A study of infectious intestinal disease in England: plan and methods of data collection

D Sethi, JG Wheeler, JM Cowden, LC Rodrigues, PN Sockett, JA Roberts, P Cumberland, DS Tompkins, PG Wall, MJ Hudson, PJ Roderick

Summary: The Committee on the Microbiological Safety of Food, set up in 1989 by the Department of Health in response to national epidemics of foodborne infection, considered the available evidence and commissioned a study of infectious intestinal disease (IID) in England. Seventy practices (with 489,500 patients overall) recruited from the Medical Research Council's General Practice Research Framework between August 1993 and January 1995 collected data for one year. The practice populations were representative of practices in England by area and urban/rural location, but with fewer small and affluent practices. There were five main components. i) A population cohort of 9776 (40% of those eligible) were enrolled to estimate the incidence and aetiology of IID in the community, and a large proportion were followed up. A median of 10% of patients on practice age-sex registers had moved away or died. ii) A nested case control component based on cases ascertained in the cohort was used to identify risk factors for IID in the community. iii) In a case control component used to identify risk factors and to estimate the incidence and aetiology of IID presenting in 34 general practices 70% of the 4026 cases returned risk factor questionnaires, 75% submitted stools, and matched controls were found for 75% of cases. iv) An enumeration component was used to estimate the incidence of IID presenting to general practitioners (GPs) in 36 practices and the proportion of specimens sent routinely for microbiological examination. v) In a socioeconomic costs component used to estimate the burden of illness of IID in the community and presenting to GPs 63% of those who returned a risk factor questionnaire also returned a socioeconomic questionnaire and were representative by age, sex, and social class.

Despite variable enrolment and compliance the study sample had sufficient power for the multivariable analysis. The characteristics associated with low enrolment and compliance must be considered in the interpretation of the main study results.


Key words: case control studies
cohort studies
compliance
diarrhoea
family practice
gastrointestinal diseases
incidence
intestinal diseases
methods
risk factors

Introduction

Increases in notifications of food poisoning and reports of cases of campylobacter, salmonella, and listeria infection, and national epidemics of foodborne infections with organisms such as Salmonella enteritidis led the Secretary of State for Health and Minister of Agriculture, Fisheries and Food to set up the Committee on the Microbiological Safety of Food in 1989 (‘Richmond Committee’)1-4. The committee reported that all indices of infectious intestinal disease (IID) had increased between 1980 and 19901 and recommended that a representative sample of all cases of IID in England should be studied to identify causal organisms, estimate the true incidence of disease and its relationship to reported cases, and determine the potential risk factors associated with the acquisition of IID and the costs incurred by individuals and by society. In response the Department of Health commissioned a study of IID in England whose methods are described in this paper.

A pilot study carried out in 1991 and 1992 tested the feasibility of the design, established the basis for the sample size calculations, and compared options...
for selection and follow up of subjects in general practice\(^5\). The degree of underascertainment of cases presenting with IID in this setting has been discussed\(^6\). The main study began in August 1993 and data were collected until January 1996. The following organisations shared responsibility for the main study: the PHLS, including its Communicable Disease Surveillance Centre (CDSC), Leeds Public Health Laboratory (PHL), and PHLS reference laboratories for specific organisms, the Centre for Applied Microbiology and Research (CAMR), the Medical Research Council Epidemiology and Medical Care Unit (EMCU) and the General Practice Research Framework, and the London School of Hygiene and Tropical Medicine (LSHTM).

**Methods**

**Setting: stratification and selection of practices**

The General Practice Research Framework is a network of over 800 practices committed to research, that care for about 10% of the registered national population. Based on sample size calculations (described below), a total of 70 volunteer practices were selected from the framework. The country was divided into three areas of similar population sizes for comparison: i) North (former health regions of Northern, Yorkshire, North Western, and Mersey), ii) Midlands and South West (East Anglia, West Midlands, Trent, South Western, and Wessex), and iii) South East (Thames Regions). The number of practices selected in each area was in proportion to the area’s total population according to the 1981 Census. Practices were selected to represent each area’s socioeconomic and urban or rural characteristics. tertiles of the population distribution of ward-based Jarman deprivation scores\(^7\) were used to stratify the area’s socioeconomic costs component to estimate the burden of illness of IID occurring in the community and presenting to GPs.

Practice recruitment was staggered between August 1993 and January 1995. Each practice collected data for one year from recruitment. Each practice dedicated a part time research nurse to the study. All 70 practices took part in the cohort component and were randomised within each stratum to take part in either the GP case control or the enumeration component (see below).

**Population cohort with nested case control component**

Two consecutive cohorts were each followed up for six months. For each cohort, 200 people selected by stratified random sampling by age and sex from the age-sex registers of 70 practices were invited to take part. The notes of people who the nurse could not contact were searched at three months to determine whether they had moved away or died. Participants attended a briefing interview, completed a baseline questionnaire, and were given a stool collection kit. They posted diary cards each week to the nurse, to declare that they had no symptoms of IID that week. Those who developed symptoms (incident cases) were asked to contact the nurse, fill in a risk factor questionnaire, and submit a stool specimen. A control matched for age and sex was selected systematically from the cohort for the nested case control component and asked to submit a stool specimen and fill in a risk factor questionnaire\(^5\). Two consecutive cohorts of six months rather than one of 12 months were used to increase the level of participation\(^5\). Age and sex strata with poor responses in the first cohort were oversampled in the second.

**General practice case control component**

Cases who presented in the surgery or on home visits were ascertained in 34 practices over one year. Those who fulfilled the case definition (box) were contacted by the nurse and asked to fill in a risk factor questionnaire and submit a stool specimen. Out-of-hours deputising agencies were contacted for details of consultations for IID. For each case up to five sex and age matched controls selected systematically from the register were invited in sequence to attend until one accepted. If the first potential control refused then a second was invited, and so on. Controls were asked to fill in a risk factor questionnaire and submit a stool specimen.
Case definition
People of all ages with loose stools or significant vomiting (more than once in 24 hours, incapacitating, or accompanied by cramps or fever) lasting less than two weeks, in the absence of a known non-infectious cause and preceded by a symptom free period of three weeks. Exclusions: people with non-infectious causes of diarrhoea - such as Crohn’s disease, ulcerative colitis, cystic fibrosis, and coeliac disease - and non-infectious causes of vomiting such as surgical obstruction, alcohol intoxication, morning sickness, infant regurgitation.

Control definition
People free of loose stools or significant vomiting for three weeks before the matched case became ill. Controls were matched to cases by age (within age group for children aged 0 to 5 months, within age group for children aged 6 to 11 months, within one year for children aged 1 to 4 years, but not below 11 months, within two years for those aged 5 to 19 years, but not below 4 years, and within five years for adults, but not below 18 years). Controls for cases over 5 years of age were also matched for sex.

Definition of IID
People of all ages with loose stools or significant vomiting (more than once in 24 hours, incapacitating, or accompanied by cramps or fever) lasting less than two weeks, in the absence of a known non-infectious cause and preceded by a symptom free period of three weeks. Exclusions: people with non-infectious causes of diarrhoea - such as Crohn’s disease, ulcerative colitis, cystic fibrosis, and coeliac disease - and non-infectious causes of vomiting such as surgical obstruction, alcohol intoxication, morning sickness, infant regurgitation.

Regional definitions
Two cases were considered to be in the same area if they were in the same practice. The regional definitions of IID were calculated for all practices in the area and for those receiving a random sample of cases and controls. Two regions were compared, the north of England and the south of England. The north included North West, North East, Yorkshire and Humberside, and Yorkshire and the Humber. The south included the remaining regions of England, excluding East Anglia.

Sample size calculations
The sample size was selected in order to estimate the incidence of IID in England with a precision of 10% on each side (95% confidence interval), to estimate the incidence in each of the three geographical areas with a precision of 20%, and to ascertain the organisms present in at least 25% of cases. This was based on the results of the pilot study, which estimated a community incidence of 13.8/100 person years, a presentation rate of 2.13/100 person years, a positive stool rate of 0.108/100 person years, and a compliance of about 75%. The sample size estimates suggested that 70 practices in the cohort study with two six-month cohorts of 90 people each would yield 6300 person years of follow up, and that 35 practices in each of the case control and enumeration components followed up for one year would yield about 280 000 person years of follow up. Within this sample it was calculated that the numbers in the GP case control analysis would be large enough to detect an odds ratio of 2 at the 1% level for exposures in at least 25% of cases. The baseline cohort questionnaire sought sociodemographic, accommodation, and food handling details. Non-respondents were sent a questionnaire that asked about family size, social class, and reasons for refusal to take part. The socioeconomic questionnaire asked about household composition, income, the impact of the illness in terms of consequent use of resources by individuals, their families, the NHS, and absence from work, and willingness to pay to avoid illness. Questionnaires were modified for child cases and controls.

Data management
Staff of the EMCU and the LSHTM coordinated data collection and management. Data were coded, double entered using Epi-Info, validated, and monitored (manually or by computers) at every stage for irregularities. Practice performance was monitored and progress reports were returned to nurses at six weeks and subsequently every three months. The reports compared the actual number of cases and controls monitored with the numbers predicted by the pilot study. Regional training nurses made scheduled visits to each practice, investigating any underperformance. Overall study performance was monitored by the IID Executive Committee, which met every three months.

Sample size calculations
The sample size was selected in order to estimate the incidence of IID in England with a precision of 10% on each side (95% confidence interval), to estimate the incidence in each of the three geographical areas with a precision of 20%, and to ascertain the organisms present in at least 25% of cases. This was based on the results of the pilot study, which estimated a community incidence of 13.8/100 person years, a presentation rate of 2.13/100 person years, a positive stool rate of 0.108/100 person years, and a compliance of about 75%. The sample size estimates suggested that 70 practices in the cohort study with two six-month cohorts of 90 people each would yield 6300 person years of follow up, and that 35 practices in each of the case control and enumeration components followed up for one year would yield about 280 000 person years of follow up. Within this sample it was calculated that the numbers in the GP case control analysis would be large enough to detect an odds ratio of 2 at the 1% level for exposures in at least 25% of cases.
TABLE 1  Study and regional (mid 1994) population distribution by area (North, Midlands and South West, and the South East), and by Jarman score tertiles

<table>
<thead>
<tr>
<th>Area</th>
<th>Study practices</th>
<th>Country</th>
<th>Population (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Population</td>
<td></td>
<td></td>
</tr>
<tr>
<td>North</td>
<td>128120 (29)</td>
<td>132542000 (27)</td>
<td></td>
</tr>
<tr>
<td>Midlands and South West</td>
<td>218788 (44)</td>
<td>212287000 (44)</td>
<td></td>
</tr>
<tr>
<td>South East</td>
<td>148758 (30)</td>
<td>140498000 (29)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>495666</td>
<td>48532700</td>
<td></td>
</tr>
<tr>
<td>Jarman</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;5 Low</td>
<td>110172 (22)</td>
<td>142236300 (30)</td>
<td></td>
</tr>
<tr>
<td>-5 to 10 Mid</td>
<td>201813 (41)</td>
<td>148705820 (32)</td>
<td></td>
</tr>
<tr>
<td>&gt;10 High</td>
<td>183681 (38)</td>
<td>179609920 (38)</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>495666</td>
<td>48532700</td>
<td></td>
</tr>
</tbody>
</table>

Results

Representativeness

GP practice population characteristics

The study practice population of 495660 accounted for 1% of the English population and was representative of the three geographical areas (table 1). The lower tertile (least deprived) of Jarman scores was slightly under-represented in the study practice population (22%) when compared with England as a whole (30%). A lower proportion of the study population attended GP practices with four or fewer partners (43%) than did the population of England (58%). The study practice population and the English population were similar as regards age and sex and urban/rural location (62% urban, 61% rural).

Community cohort

Forty-five per cent of the study population enrolled were male compared with 49% of the national population (table 2, ONS mid-1994 estimate). Enrolment was proportionately lowest in the 15 to 24 age group when compared with the national population. Fifty-six per cent of the cohort were from all 61 local research ethics committees. Written informed consent was obtained from cases and controls by each practice research nurse.

Enrolment in the community cohort

A total of 27 651 people were invited to take part in the cohort study, 35% of whom (9776) enrolled and 24% (6686) declined (figure 2). Together these 16 462 people (59% of the total invited) constituted the total known to be eligible for enrolment. Eight per cent (2177) of the remaining 11 189 were known to be ineligible because they were no longer registered with the general practice, having moved away or died. The nurse could not contact 3844 (14%) of the remaining 9012 (33%) people, and the reason for non-enrolment was not obtained for the remaining 5168 (19%): (the forms of 577 were lost and for 4591 the reason was not recorded).

Ineligibility was higher among males and those aged 15 to 34 years. Sixty-one per cent of refusers provided sociodemographic details. Twenty-five per cent of refusers were economically inactive, compared with 9% of those who enrolled.

Estimate of practice list inflation

Practice records of 3007 of the 3844 people (figure 2) not contacted were searched and 430 (14%) were found to be ineligible, having moved or died. Ten per cent of the 1238 of the 4591 people whose reason for non-enrolment was not recorded and whose notes were searched were ineligible. We assumed that 10% of the 577 people whose forms were lost would also be ineligible. A total of 3252 were estimated to be ineligible (figure 2), a crude proportion of 12%.

The multiple regression model applied to the 3007 people not contacted showed that ineligibility was commonest in areas with a mid-tertile Jarman score, rural areas, and urban areas in the South East. This was used to predict the proportion ineligible by practice. A median of 10% of those invited were ineligible (5th and 95th centiles 5 and 24.7). The median adjustment used to correct for list size inflation was therefore 90%. This corrected list size was then used as the denominator to calculate the presentation rates of IID in the case control and enumeration components.

The overall enrolment after correcting for ineligible patients on practice lists was 40%.

Compliance

GP case control component

A median of 70% of cases returned questionnaires and a median of 75% submitted stool specimens (table 3). A median of 50% of potential controls agreed at the first invitation but manual classes compared with 65% in England (1991 census). The population was similar in ethnic composition to the national population (1991 census). Seventy-two per cent of the cohort were married compared with 58% nationally and 15% were single compared with 26%. A smaller proportion of the cohort were economically active (58% vs 61%) and a larger proportion had retired from work (23% vs 19%). A larger proportion were homeowners (80% vs 72%).

TABLE 2  Age and sex distribution of community population taking part in cohort study, compared with population of England (mid 1994 estimate)

<table>
<thead>
<tr>
<th>Age (yrs)</th>
<th>Cohort population</th>
<th>Population of England in thousands</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Male (%)</td>
<td>Female (%)</td>
</tr>
<tr>
<td>&lt;1</td>
<td>48 (1)</td>
<td>47 (1)</td>
</tr>
<tr>
<td>1-4</td>
<td>254 (6)</td>
<td>277 (6)</td>
</tr>
<tr>
<td>5-9</td>
<td>360 (9)</td>
<td>382 (8)</td>
</tr>
<tr>
<td>10-14</td>
<td>315 (8)</td>
<td>298 (6)</td>
</tr>
<tr>
<td>15-24</td>
<td>289 (7)</td>
<td>396 (8)</td>
</tr>
<tr>
<td>23-34</td>
<td>469 (11)</td>
<td>719 (14)</td>
</tr>
<tr>
<td>35-44</td>
<td>519 (12)</td>
<td>718 (14)</td>
</tr>
<tr>
<td>45-54</td>
<td>652 (16)</td>
<td>741 (15)</td>
</tr>
<tr>
<td>55-64</td>
<td>544 (13)</td>
<td>618 (12)</td>
</tr>
<tr>
<td>65-74</td>
<td>479 (12)</td>
<td>503 (10)</td>
</tr>
<tr>
<td>75+</td>
<td>228 (5)</td>
<td>291 (6)</td>
</tr>
<tr>
<td>Totals</td>
<td>4157 (45)</td>
<td>4990 (55)</td>
</tr>
</tbody>
</table>

TABLE 1  Study and regional (mid 1994) population distribution by area (North, Midlands and South West, and the South East), and by Jarman score tertiles
FIGURE 2 Numbers who enrolled, refused, and estimated to be eligible for cohort component practices

<table>
<thead>
<tr>
<th>Known to be eligible</th>
<th>Not known if eligible or not</th>
<th>Known to be ineligible</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enrolled</td>
<td>Refusers</td>
<td></td>
</tr>
<tr>
<td>9776 (35%)</td>
<td>6686 (24%)</td>
<td></td>
</tr>
<tr>
<td>No contact (3844) +</td>
<td>Reason not recorded (4591) +</td>
<td>Unknown* (577)</td>
</tr>
<tr>
<td>9012 (33%)</td>
<td>9012 (33%)</td>
<td>2177 (8%)</td>
</tr>
</tbody>
</table>

Estimated eligible 7937
Estimated ineligible 1075

Total estimated eligible 24399
Total estimated ineligible 3252

*Unknown: forms lost at two practices

the proportion declined to 37%, 27%, 18%, and 13% for the second, third, fourth, and fifth invitations, respectively. Higher proportions of controls than cases completed questionnaires and submitted stool specimens. Compliance of both cases and controls was lowest among males and the 15 to 24 year age group. Regression modelling showed that high Jarman score, urban location, and location in the South East were the practice characteristics associated with lower questionnaire and stool compliance for cases and control, and poorer case control matching.

Seventy-six per cent (5254/6897) of subjects in the case control component submitted stools for analysis. More than 90% of all specimens were tested to stage 5 (see table 1, page 109). A third of specimens were received for analysis within four days of onset of illness and 63%, 77%, and 85% within 7, 10, and 14 days, respectively. Ninety-five per cent of specimens were received within four days of voiding.

<table>
<thead>
<tr>
<th>Proportion (%) returned per practice</th>
<th>Number returned</th>
<th>Proportion (%) returned per practice</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GP case control component</td>
<td></td>
</tr>
<tr>
<td>Cases (of 4026)</td>
<td>2642</td>
<td>70 40 100</td>
</tr>
<tr>
<td>Risk factor questionnaires</td>
<td>2962</td>
<td>75 54 100</td>
</tr>
<tr>
<td>Stools submitted</td>
<td>1652</td>
<td>43 17 61</td>
</tr>
<tr>
<td>Socioeconomic questionnaire</td>
<td>2429</td>
<td>88 68 100</td>
</tr>
<tr>
<td>Controls (of 2871)</td>
<td>2392</td>
<td>85 56 92</td>
</tr>
<tr>
<td>Stools submitted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk factor questionnaires</td>
<td>648</td>
<td>81 – 100</td>
</tr>
<tr>
<td>Stools submitted</td>
<td>761</td>
<td>94 57 100</td>
</tr>
<tr>
<td>Socioeconomic questionnaire</td>
<td>555</td>
<td>67 – 92</td>
</tr>
<tr>
<td>Controls (of 675)</td>
<td>613</td>
<td>100 67 100</td>
</tr>
<tr>
<td>Stools submitted</td>
<td>555</td>
<td>88 50 100</td>
</tr>
<tr>
<td>Enumeration component (of 4744)</td>
<td>2182</td>
<td>48 23 72</td>
</tr>
</tbody>
</table>

Baseline questionnaire return, follow up of population cohort, and nested case control component

Ninety-five per cent of the cohort component returned baseline questionnaires. Eighty-two per cent completed more than 23 weeks of follow up and 61% completed the full 26 weeks. Overall, stool and questionnaire compliance was higher in the nested case control component (table 3) than in the GP case control component. Low compliance with questionnaire return, stool submission, and case control matching was associated with 10 to 24 year age group, males, urban location, the South East, and practices with high Jarman scores.

Returning the socioeconomic questionnaire

Compliance was higher in the cohort component (median 67%) than the enumeration (48%) and GP case control (43%) components. Overall compliance was not high but 63% of those who returned risk factor questionnaires also returned a socioeconomic questionnaire. The age, sex, and social class distribution was similar to that in the case control and cohort studies, suggesting that the responses were representative.

Discussion

Representativeness

The study practice population of nearly half a million accounted for about 1% of the population of England. Stratified recruitment of the 70 practices achieved a study population representative of the national population by age, sex, geographical area, and urban/rural composition. The least deprived areas and smaller practices were slightly underrepresented, but this reflected the composition of the General Practice Research Framework rather than the sampling strategy. This is the first time IID has been studied in such a large and representative practice population sample, and the framework appeared to provide a suitable sampling frame. Previous population based studies in the United Kingdom (UK) were restricted.
to between one and seven practices, mainly in urban areas, and were unlikely therefore to represent the national population\textsuperscript{13-16}.

Despite a corrected enrolment rate of 40%, the characteristics of the cohort population were very similar to the population of England. Enrolment was slightly lower in people aged 15 to 24 years, males, and in manual classes than the structure of the national population would imply. A larger proportion were married, retired, and owned their own homes. Others have found lower response rates in subjects who are unmarried, unemployed, or poorly educated\textsuperscript{17}. A population based community cohort study in the Netherlands also found that people aged 19 to 35 years and males were less likely to take part than the reference population\textsuperscript{18}, but the Dutch sample also underrepresented people over 65 years, a problem not encountered in our study. Studies from the United States have used highly selected cohort populations\textsuperscript{19-21}. One studied a cohort of 61 families with young children, whose main wage earners had professional, managerial, or sales occupations\textsuperscript{19}. The results were therefore less generalisable to the whole population than ours. Previous larger scale community surveys in the UK have used retrospective recall alone to detect cases, which may be subject to bias\textsuperscript{16,22}.

**General practice list inflation**

In the GP practice components it was assumed that the total practice population would be followed for the year of participation, but a correction was made for practice list size inflation. This correction used evidence collected during recruitment of the cohort, which showed that a proportion of the names on practice lists had moved out of the area or died. The median list inflation was calculated to be 10%, within practice lists had moved out of the area or died. The evidence collected during recruitment of the cohort, which showed that a proportion of the names on practice lists had moved out of the area or died. The median list inflation was calculated to be 10%, within the year of participation, but a correction was made for practice list size inflation. This correction used the total practice population would be followed for the year of participation, but a correction was made for the total practice population would be followed for the year of participation, but a correction was made for practice list size inflation. This correction used evidence collected during recruitment of the cohort, which showed that a proportion of the names on practice lists had moved out of the area or died. The median list inflation was calculated to be 10%, within the practice lists, and therefore reflected variations between practices.

A second correction to the incidence in general practice was made for suspected underascertainment of cases – that is, failure to report a study case to the coordinating centre at the EMCU\textsuperscript{6}. This was estimated by using practices with computerised diagnosis to identify patients who should have been ascertained but were not. A median of 64% of cases were ascertained: the proportion varied according to practice characteristics: number of partners, urban or rural location, study component, and previous research experience\textsuperscript{6}. Estimates from the practices selected were extrapolated to the remaining practices, according to their individual characteristics. The representativeness of the population and the adjustments for list inflation and underascertainment should have increased the accuracy of the corrected IID presentation rate.

**Compliance and follow up**

Compliance in this study compared favourably with other cohort studies of IID\textsuperscript{18}. The corrected enrolment was 40%, a large proportion of baseline questionnaires were returned, and 82% completed more than 23 weeks. A Dutch study described an enrolment of 36%, 86% of whom completed more than 14 of the 17 weeks of total follow up\textsuperscript{18}. Our follow up data are likely to be highly accurate as they were derived using weekly report cards. This is important as the person years of follow up formed the denominator for calculating incidence. Two consecutive cohorts of six months each were used as opposed to one of 12 months, to increase the level of participation\textsuperscript{6}. Recruitment of controls and compliance in the nested case control component was very high, thus reducing potential bias.

In the GP case control component 70% of cases returned questionnaires and 75% submitted stool specimens. As presentation rather than completion of the questionnaire was the means by which cases were ascertained, no adjustment for presentation rates had to be made for non-responders. In another study 88% of questionnaires were completed and 67% of the cases provided stool specimens\textsuperscript{15}. Although we monitored the return of questionnaires as a regular quality assurance check, the compliance we achieved was not as high. Seventy-five per cent of controls were matched to cases. Compliance among controls was higher than among cases, probably because they had chosen to take part in the study. For cases and controls, compliance and case control matching were both lowest in the 15 to 24 year age group and in males. The practice characteristics associated with lower compliance identified by regression modelling were high Jarman score, urban location, and location in the South East. Although lower compliance in the 15 to 24 year age group could have caused bias, information collected showed that cases and controls were affected equally. Matched analysis of risk factors would also have helped to minimise confounding. Respondents to the socioeconomic questionnaire had a similar distribution of social class, age, and sex to that in the case control and cohort studies, suggesting that the sample was representative.

**Implications of the study**

The robustness of the overall design demonstrated by the representativeness of the total sample population and reasonable compliance with stools and risk factor questionnaires suggests that efforts made to avoid bias in this study were reasonably successful and that its results will be applicable to England as a whole\textsuperscript{28}. Comparison of this study with other studies of gastroenteritis in the population supports this assertion, as discussed above\textsuperscript{13-16,19-21}. The corrections made for practice list size and for underascertainment will make adjustments for presentation rates in general practice more accurate\textsuperscript{6}. Despite the strategy of matched analysis potential biases the results may not be entirely representative of urban practices with high Jarman scores, and this may need to be taken into account when interpreting the main results. Researchers conducting observational
studies need to be aware of these factors in order to develop strategies to improve and allow for low compliance.

Acknowledgements
We thank Professor TW Meade, the EMCU staff, and M Goldsborough, A Williams, L Hands, E Marshall, P Allen, F Symes, S Fox, and J Elwood for their invaluable help.


References
A study of infectious intestinal disease in England: microbiological findings in cases and controls

DS Tompkins, MJ Hudson, HR Smith, RP Eglin, JG Wheeler, MM Brett, RJ Owen, JS Brazier, P Cumberland, V King, PE Cook

Summary: A study was undertaken to identify the microorganisms and toxins in stool specimens associated with infectious intestinal disease (IID) among cases in the community and presenting to general practitioners (GPs) and in asymptomatic controls. Population based cohorts were recruited from practice lists in 70 practices and followed for 26 weeks (cohort component). Seven hundred and sixty-one cases of IID identified from the cohorts, 2893 cases who presented to GPs in 34 of the practices (GP component), and age/sex matched control subjects (555 and 2264, respectively) submitted stool specimens by post for comprehensive microbiological examination.

Campylobacter spp (12.2% of stools tested), rotavirus group A (7.7%), and small round structured virus (SRSV) (6.5%) were the organisms most commonly detected in the GP component. SRSV was identified in 7.0% of cases in the community cohort. No target microorganisms or toxins were identified in 45.1% and 63.1% of cases in the two components. Aeromonas spp, Yersinia spp, and some enterovirulent groups of Escherichia coli were detected as frequently in controls as in cases.

The higher frequency of detection of campylobacter, salmonella, and rotavirus among cases who presented to GPs than among those in the community suggests that those pathogens cause more severe illness. No enteropathogens were detected from a large proportion of cases although comprehensive standard methods were used to seek them.

Introduction

Current data on the significance of enteropathogenic microorganisms in England and Wales are derived largely from routine voluntary reporting of specified isolates by clinical microbiology laboratories to the PHLS Communicable Disease Surveillance Centre (CDSC). These reports are thought to represent a fraction of the true incidence of infections as only a proportion of cases seek medical attention, only a subset of these submit a stool specimen for analysis1, and there are also deficiencies and inaccuracies in reporting by laboratories2. Further limited information is derived from CDSC’s surveillance scheme for general outbreaks (those affecting members of more than one household) of IID2.

Previous studies undertaken to define microbiological causes of gastrointestinal disease have sought to identify only a limited number of the recognised or putative enteropathogens in faecal specimens submitted from general practice, but have not attempted to define the relative contributions of enteropathogenic bacteria, protozoa, and viruses in the community as a whole3-5.

Following a recommendation of the Richmond Committee6, a complex collaborative study was...
undertaken to determine the true incidence of infectious intestinal disease (IID) in England and to estimate the incidence of gastrointestinal disease in the community attributable to a microbiological cause. The feasibility of the methodology adopted was confirmed in a pilot study. This paper presents the main microbiological results of the study. Further details of the methods are included in an accompanying article, and other results including incidence, socioeconomic costs, and risk factors will be published in a detailed report.

Methods

Subjects
Cases were defined by symptoms of loose stools or vomiting (details and exclusions page 103). Single stool specimens were obtained from cases and controls in two study components – a population cohort component (with a nested case control element) and a GP case control component. Thirty-four out of 70 GP surgeries throughout England selected to be representative of the population as a whole took part in the GP case control component. As practice recruitment was staggered, specimens were received between August 1993 and January 1996.

Laboratory methods

Stool specimens were submitted by first class post to Leeds Public Health Laboratory (PHL). Investigations of small specimens were prioritised as about 9 grams of faeces were needed for completion of investigations for all target organisms and toxins (table 1). An aliquot of each specimen of sufficient volume was sent to the PHLS Laboratory of Enteric Pathogens (LEP) for examination for the pathogenic or potentially pathogenic (enterovirulent) groups of Escherichia coli using DNA probes. A second aliquot was frozen at -70°C in cryoprotective broth to be archived for further studies.

Conventional selective and enrichment culture techniques were used for the isolation of bacterial enteropathogens. Bacillus spp. spores and Staphylococcus aureus were counted, as counts of >10⁷/g and >10⁸/g of faeces, respectively, were considered significant. Isolates were sent to PHLS reference laboratories for confirmation of identity and typing.

Clostridium difficile cytotoxin (toxin B) was detected using Vero cells. C. perfringens enterotoxin was detected using an agglutination assay (PET-RPLA; Oxoid, Basingstoke, Hampshire) and positive results confirmed using an in-house enzyme immunoassay (EIA) at the Food Hygiene Laboratory of the PHLS Central Public Health Laboratory (CPHL). Conventional light microscopy of a wet film and formol-ether concentrate was used to detect ova, cysts, and parasites. An auramine-stained smear was examined using fluorescence microscopy for cysts of Cryptosporidium parvum and Cyclospora cayetanensis.

Transmission electron microscopy (EM) was used to detect viruses, with EIAs for rotavirus group A (Rotascreen; Microgen, Camberley, Surrey) and adenovirus types 40/41 (Adenoclone type 40/41; Cambridge Biotech, Worcester, MA, USA). All astrovirus identifications were confirmed by culture and specific fluorescent antibody testing (Oxford PHL). Rotaviruses detected by EM but not by EIA were subsequently confirmed as rotavirus group C by the Enteric Virus Unit of CPHL using polyacrylamide gel electrophoresis.

Written clinical reports were posted to participating surgeries and all clinically significant findings were reported by telephone. Data on microbiological findings were analysed at the London School of Hygiene and Tropical Medicine.

Statistical methods

The frequency with which organisms were identified in controls was standardised by age to reflect the age distribution of the general population rather than that of the cases. The direct standardisation method was applied, using the mid-1994 population age distribution of England.

Results

All 6743 stools received were cultured for all bacterial target organisms and 90% of them were of sufficient volume for investigation to stage 5 (table 1). Specimens were obtained between 0 and 73 days after the onset of symptoms (median 3 days, mode 1 day). Specimens were received by Leeds PHL between 0 and 45 days after being obtained (median 2 days, mode 1 day).

Target organisms or toxins were identified in 54.9% of cases in the GP component and 36.9% in the cohort component (tables 2 and 3). Small round structured virus (SRSV) was the most commonly identified target organism in cases of IID in the community cohort (7% of cases positive). Campylobacter spp. and rotavirus group A were more common than SRSV in cases in the GP component. The distribution of microorganisms

### TABLE 1 Target organisms sought and priority list for microbiological investigations

<table>
<thead>
<tr>
<th>Priority of testing</th>
<th>Procedure</th>
<th>Target organism sought</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stage 1</td>
<td>Bacteriological culture</td>
<td>Campylobacter sp</td>
</tr>
<tr>
<td>Stage 2</td>
<td>Bacteriological culture</td>
<td>Aeromonas sp, Bacillus sp, Clostridium difficile, Salmonella sp, Shigella sp, Staphylococcus aureus, Vibrio sp, Yersinia sp</td>
</tr>
<tr>
<td>Stage 3</td>
<td>Bacterial culture</td>
<td>Escherichia coli Q157, Giardia intestinalis</td>
</tr>
<tr>
<td>Stage 4</td>
<td>Direct microscopy</td>
<td>Entero-virulent E. coli, Cryptosporidium parvum, Cyclospora cayetanensis</td>
</tr>
<tr>
<td>Stage 5</td>
<td>Virology (electron microscopy and enzyme immunoassay)</td>
<td>Adenovirus, astrovirus, calicivirus, rotavirus, SRSV (Norwalk-like)</td>
</tr>
<tr>
<td>Stage 6</td>
<td>Toxic tests culture counts for vegetative and spores</td>
<td>C. difficile, C. perfringens, B. cereus, S. aureus</td>
</tr>
<tr>
<td>Stage 7</td>
<td>Concentration and for ova, cysts, and parasites</td>
<td>Protozoa and helminths</td>
</tr>
<tr>
<td>Stage 8</td>
<td>20% frozen suspension</td>
<td>Archiving at CAMR</td>
</tr>
</tbody>
</table>
in controls, when standardised by age, illustrates the carriage of enteropathogenic microorganisms by asymptomatic individuals in the general population of England (tables 2 and 3).

The relative proportions of frequencies of detection in cases and controls in the GP component (where the numbers were larger) are compared in table 4, highlighting the differences in proportions between different microorganisms. High case:control ratios were seen with recognised enteropathogens such as campylobacter, salmonella, *C. parvum*, and the enteric viruses. More than one target organism or toxin was identified in 11.3% of cases presenting in the GP component, 6.4% in the cohort component, and in less than 2% of all controls (table 5). Many combinations were observed but none predominated.

The percentages of isolates identified only after enrichment procedures in cases and controls, respectively, were 31.5% and 58.3% for salmonella, 4.5% and 0% for campylobacter, 78.5% and 84.7% for aeromonas, and 64.9% and 83.1% for yersinia. The effect of enrichment on isolation rates became more pronounced as time between the onset of symptoms and the receipt of the specimen in the laboratory increased for both salmonella and campylobacter (data not shown).

The speciation and typing of bacterial isolates were consistent with data from national surveillance, with *C. jejuni* and *C. coli* being the most frequently identified campylobacters (88% and 9% respectively) and *S. enteritidis* and *S. typhimurium* the commonest salmonellas (55% and 20%, respectively).

The percentages of faecal specimens positive for selected target organisms and toxins in different age groups (table 6) confirmed previous observations. Viral infections were commoner in infants and young children and salmonella and campylobacter infections commoner in older children and adults.

### Discussion

This is the first British investigation to use such a wide range of microbiological methods to study IID in a large and representative population. Similar physician-based and population cohort studies with comprehensive microbiological investigations are in progress in the Netherlands. Isolation rates for

---

**TABLE 2 Target organisms identified in the GP case control component**

<table>
<thead>
<tr>
<th>Organism or toxin not detected</th>
<th>1305</th>
<th>2893</th>
<th>45.1</th>
<th>1834</th>
<th>2264</th>
<th>81.0</th>
<th>85.2</th>
</tr>
</thead>
</table>
| *percentage standardised by the population age distribution (England, mid 1994)*
| *Organisms not detected in specimens in this study include; Arcobacter butzleri, A. skirrowii, Balantidium coli, Cyclospora cayetanensis, Enteroinvasive Escherichia coli, Entamoeba histolytica, Helicobacter cinaedi, H. fennelliae, nematodes, Vibrio parahemolyticus*
Campylobacter and salmonella were higher in this study than have been reported previously. No pathogens were detected in 45% of cases of IID in the GP component and 63% of cases in the community cohort component, however, and there are several possible reasons for this. Some microorganisms identified as causes of IID (for example, *Listeria monocytogenes*, *microsporidia*) were not sought in this study and other unrecognised pathogens may exist. Non-infective causes of intestinal disease can evoke symptoms similar to those of IID, although our case definition sought to exclude these causes. There were some technical reasons for potential under-recovery: a single specimen only was examined from each case, 30% of specimens were too small for complete analysis, and delays between onset of symptoms and receipt of specimens could have impaired viability and therefore detection by the methods used. EM detection of virus particles is relatively insensitive. Use of sensitive molecular methods such as polymerase chain reaction (PCR) amplification would probably have increased detection. The potential remains for further investigation of the archived material using such molecular techniques.

Viruses were confirmed as the commonest causes of IID in the community, particularly in children, with SRVS the most frequently identified target organism in the population cohort component. Evidence from outbreaks, studies of children admitted to hospital with diarrhoea, and failure to identify bacterial pathogens in the majority of cases all suggest that viruses are responsible for many cases of community-acquired IID. The contribution of SRVS and other viruses to widespread morbidity has now been confirmed. In contrast to the other viruses, SRVS were identified more frequently in cases in the population cohort than in the GP component. This is consistent with the finding that SRVS infection generally causes less severe illness than infection with other recognised pathogens such as rotavirus, salmonella, and campylobacter. Rotavirus group A was the target organism most frequently identified in children up to the age of 5 years in the GP component. It has been estimated that about 18000 children under 5 years of age are admitted to hospital in England and Wales each year for rotavirus infection. A rotavirus vaccine...
was licensed for use in the United States in August 1998 and approval for use in the United Kingdom is awaited.

C. jejuni was the most frequently identified target organism in the GP component, with group A rotavirus the runner-up, reflecting patterns seen in national surveillance data, although the two sources cannot be compared directly. In routine diagnostic practice many of the target organisms and toxins are sought only when travel history, food history, or clinical features indicate a need for specific tests. The relationship between the results of this study, including results of other aspects of this study, and national surveillance data are presented elsewhere.

Salmonellas were isolated more frequently from cases in the GP component (5%) than from cases in the community cohort component (1.1%), reflecting the severity of disease associated with this pathogen. VTEC O157 was isolated and also detected by DNA probes in three cases only in the GP component (0.1%), indicating that cases occur rarely, although affected individuals may have severe disease. Enteroaggregative E. coli, a heterogeneous group associated with diarrhoea in children and adults, was the most commonly detected enterovirulent E. coli in GP component cases.

The rarity of recognised enteropathogens in control subjects was expected, but the relatively high rates of detection of aeromonas, yersinia, diffusely adherent E. coli, and attaching and effacing E. coli in controls in both components of the study suggests that their detection in cases was not always related to IID. Typing of these organisms did not help to identify pathogenic subsets. Aeromonas and yersinia in most cases and controls were isolated only after enrichment, indicating that these microorganisms were present in small numbers in stool specimens. Further studies on archived isolates may help to identify differences in virulence characteristics and improve our understanding of the role of these microorganisms in IID.

**TABLE 5** Frequency of multiple organisms in faecal specimens (percentage)

<table>
<thead>
<tr>
<th>Organisms identified</th>
<th>GP component cases (%)</th>
<th>Community cohort cases (%)</th>
<th>Controls (%)</th>
<th>Total (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n=2893)</td>
<td>(n=761)</td>
<td>(n=2819)</td>
<td>(n=6473)</td>
</tr>
<tr>
<td>0</td>
<td>1305 (45.1)</td>
<td>480 (63.1)</td>
<td>2296 (81.4)</td>
<td>4081 (63.0)</td>
</tr>
<tr>
<td>1</td>
<td>1261 (43.6)</td>
<td>232 (30.5)</td>
<td>478 (17.0)</td>
<td>1971 (30.5)</td>
</tr>
<tr>
<td>2</td>
<td>276 (9.5)</td>
<td>48 (6.3)</td>
<td>41 (1.5)</td>
<td>53 (0.8)</td>
</tr>
<tr>
<td>3</td>
<td>48 (1.7)</td>
<td>1 (0.1)</td>
<td>4 (0.1)</td>
<td>53 (0.8)</td>
</tr>
<tr>
<td>4</td>
<td>3 (0.1)</td>
<td>–</td>
<td>–</td>
<td>3 (0.1)</td>
</tr>
</tbody>
</table>

**TABLE 6** Percentage of stools positive, by age group, for cases with selected target organisms in the GP case control (GPCC) and population cohort (PC) components

<table>
<thead>
<tr>
<th>Organisms identified</th>
<th>&lt; 1 year</th>
<th>1-4 years</th>
<th>5-14 years</th>
<th>15-74 years</th>
<th>&gt; 74 years</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>GPCC 302</td>
<td>PC 31</td>
<td>GPCC 606</td>
<td>PC 156</td>
<td>GPCC 221</td>
</tr>
<tr>
<td>Campylobacter spp</td>
<td>2.0</td>
<td>6.4</td>
<td>5.4</td>
<td>6.4</td>
<td>11.3</td>
</tr>
<tr>
<td>Closidium difficile cytotoxin</td>
<td>7.2*</td>
<td>28.4*</td>
<td>1.7</td>
<td>0.9</td>
<td>0.6</td>
</tr>
<tr>
<td>Closidium perfringens enterotoxin</td>
<td>4.0</td>
<td>0.0</td>
<td>5.6</td>
<td>1.9</td>
<td>4.5</td>
</tr>
<tr>
<td>Salmonella spp</td>
<td>2.3</td>
<td>0.0</td>
<td>2.5</td>
<td>1.9</td>
<td>5.9</td>
</tr>
<tr>
<td>Cypstosporidium parvum</td>
<td>0.7</td>
<td>0.0</td>
<td>2.8</td>
<td>1.3</td>
<td>5.0</td>
</tr>
<tr>
<td>Adenovirus types 40, 41</td>
<td>6.9</td>
<td>7.1</td>
<td>10.3</td>
<td>6.1</td>
<td>1.0</td>
</tr>
<tr>
<td>Astrovirus</td>
<td>1.9</td>
<td>7.1</td>
<td>6.7</td>
<td>4.0</td>
<td>3.0</td>
</tr>
<tr>
<td>Rotavirus group A</td>
<td>21.3</td>
<td>10.3</td>
<td>17.3</td>
<td>8.8</td>
<td>6.7</td>
</tr>
<tr>
<td>RSRS and calicivirus</td>
<td>14.7</td>
<td>17.8</td>
<td>14.8</td>
<td>13.5</td>
<td>6.0</td>
</tr>
</tbody>
</table>

* percentage in controls GPCC – 16.6, PC – 21.0
RSRS: Small round structured virus (Norwalk-like)
C. perfringens enterotoxin was confirmed as an important cause of IID in the community and was commoner in GP patients (4%) than in the population cohort (1.2%). C. difficile cytotoxin was most frequently identified in both cases and controls under 2 years of age, when the organism is a common feature of normal faecal flora, but was also found in 21 older cases who presented to GPs, associated with prior use of antibiotics. Current efforts to reduce unnecessary prescribing of antibiotics may reduce the frequency of associated episodes of diarrhoea.

This study has helped to improve understanding of the microbial causes of IID in England and highlighted the differences between the spectrum of disease in the community as a whole and among those who present to GPs. Results of this study will help medical microbiologists to optimise laboratory protocols for the investigation of stool specimens submitted by GPs. The relative numerical importance of different microorganisms has been clarified but this should be considered alongside the severity of illness that they cause. Work on archived stool specimens and bacterial isolates will help to provide further useful information.

Acknowledgements

We thank the staff of the laboratories that took part in the study, the many individuals who provided helpful comments and guidance, and the surgeries in the MRC’s General Practice Research Framework (listed on page 107).

References

10. Smith HR, Scotland SM. Recent developments in laboratory techniques for the detection of diarrhoeagenic Escherichia coli. PHLS Microbiology Digest 1994; 11: 7-12.