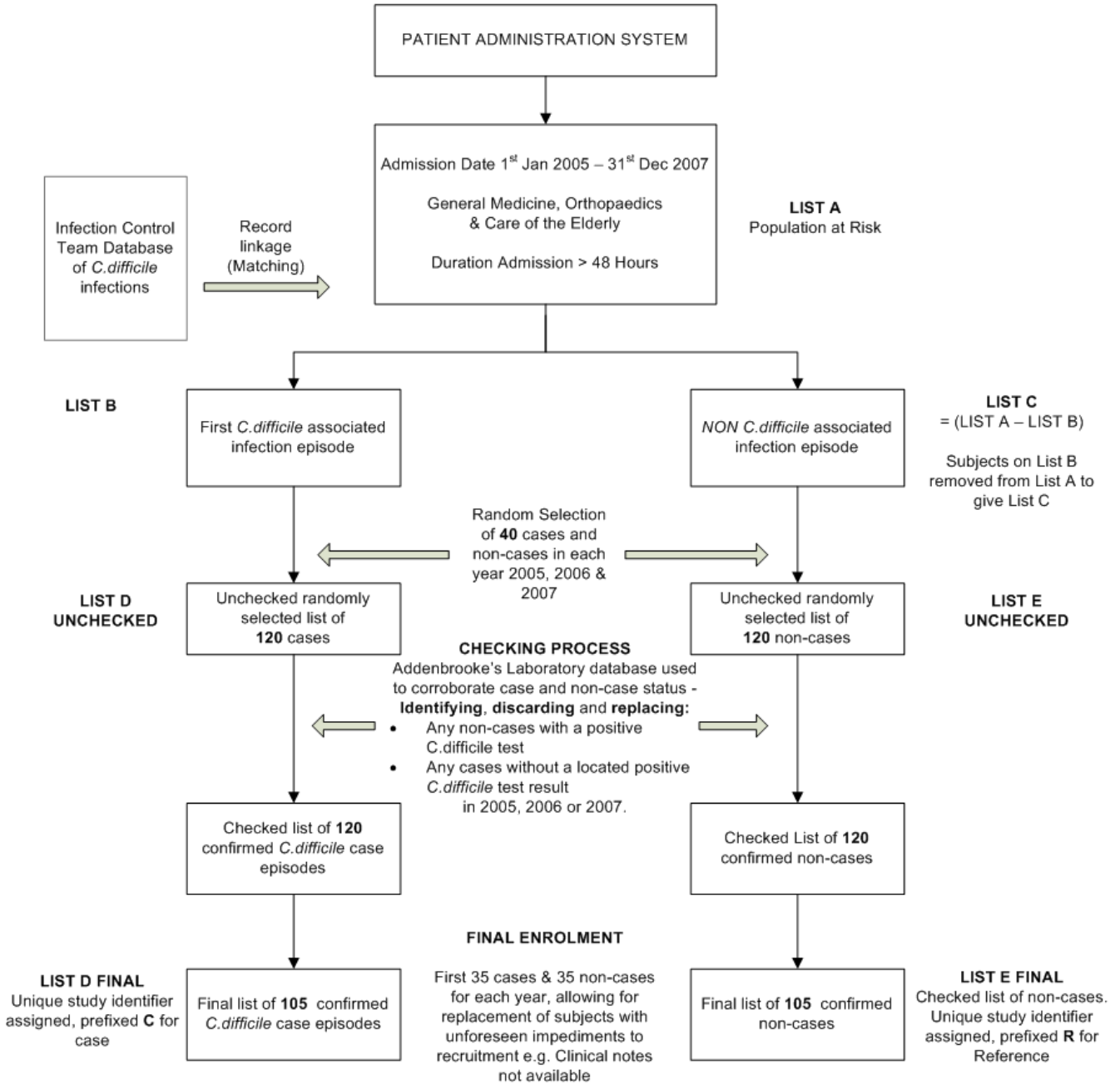
A search for multiple admissions with *C. difficile* in individual patients will be made. If this has occurred, second and subsequent episodes of *C. difficile* admission episodes will be deleted from the list to create a list of first *C. difficile* infected case admissions (List B).

7.7.4 Selection of *C. difficile* associated admission episodes from list B to create a random list of *C. difficile* infection associated episodes (list D)

List B will be sorted by admission date and divided into three sub lists of case admissions with admission dates during 2005, 2006 and 2007.

The sub list for each year will be sorted in ascending alphabetical order of first name of the case because this will ensure that the list order is unrelated to admission date. Thirty five case admission episodes will be randomly selected from each of the annual sub lists of admissions using a sequence of random numbers generated by a computer random number generator and algorithm.

Clostridium difficile Life Table Study Data Flows Addenbrooke's Hospital



The case admission records identified by the random numbers will be cut from the annual sub lists derived from B and pasted to a new list, list D which is the randomly selected list of case admissions for enrolment. List D will be checked using PII against the Hospital Microbiology Laboratory data base to identify the presence of a positive *C. difficile* laboratory test result for that patient during the index admission episode. If a positive test result cannot be located, that admission episode will be discarded and replaced by a new randomly selected case admission episode selected from the sublist for that year. This process will be repeated until a checked list of confirmed first *C difficile* associated infection episodes has been created (Checked list D). The list will be sorted by first name ensuring it is not ordered by admission date. Unique study reference numbers will be assigned to case records in final list D comprising the prefix "C" for Case, followed by a sequence of three numbers C001 to C120. This is final list D from which case admission episodes will be recruited.

The target recruitment is thirty five case admissions in each year, but over sampling to identify 40 admission episodes in each of the three years (Total 120 case episodes) will be undertaken to permit replacement in the event that a potential case admission cannot be recruited for some unforeseeable reason, such as failure to locate medical notes. Where replacements are undertaken, the next record in the random list will be used and the reason for replacement will be recorded. When 35 cases have been recruited for a year, no further cases will be recruited for that year. Recruitment to the remaining years will proceed working through final List D in sequence until 35 cases have been recruited from the list for each of the years 2005, 2006 and 2007.

7.7.5 Creation of a list of admission episodes not related to having acquired *C. difficile* infection: List C

Admission records associated with a diagnosis of *C. difficile* infection by the Addenbrooke's Hospital Infection Control team in the download from its data base to create list B, will be identified in the list of total admissions defining the population at risk (LIST A) and will be removed from list A to create a list of admissions in which no episodes of *C. difficile* infection diagnosed by the infection control team occurred. (List C)

7.7.6 Selection of a random sample of non-case admission episodes to create the list of non-cases for recruitment to the study (List E)

The list of non-case admission episodes (List C) will be ordered by year of admission and separated into three sub lists for each year, 2005, 2006 and 2007.

The sub list for each year will be sorted in ascending alphabetical order of first name of the patient. Thirty five non-case admission episodes will be randomly selected from each of the annual sub lists of admissions using a sequence of random numbers generated by new runs of the same computer random number generator and selection algorithm used for case selection.

Corroboration of being a non-case subject will be made by reference to the Addenbrooke's Hospital Microbiology Department data base. A search will be made in the Addenbrooke's Microbiology Laboratory data base for each of the randomly selected patient records believed to be from a non-case on the basis of non inclusion within the case list derived from the Infection control team data base. If a positive *C. difficile* test result is located for this patient subject, then this record will be deleted from the list of non *C. difficile* patient subjects and replaced by another admission episode randomly selected from the sublist of admissions for the same year. The list of checked randomly selected non case admissions for each year will be cut and pasted back into a single list and sorted on first name so it is not ordered by admission date to create the final list of checked list of non- *C. difficile* patient subjects, list E.

Unique study reference numbers will be assigned to non-case records in list E comprising the prefix “R” for Reference, followed by a sequence of three numbers R001 to R120.

The target recruitment is thirty five non-case admissions in each year, but over sampling to identify 40 non-case admission episodes in each of the three years (Total 120 non-case episodes) will be undertaken to permit replacement in the event that potential non-case admissions cannot be recruited for some reason, such as failure to locate medical notes. Where replacements are undertaken, the next record in the random list will be used and the reason for replacement will be recorded. When 35 non case admissions have been recruited for a year, no further non case admissions will be recruited for that year. Recruitment to the remaining years will proceed working through final List E in sequence until 35 non case admissions have been recruited from the list for each of the years 2005, 2006 and 2007.

7.7.7 Non-case definition

Non-cases are defined as individuals in the study population at risk for which no record of having been a case of *C. difficile* infection could be identified in the Addenbrooke’s infection Control team data base, and no record of a positive *C. difficile* test result could be identified in the Addenbrooke’s Hospital Microbiology Laboratory data base for the index admission episode.

8. Recruitment and data abstraction

8.1 Summary

The hospital in patient notes will be obtained of study subjects from the medical records department and reviewed in a designated area within the Addenbrooke’s Hospital clinical microbiology Department.

8.2 Abstraction of clinical information to Questionnaire 1.

The clinical data abstraction questionnaire, Questionnaire 1. is provided as a separate file to this protocol. Questionnaire 1. will be completed using all available elements of the patient records including clinical notes, nursing notes, admission and discharge summaries and drug charts, radiology and laboratory results. Questionnaire 1. records age, sex, smoking history, alcohol consumption history, medical conditions and treatments during index admission. Primary and subsidiary diagnoses will be coded at the time of analysis into the major diagnostic categories as used in the CDAD 027 versus non-027 ribotyping study¹ as shown in the table.

Diagnostic Group	Code
Cardiovascular	1
Respiratory	2
Gastrointestinal	3
Urinary & renal	4
Genital	5
CNS	6
Psychiatric	7
Metabolic	8
Trauma	9
Malignancy	10
Skin	11
Musculo- skeletal	12
Unspecified	13
Diabetes	14

Antibiotic prescribed during and in the 8 weeks (56 days) prior to admission will be recorded and categorised as initiating, except metronidazole and oral glycopeptides which will be categorised as protective and curative. Probiotics exposure will also be recorded. Other drug treatments, immune suppression, drugs suppressing gastric acid, gastric intubation and features of *C. difficile* Associated Disease (CDAD) including evidence of shock, fever, pseudo membranous colitis, toxic megacolon and colectomy will also be recorded (Questionnaire 1.). Index of Multiple Deprivation 2007 Score (IMD2007) will be assigned each subject on the basis of post code of residence at the time of Index Admission.

The case or non-case status of patients whose notes are reviewed, will be masked to the notes reviewers as far as possible, by taking care to avoid external labelling, ordering and placing notes in a manner from which non-case or case status could be inferred, by collection of general medical and drug information first and leaving data fields which indicate case or non-case status for collection at the end of Questionnaire 1.. (See also the document “Instructions and notes on completing, transporting and storing sections one and sections two to five of Questionnaire 1.”). The investigators and medical microbiology trainees who will undertake data abstraction are experienced in such work and have extensive clinical and microbiological experience.

PII from page 1 questionnaire 1 will be entered into Data base 1. Clinical and mortality data will be entered into Data base 2. Both data bases are password protected encrypted EPIDATA data bases and can be linked in Random Access memory using the unique study reference number as key for data entry and data cleaning. This is discussed further in the section on data security (section 10) below.

9. Long Term Follow up for registered deaths in the study cohort by the Medical Research Information Service of the National Health Service Information Centre (NHS-IC)

The Medical Research Information Service is part of the National Health Service Information Centre. (NHS-IC). Application to NHS-IC to link data and provide mortality data is approved by the National Information Governance Board (NIGB). A search for registered deaths of study subjects and a copy of the Death Certificate will be requested for study subjects from the NHS-IC in accordance with NHS-IC standing procedures.²

A copy of Data Base 1 will be prepared in Excel and encrypted and copied to a CD and sent by secure courier to the NHS-IC.

This file will contain the Personal Identifying Information collected on page one of Questionnaire 1 comprising full forename, full family name, date of birth and NHS number.

NHS-IC will be requested to undertake **Long Term Follow Up** of study subjects' mortality experience by matching study subjects with the National Death Register using NHS-IC standard procedures based on full family name, full forename, date of birth and NHS number.³

In Long Term Follow Up, matching will be sought for all subjects within the cohort and the National Death Register on an ongoing basis by NHS-IC and a summary of newly registered deaths within the cohort will be made by NHS-IC to the EEREU each quarter year.

NHS-IC will send reports of individual subject deaths including date of death and a copy of the death certificate showing registered causes of death in a secure electronic format to Dr M. Reacher, Lead Investigator, CDLTS, East of England Regional Epidemiology Unit, Institute of Public Health, Cambridge, CB2 0SR. The NHS-IC reports may be abstracted by study staff using Questionnaire 2, for entry into Data Base 2.

Because NHS-IC will be regularly searching the National Death Register for study subjects, the date of the most recent report of study subject long term follow up by NHS-IC to EEREU will be the end point of observation of the cohort in life table analyses. The primary and secondary causes of death in the death certificate will also be entered and analysed.

10. Security of confidentiality in handling Patient Identifiable Information (PII)

Patient records will be handled in strict medical confidence at all times and in a manner compliant with Caldicott Standards for the handling of PII.

Because this study is a review of existing data sources, no direct contact with patients, carers or clinical providers will be required at any time and such contacts will not be made.

PII is required in this study for record linkage (also termed record matching) between the following data sources

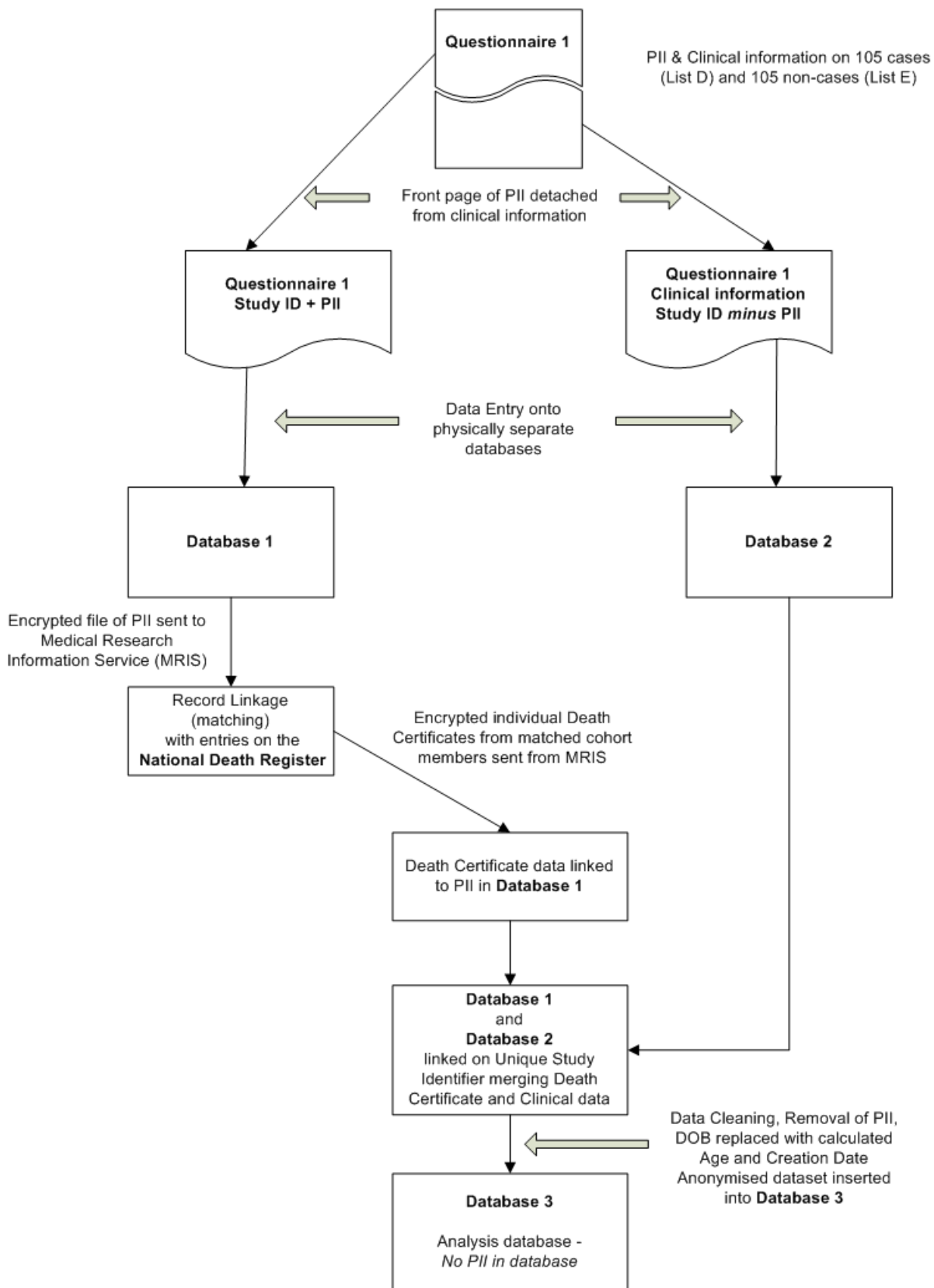
1. The admissions data base of the Patient Administration System (PAS)
2. The Infection Control Team data base
3. The Addenbrooke's Microbiology Laboratory data base
4. The National Death Register.

The first page of Questionnaire 1 contains the data fields for PII and the Unique Study Reference Number and is input into Data Base 1.

The remaining pages of Questionnaire 1 contain data from clinical notes review and laboratory test results and are input into Data Base 2. Data Base 2 also contains death registration data derived from Long Term Follow Up of the Study Cohort by the NHS-IC Medical Research Information Centre (MRIS). NHS-IC reports of matches of registered deaths will be made to the study coordination centre at the EEREU and may be summarised by study staff on questionnaire 2 for Data entry into Data Base 2.

See the figure "Clostridium difficile Life Table Study Record Linkage of study cohort members with the National Death Register and Study centre."

***Clostridium difficile* Life Table Study** **Record Linkage of Study Cohort Members with** **the National Death Register &** **Study Centre**



PII will be kept separated from clinical information in paper and in electronic records except when brought together by investigators when working on the conduct of the study within Addenbrooke's Clinical Microbiology Laboratory, The Office for the Addenbrooke's Hospital Infection Control Team and the Health Protection Agency East of England Regional Epidemiology Unit (EEREU) located within the Cambridge University Institute of Public Health (IPH) adjacent to the Addenbrooke's site. All three Sites and Departments are Caldicott compliant in handling PII. The HPA Centre for Infections Statistic Unit is also Caldicott compliant, but will not receive PII which is absent from Data Base 2.

On completion of Questionnaire 1 in the Microbiology Laboratory and the Infection Control team office, the PII containing first page of Questionnaire 1 will be detached from the second part of questionnaire 1, which contains clinical and microbiological data.

The PII containing first pages will be transported by a member of the study team from the Addenbrooke's Clinical Microbiology Laboratory to the EEREU.

Within the EEREU, the information contained in the PII containing first page of Questionnaire 1 will be entered into a password protected EPIDATA data base - Data Base One, held in a password protected folder on the EEREU secure Local Area Network. The paper copies will be stored in a locked filing cabinet within the EEREU.

Sections 2 to 5 of Questionnaire 1 containing clinical and microbiological information will be transported to, and stored separately from, Section 1 in the EEREU. The data from sections 2 to 5 of Questionnaire 1 will be entered into an EPIDATA data base, Data Base 2 which will be stored in a directory separate from the PII containing file Data Base 1.

Data Base 2 also contains the Data fields for Death Registration reports made from Long term follow up by the NHS-IC. NHS-IC reports to EEREU may be summarised by study staff in the EEREU on Questionnaire 2 to assist entry into Data Base 2.

Data Base 1 containing PII and Data Base 2 containing clinical and mortality information, will only be linked in Random Access Memory using the unique study number as the key, during data entry and data editing. The Age of subjects at study admission will be calculated in Data base 1 from date of admission and date of birth.

Sex and the calculated age at specified dates will be determined from date of birth, but date of birth itself, will be copied to Data Base 2.

Data Base 3 is the data base for the epidemiological analysis; it is a copy of the cleaned and checked version of Data base with calculated age at admission and calculated age at death replacing date of death so that it holds no PII.

The completed first page of Questionnaire 1 with PII will be retained by EEREU until one year after completion of the baseline study analysis in 2010, that is by 31 December 2012 then shredded by a confidential office shredding service.

The Personal Identifying Information on page 1 of Questionnaire 1 will be entered into Data Base 1. Long term follow up of survival of the cohort is planned at five years following the baseline analysis in 2010, that is in 2015.

At the five year long term follow up date Data Base 2 will be updated with death registration reports made by long term follow up of the cohort by the NHS-IC.

A decision will be taken in 2014 if a further follow up to 2020 is required. The NHS-IC will be informed immediately of the decision to terminate long term follow up of the cohort. When the survival analysis for the final follow up round has been undertaken, Data Base 1 containing PII will be deleted at the EEREU and the copy of the PII from this held by NHS-IC will also be deleted. These data bases will be deleted by the end of the year following the last follow up round. That is by 31 December 2016 for the 2015 follow up round and by 31 December 2021 if a 2020 follow up round was agreed. Data base 2 will be retained for the same duration. Fully anonymised versions, of data base 3, which contain no PII will be retained and may be made available for training and teaching purposes.

11. DATA handling and analysis

Data will be double entered and checked for inconsistencies and corrected using the validate command in EPIDATA.

The final cleaned version of Data Base 3 (which holds no PII) will be held in EEREU and a copy sent by secure courier on CD to the study statistician, Mr Neville Verlander at the Statistics Unit HPA Cfi London where the statistical analysis will be undertaken.

Data Base 3 will be exported into STATA for analysis.

The person time of observation for survival analysis of subjects will be from index admission date to Addenbrooke's Hospital to the date of registered death. Where no record of death by matching undertaken by NHS-IC to the National Death Register is observed, that individual will be censored within the survival analysis.

Survival of cases and non-cases will be compared in Kaplan Meier survival plots of the proportion of the cases and non-cases surviving from the date of index admission to the registered date of death. The statistical significance of difference in the uncorrected survival curves will be tested by log rank test. The impact of co-morbidities, risk factors and treatments on survival will be investigated using Cox proportional hazards models. Single variable analysis will be undertaken in a proportional hazards regression approach with one explanatory variable at a time. Stepwise model building will then be undertaken with removal of non-significant variables. Age and sex will be retained in all iterations regardless of significance. Antibiotics prescribed in the 8 weeks prior to admission and during admission will be categorised as initiating except metronidazole and oral glycopeptides which will be categorised separately as protective. Probiotic ingestion will also be included in the models.

After publication of the results of this study, Copies of Data Base 3 may be made available to other workers for teaching purposes.

12. Ethical considerations

The study will answer accurately, for the first time: Is *C. difficile* infection associated with shortened life expectancy?

Shortened life expectancy associated with *C. difficile* infection is possible and a 2.4 fold higher risk of death at one year in *C. difficile* infected compared to non infected patients was observed in a pilot study at the Hospital in the East of England in 2007.

A sound measure of *C. difficile* associated excess mortality is needed to inform the priority that *C. difficile* prevention and control is afforded within the UK National Health Service and internationally, including innovations such as improved drug treatments and vaccines.

The study will use clinical and vital registration data which already exist and may be considered an audit of patient outcomes.

The study does not require contact with subjects, their relatives or medical attendants.

The process of enrolment will not compromise the medical confidentiality of subjects nor adversely affect their medical or social care at this or any future time.

Subject records will be handled in complete confidentiality in compliance with Caldicott standards for the handling PII.

The study data set will be fully anonymised.

Ethical committee approval will be sought.

13 Governance issues

The Investigators involved in patient data abstraction (Reacher, Brown, Farrington, Jones and Trundle) will be assisted by trainees in Medical Microbiology at the Addenbrooke's Hospital Department of Microbiology and by nurses working in the Addenbrooke's Hospital Infection Control team. All deal with data collected for this study as a matter of routine within their job description and contracts with the Cambridge University Hospitals NHS Foundation Trust.

14. Duration of the study

The baseline study is anticipated to take 18 months from obtaining ethical approval to publication in a leading peer review journal.

15. Funding and staff time

Dr Mark Reacher will devote 15 % WTE for year one and year two.

Dr Nicholas Brown and Dr Mark Farrington will devote 5% WTE time in year 1

Ms Cheryl Trundle are will devote 5% WTE time in year 1.

Dr Philip Jones will devote 5% WTE for 6 months in year 1.

Mr. Neville Verlander, study statistician will devote 10% WTE in year one and 50% WTE during statistical analysis of the study, which will take approximately two months.

Dr Gareth Hughes, HPA Health Care Associated Infections Epidemiological Scientist, will devote 25% WTE to the study in year one.

Specialist Registrar Trainees in Medical Microbiology with contracts at Cambridge University Hospitals NHS Foundation Trust may be nominated by Dr. Nicholas Brown and Dr. Mark Farrington to join this project as required.

It is anticipated that this study will be undertaken within existing HPA and NHS core funding.

16. Publishing Intention

Peer reviewed article in a leading UK or international medical journal

Reference

¹ Morgan OW, Rodrigues B, Elston T, Verlander NQ, Brown DF, Brazier J, Reacher M.
Clinical severity of *Clostridium difficile* PCR ribotype 027: a case-case study
PLoS ONE. 2008 Mar 19;3(3):e1812 <http://www.ncbi.nlm.nih.gov/pubmed/18350149?dopt=Citation>

² NHS-IC MRIS Information sheet. How to send us your data. <http://www.ic.nhs.uk/services/medical-research-information-service/what-we-can-do-for-you/how-to-send-us-your-data>

³ NHS-IC MRIS Information Sheet The Matching Process. <http://www.ic.nhs.uk/services/medical-research-information-service/what-we-can-do-for-you/the-matching-process>