

The University of Salford

**Statistical Inference in Epidemiology:
Analysis of Gastro-intestinal Disease
in Day-nurseries**



Mariam Ahmad Al-Mannai

Centre of Operations Research and Applied Statistics
Department of Computer and Mathematical Sciences

Submitted in Partial Fulfilment of the Requirements of
the Degree of Doctor of Philosophy

March 98

Dedication

I would like to dedicate this work to my parents,
Ahmad Mohammed Al-Mannai and Fatima Salman Ajlan.

Contents

1	Introduction	3
2	Data Collection and Laboratory Methods	13
2.1	Data Collection Method	13
2.2	Laboratory Methods	16
3	Microbiological and Medical Background	20
3.1	Viruses	20
3.1.1	Adenovirus	20
3.1.2	Calicivirus (Norwalk agent) and other small rounded gastroenteritis viruses	24
3.1.3	Coronavirus	26
3.1.4	Enteroviruses	28
3.1.5	Reovirus	33
3.1.6	Rotavirus	34
3.2	Bacteria	36
3.2.1	Campylobacter	36
3.2.2	Escherichia Coli	39
3.2.3	Shigella	42
3.2.4	Yersinia	44
3.2.5	Spirochetes	46

3.3	Protozoa	46
3.3.1	Cryptosporidium	46
3.3.2	Giardia intestinalis	47
4	Statistical Overview	50
5	Elementary Statistical Analysis	83
5.1	Introduction	83
5.2	The Data	83
5.3	Elementary Statistics	84
5.4	Logistic Regression	91
5.4.1	Microbiological Factors	91
5.4.2	Social Factors	98
5.5	Proportions of asymptomatic carriage	101
6	Bias Correction	109
6.1	Introduction	109
6.2	Correction for cases missed due to low sensitivity	110
6.3	Correction factor ignoring edge effects	112
6.4	Correction factor including edge effects	116
6.5	The distribution of observed infection period	121
7	Survival Analysis and R_0 Calculation	123
8	Conclusions and Further Work	142
A	Glossary	144
B	Publication	165
C	Literature Review	191
C.1	Logistic Regression	191

C.2	The basic reproductive rate R_0	194
C.3	Survival analysis and Weibull distribution	201
D	References	205

List of Tables

4.1	Summary of the data	63
4.2	Distribution of symptoms between the two schools . . .	66
4.3	Occurrences of agents in the two schools	67
4.4	Frequencies of occurrences of symptoms and agents . . .	68
4.5	Distribution of symptom occurred simultaneously	69
4.6	Distribution of children by parents and parents employ- ment rates	69
4.7	Enrolment, attendance and specimen collection.	78
4.8	Incidence density of reported gastro-enteritis.	78
4.9	The study population. Age at entry to the study, years (%). Ages of 2 males and 1 female are missing.	80
4.10	The study population. Percentage of siblings.	81
4.11	Association between symptoms and isolation of specific organisms during the same fortnight.	81
5.1	One row of the Data as it appears in the SPSS data file.	85
5.2	Mean and standard deviation of the number of fort- nights of occurrence of the different agents of infection per child	87
5.3	Mean and standard deviation of the number of episodes of symptoms per child	87
5.4	Pearson correlation coefficients (1)	88

5.5	Pearson correlation coefficients (2)	89
5.6	Pearson correlation coefficients (3)	90
5.7	Pearson correlation coefficients (4)	91
5.8	Logistic regression Coefficients	96
5.9	Expected numbers of episodes of symptoms attributable to various agents	99
5.10	Distribution of age by sex	101
5.11	Distribution of agents by age for females	102
5.12	Distribution of agents by age for males	103
5.13	Distribution of agents by age for females when at least one symptom was present	104
5.14	Distribution of agents by age for males when at least one symptom was present	105
5.15	Proportion of asymptomatic carriage by age for females	107
5.16	Proportion of asymptomatic carriage by age for males	108
7.1	Coefficients estimates	138
7.2	χ^2 values	138
7.3	Table χ^2 and degrees of freedom values	139
7.4	R_0 estimates	139
B.1	The study population. Age at entry to the study, years (%). Records of ages of 2 males and 1 female are missing.	188
B.2	The study population. Percentage of siblings.	188
B.3	The study population. Single parent and employment status.	188
B.4	Incidence density of reported gastroenteritis.	188
B.5	Enrolment, attendance and specimen collection.	188

B.6	Association between symptoms and first isolation of specific organisms during the same fortnight. * denotes $p < 0.05$, Fisher's Exact Test. Note absence of Salmonella.	189
B.7	Proportion of first symptom episodes attributable to significantly associated organisms.	189
B.8	Infection rates as number per person-year, percentages symptomatic, and symptomatic infection rates as number per person year after a crude bias correction.	190
B.9	Agents tested for and their detection methods.	190
B.10	Significance tests for difference between age/sex group incidence density of Gastroenteritis no. Chi square tests for difference in proportions.	190

List of Figures

4.1	An example of a child's record	52
4.2	Abdominal pain by fortnight	53
4.3	Diarrhoea or Loose Stool by fortnight	53
4.4	Temperature by fortnight	54
4.5	Vomiting by fortnight	54
4.6	Adenovirus by fortnight	55
4.7	Astrovirus by fortnight	55
4.8	Calicivirus by fortnight	56
4.9	Campylobacter by fortnight	57
4.10	E. Coli by fortnight	57
4.11	Cororna virus by fortnight	58
4.12	Cryptosporidium by fortnight	58
4.13	Coxsackie A virus by fortnight	59
4.14	Coxsackie B virus by fortnight	59
4.15	Echovirus by fortnight	60
4.16	Enterovirus by fortnight	60
4.17	Giardia intestinalis by fortnight	61
4.18	Reovirus by fortnight	64
4.19	Rotavirus by fortnight	64
4.20	SRSV by fortnight	65
4.21	S. Sonnei by fortnight	65

4.22	Yersinia by fortnight	70
4.23	Adenovirus serotypes for school 1 by fortnight	71
4.24	Adenovirus serotypes for school 2 by fortnight	71
4.25	E. Coli serotypes for school 1 by fortnight	72
4.26	E. Coli serotypes for school 2 by fortnight	72
4.27	Coxsackie B virus serotypes for school 1 by fortnight . . .	75
4.28	Coxsackie B serotypes for school 2 by fortnight	76
4.29	Echovirus serotypes for school 1 by fortnight	76
4.30	Echovirus serotypes for school 2 by fortnight	77
4.31	Distribution of specimens collected per fortnight	79
4.32	Distribution of number fortnights enrolled	79
5.1	Logistic regression output.	93
7.1	Observed versus predicted number of infections of <i>Ad-</i> <i>enovirus</i> for both sexes by age band	131
7.2	Observed versus predicted number of infections of <i>Ad-</i> <i>enovirus</i> for females by age band	131
7.3	Observed versus predicted number of infections of <i>Ad-</i> <i>enovirus</i> for males by age band	132
7.4	Observed versus predicted number of infections of <i>E.</i> <i>coli</i> for both sexes by age band	132
7.5	Observed versus predicted number of infections of <i>E.</i> <i>coli</i> for females by age band	133
7.6	Observed versus predicted number of infections of <i>E.</i> <i>coli</i> for males by age band	133
7.7	Observed versus predicted number of infections of <i>Cox-</i> <i>sackie B</i> for both sexes by age band	134

7.8	Observed versus predicted number of infections of <i>Coxsackie B</i> for females by age band	134
7.9	Observed versus predicted number of infections of <i>Coxsackie B</i> for males by age band	135
7.10	Observed versus predicted number of infections of <i>Echovirus</i> for both sexes by age band	135
7.11	Observed versus predicted number of infections of <i>Echovirus</i> for females by age band	136
7.12	Observed versus predicted number of infections of <i>Echovirus</i> for males by age band	136

Acknowledgements

My greatest gratitude and sincerest appreciation is given to Dr. R. D. Baker for her supervision, useful guidance and her endless encouragement throughout the project.

I would like to thank Dr. R. H. Stevens, Dr. T. Riordan, Dr. A. Curry and Mrs. A. Walker for providing the data studied here.

I would also like to thank the University of Bahrain for their financial support during the period of my studies at the United Kingdom.

My last but certainly not least gratitude is to my parents and family for their continuous encouragement and support.

Summary

This study was aimed at giving a picture of gastro-enteritis among children in the community as opposed to in hospital, and also studying the main agents of infection and the amount of asymptomatic carriage of disease. A follow-up study of a one-year period was conducted in two different day nursery schools in a socially deprived area of Greater Manchester. A total of one hundred and sixty-seven children were enrolled in the study. Faecal specimens were collected fortnightly from the children and were tested for certain agents of infection. Three different types of microorganisms and twenty-one agents were isolated. Presence of any symptoms of illness was recorded. Demographic information on the children and their families was also acquired. A set of simple statistical analyses were conducted on the data such as calculation of means and standard deviations of the number of times each agent was isolated per child or an episode of a particular symptom was reported. Correlation coefficients between the number of times each agent was isolated per child and occurrence of symptoms were also calculated. Further comprehensive statistical analyses such as fitting logistic regression models and survival analysis were also conducted. This was aimed at estimating the attributable fraction of gastro-intestinal disease due to different agents, the force of infection and the basic reproductive rate R_0 , where feasible. This study was useful in two ways :

- in gaining an understanding of the causes of childhood diarrhoea in the community;
- and

- mathematically, some useful techniques for drawing inferences from "messy" epidemiological data have been developed.

Chapter 1

Introduction

In the contemporary era the study of statistics had been widely applied in many areas such as business, education, psychology, economics, medicine, health and epidemiology. Here, we present an example of applying statistics in the field of epidemiology and infectious diseases. The study involved two nurseries each of which was studied for a year with the intention of obtaining faecal specimens from each child every two weeks irrespective of symptoms. The study was conducted between 1985 and 1987 in the city of Salford.

The purpose of this study was to establish a relationship between social deprivation and frequent childhood diarrhoea. The study was originally conducted by two Public Health Officers from Manchester Public Health Laboratories, Dr. Terrance Riordan and Dr. Allan Curry. A nursery nurse, Mrs. Aprille Walker, was employed to assist in specimen collection and gathering information about the subjects of the study during the experimental period. The results of this part of the study were entered into a database on an Apricot computer and were taken to Exeter by Dr Riordan when he transferred to the Public Health Laboratories there. In 1995, because of the incompatibility of

the database and the Apricot floppy disks with modern systems of the data were printed out as hard copy. The author typed the data into a modern PC. It was then the author's responsibility to recode the data in another format as can be seen in chapter 4.

One hundred and sixty-seven children submitted one or more specimens with a total observation period of 94.7 children years. This overall figure compares with 131/147 children with gastro-intestinal illness where ill children were more likely to have specimens sent.

Day nurseries were chosen for this study because of the known high risk of enteric infections in such population and also because such a population is sufficiently stable to permit prospective studies to be carried out. The specimens were examined for a number of agents. This would provide much information of general relevance on the morbidity associated with infection by these organisms and their epidemiology. It would also give information on the importance of these organisms in day nurseries and in particular on how to prevent the spread of these organisms within a day nursery.

This chapter contains a description of the rest of the thesis and a list of the basic definitions that are needed throughout the course of research.

Chapter 2 discusses the data collection and laboratory methods as described by the experiment conductors. The author had no contribution to this since this was done long before her enrolment as a research student at the University of Salford. It is divided into two sections. The first section discusses the methods used in collecting the specimens from the study children. The second section discusses the laboratory methods used in detecting the agents found in the children's faecal specimens.

Chapter 3 contains a survey of some of the literature about the agents of infection isolated from the children during the data collection period. These organisms are divided into three types, viruses, bacteria and protozoa. This chapter is divided into three sections. Each section discusses one kind of these organisms.

Chapter 4 describes the data graphically and by a set of tables. It contains a table that gives the overall statistics of the study such as the number of children enrolled in each school and the number of children of each sex. Another table contains a list of the micro-organisms found during the laboratory analysis of the children's stool samples and a second table that lists the frequency of the number of symptoms children induced during the course of the experiment. It also contains graphical representations of the frequency distributions of the different symptoms induced and agents detected per child per fortnight. A table of the association between the different agents and symptoms is also included in this chapter. This chapter concludes with a frequency distribution of the number of fortnights enrolled per child and the number of specimens collected per fortnight.

Chapter 5 contains statistical tests and calculations performed on the data. It is divided into five sections. First, it discusses the data, then it covers the elementary statistics calculated from the data such as the means and standard deviations of the number of fortnights of the occurrences of the different agents of infection per child and the number of episodes of symptoms per child per fortnight. It also contains a table of the correlation coefficients of the occurrences of agents and symptoms and the different social factors such as number of parents the child lived with and the number of parents employed, the child's number of siblings and birth order. The fourth section contains the

results of fitting logistic regression models using the episodes of symptoms per child as the dependent variables and the presence of agents as the independent factors as one set of models. The second set of models used the same dependent variables and social factors as the independent variables. The final statistical analysis done in this chapter was calculating the proportions of asymptomatic carriage in the data.

Chapters 6 and 7 contain the results of fitting a Weibull survival model to the data and estimates of the basic reproductive rate for some of the agents of infection found during the specimen collection period.

Chapter 8 contains the conclusions reached during this study and suggestions for any further work.

Appendix A contains a glossary of medical terms used here, and appendix B contains an article that is to be published shortly that was done in collaboration with the original experiment conductors. Appendix C contains a review of the literature published about the methods of data analysis used in this study.

Definitions

The basic reproductive rate R_0

The basic reproductive rate, R_0 , is defined as the mean number of secondary cases of infection arising when one infected case is introduced to a totally susceptible population. R_0 could be directly measured by observing the number of secondary cases brought about by a single infected individual, but such direct measurement is almost never feasible and thus R_0 is normally estimated indirectly.

If $R_0 > 1$ the disease is capable of invading and establishing itself within the host population.

There are many methods for indirectly estimating R_0 . Anderson and May (1991) have developed several methods for estimating R_0 for both micro- and macro-parasites. A few estimates were given for different possible situations, for example there is a method of estimating R_0 when the population is assumed to mix homogeneously. Another estimate for R_0 is made under the assumption that an individual will gain life long immunity against the parasite. The values of R_0 for some of most common illnesses were estimated.

The main reason for estimating R_0 is to determine how large a fraction of the population must be immunised by vaccination against the parasite in order to eradicate the disease. To reduce R_0 to under 1, the population x vaccinated must satisfy

$$R_0(1 - x) < 1$$

or

$$x > 1 - \frac{1}{R_0}.$$

Survival analysis

Survival analysis is defined as a collection of statistical procedures for data analysis for which the outcome variable is ‘time until an event occurs’ Kleinbaum (1995). By time is meant years, months, weeks or days from the beginning of the follow-up of an individual until an event occurs, alternatively, time can refer to the age of an individual when an event occurs. An event is defined as death, disease incidence, relapse from remission, recovery or any designated experience of interest that may happen to an individual.

Survival analysis in the medical field deals with the following problems:

1. estimation of failure time distributions;
2. comparison of survival of different groups of subjects and estimation of treatment effects;
3. prognostic evaluation of different variables jointly or singularly considered, such as biochemical, histological and clinical characteristics.

Survival data arise when the aim is to study the time elapsed to some particular starting point to the occurrence of an event. The starting observation is usually a medical intervention such as first diagnosis of a given disease, a surgical intervention or the beginning of a treatment in a clinical study. In epidemiological studies, the starting point may be birth or the beginning of exposure to some risk factor. Survival analysis is useful whenever the researcher is interested not only in the frequency of occurrence of a certain type of event, but also in the time process underlying such occurrence.

Survival and Hazard functions

In summarising survival data, there are two functions of interest, the survival function and the hazard function.

The actual survival time of an individual, t , can be regarded as the value of a variable, T , which can take any non-negative value, where T is a random variable.

Suppose that T has a probability density with underlying probability density function $f(t)$. The distribution function of T is then given by

$$F(t) = P(T \leq t) = \int_0^t f(u)du.$$

The survival function $S(t)$ is defined to be the probability that an individual survives from the time origin to some time beyond t , so

$$S(t) = 1 - F(t).$$

The hazard function is the pdf that an individual dies at time t conditional on his/her survival to that time. This definition can be rephrased as follows: the hazard function is the pdf that the event of interest occurs at time t given that it did not occur prior to that time. To obtain a more formal definition of the hazard function, consider the probability that a random variable associated with an individual's survival time T , lies between t and $t + \delta t$ conditional on T being greater than or equal to t written

$$P(t \leq T \leq t + \delta t \mid T \geq t).$$

The hazard function $h(t)$ is defined as

$$h(t) = \lim_{\delta t \rightarrow 0} \frac{P(t \leq T \leq t + \delta t \mid T \geq t)}{\delta t}.$$

From this definition the following relationships between the hazard and survival functions are obtained:

$$P(t \leq T \leq t + \delta t \mid T \geq t) = \frac{F(t + \delta t) - F(t)}{S(t)}$$

$$h(t) = \lim_{\delta \rightarrow 0} \left(\frac{F(t + \delta t) - F(t)}{\delta t} \right) \frac{1}{S(t)}$$

$$h(t) = \frac{\frac{dF(t)}{dt}}{S(t)}$$

$$h(t) = \frac{f(t)}{S(t)}$$

$$h(t) = -\frac{d \log(S(t))}{dt}.$$

This implies that

$$S(t) = \exp\left(-\int_0^t h(u)du\right).$$

Let

$$H(t) = \int_0^t h(u)du$$

then

$$S(t) = \exp(-H(t))$$

$H(t)$ is called the integrated or cumulative hazard. It can be obtained by the formula

$$H(t) = -\log S(t)$$

In the analysis of survival data the survival and hazard functions are estimated from the observed survival times,

The Weibull distribution

The Weibull distribution is a probability distribution that plays a central role in the analysis of survival data Collette (1994). It was introduced by Weibull in 1951 in the context of industrial reliability testing. It is considered as central to the parametric analysis of survival data as the normal distribution is in linear modelling. The two parameter Weibull distribution has the hazard

$$h(t) = \alpha\beta(\alpha t)^{\beta-1}$$

where α and β are both greater than zero.

This model allows the hazard to depend on time through the exponent $\beta - 1$. It is monotone decreasing if $\beta < 1$, increasing if $\beta > 1$, and constant if $\beta = 1$.

In Type II censoring, observation ceases after predetermined number d of failures, so that c becomes a random variable. Type II censoring is a useful technique for economical use of effort in industrial life-testing. A crucial condition is that, conditionally on the values of any explanatory variables, the prognosis for any individual who is censored at c_i should not be affected if the individual is censored at c_i . That is, an individual who is censored at c should be representative of all those subjects with the same values of the explanatory variables who survive to c .

The simplest way to ensure this is to take the c_i to be, in principle, predetermined constants. Note, however, that often the c_i will not be known to the investigator in advance, and that the unrealised c_i corresponding to observed failure may never become known. The above condition is also satisfied if the potential censoring times are random variables c_i , which are independent of the T_i . Type II censoring is an example of a more general scheme in which, loosely speaking, censoring can depend on the past history, but not the future, of the whole process. We may call this evolutionary censoring.

Chapter 2

Data Collection and Laboratory Methods

2.1 Data Collection Method

The study was conducted to determine the association between social deprivation and frequent childhood diarrhoea. The study population was chosen from two Social Department day-care nursery schools which provide free of charge places for children who had been referred to the Social Services Department by a Health Visitor or a Social Worker as giving a cause of concern. Such causes of concern include: the child's name being on Child's Protection Register, parents having difficulties coping, for example single parent families, and children with disabilities. Those nurseries were located in a socially deprived area of Greater Manchester in order to maximise the rate of isolation of faecal pathogens.

A series of presentations was organised to seek the support of the nursery staff stating the aims and methods of the study. The parents of all the children who were currently enrolled were sent covering letters

and consent forms. Also the parents of the children who had attended after the study began were asked to consent. Nursery staff were asked to speak to the parents who did not return the consent forms when they brought their child to the nursery.

A research nurse was required for the following reasons: for the study to be successful it is essential that specimens and standardised clinical data are obtained during the study. Even with highly cooperative nursery staff, it would not be possible to obtain regular faecal samples from all the study children without outside assistance. Also, it was known from past experience that when children in day nurseries developed gastro-enteritis and were absent, the nursery staff would usually be given minimal details of illness, which would be quite unacceptable for a study of this type. It was therefore essential to have a person who was assigned to the study, who could coordinate specimen collection and visited the children with gastro-enteritis at home and obtained detailed and standardised information on the illness from the child's parents, and at the same time obtain stool specimens from the child and family if appropriate.

In summary, the research nurse would have a crucial central role in the study, providing the essential link between the nurseries and the laboratory without which the study would not be possible.

The nursery nurse, who was present at least two days a week, was responsible for collecting faecal specimens and collating microbiological and demographical data.

The aim of specimen collection was to obtain faecal specimens for each child who was enrolled in the nursery every two weeks for a one year follow up period or until the child leaves the nursery. Hence the data consist of a cohort of observations abridged into 26 fortnights.

If symptoms of gastro-enteritis were reported during a child's fortnight, this was recorded as a single episode. Multiple episodes within a child fortnight were not permitted. This may represent a slight under recording of episodes as previous work suggested that symptoms recurring after a four day interval can be regarded as a separate episode (Baqui *et al.* 1991).

The two week period for faecal specimen collection was chosen because it has been shown by population surveys to be the optimal trade-off between under-reporting due to memory failure associated with longer periods of recall and over-reporting, which has been attributed to the reporting of short duration symptoms episodes that turn out not to be judged as illness once a few days have elapsed.

If symptoms of gastro-enteritis occurred in a fortnight, every effort was made to collect one or more diarrhoeal specimen and all isolates resulting from these specimens were recorded. When the child were absent from the nursery because of illness, enquiries were made and Environmental Health Department was notified if gastro-enteritis was suspected.

Specimens were obtained from children in diapers and from preferably a potty or from the lavatory bowl in children who had gained bowel control. There was a separate lavatory bowl for each nursery class, and three nursery classes per school. Stool collection charts were placed in each lavatory to enable staff to participate in stool collection. This practice in both schools was to encourage all the children to visit the toilet together at set times of the day. Even though the study nursery nurse was present on at least two days per week, it was therefore necessary for some help to be given by staff to achieve complete specimen collection. Fresh specimens were taken personally by the

nursery nurse to the laboratory each day avoiding the need to enoculate the late specimens into transport medium. Cover was provided by the Environmental Health Officers for the nurse's vacation and sickness absences. The Environmental Health Officer would collect the specimens and transport them to the laboratory in the same way. If it was known that the child was absent from the nursery because of gastro-enteritis, the Environmental Health Officer would visit the household and attempt to obtain a diarrhoeal faecal specimen.

Data about the child's family were obtained from the school records. Histories of symptoms of gastro-enteritis were obtained from the report of parents and/or staff.

2.2 Laboratory Methods

Stools were routinely cultured for *Enteropathogenic Escherichia Coli*, *Salmonellas*, *Campylobacter*, *Shigellas* and *Yersinia*. They were also examined microscopically for ova, cysts and parasites.

Isolation of *Salmonellas* and *Shigellas*

Specimens were plated directly on Deoxycholate Citrate Agar (DCA) and also inoculated into a gelatine broth which was sub-cultured with DCA plate after 18 hours' incubation. Subspecies of *Salmonellas* and *shigellas* were examined using standard serological and biochemical methods.

Isolation of *Campylobacter*

Specimens were plated on a medium containing blood agar base, 5mg/L. Plates were incubated for 48 hours at 41°C under reduced oxygen tension. Colonies with suggestive morphology are examined by Gram stain and oxidase.

Verotoxin testing of *E. Coli* isolates

Specimens were plated onto MacConkey agar. After incubating, five lactose-fermenting colonies are picked to blood agar for testing for agglutination with standard enteropathogenic "O" antisera. Organisms showing agglutination to titre with specific antisera were confirmed to be *E. Coli* by standard biochemical tests.

Strains of enteropathogenic serotypes of *E. Coli* were tested for verotoxin production by the following method. The strain would be inoculated into broth and incubated overnight at 37°C. The broth would then be filtered to remove bacterial cells and the filtrate tested for cytotoxic activity in Vero cell line. Two dilutions of each filtrate would be added to Vero cell monolayers maintained in minimal essential medium plus foetal calf serum. The cells would be incubated at 37°C and examined daily for four days. Any culture showing a cytotoxic effect was referred to the Division of enteric pathogens at Colindale for confirmation. Although the method described is less sophisticated than that of O'Brien *et al.* (1982) it was sufficiently sensitive to be used for the proposed study.

Giardia intestinalis

Specimens were examined for the protozoal cysts by light microscopy. Examination involved both direct examination of saline suspensions, and examination following concentration using a formalin ether method of Allen and Ridley (1970).

Cryptosporidium

Smears of faeces were prepared and fixed then stained using a

phenolauramine method Nichols and Thom (1984). The smears were then examined for fluorescent oocysts using an ultraviolet microscope. Positive results were confirmed using a modified Ziehl Neelsen method.

Electron Microscopy for viruses

A 20% faecal suspension was prepared in phosphate buffered saline with added bicarbonate. The suspension was centrifuged for 20 minutes at 2000g. The supernatant was re-spun at 65000g for one hour, the supernatant was poured off and the pellet resuspended in the remaining fluid. A drop of suspension was placed on a clean waxed surface and a formvar coated electron microscope grid was floated on the drop for five minutes. The grid was then washed in a drop of distilled water and blotted dry on filter paper. It was then stained with a drop of 3% phosphotungstic acid (pH 6-6.5) for ten seconds. The grid was then blotted dry and examined in the electron microscope at 63000x magnification.

All morphologically different virus types, seen on scanning the grid, were photographed for subsequent reference.

Cell Culture for virus isolation

A pea-sized portion of faeces was transferred to a bijou bottle of phosphate buffered saline. The bijou was shaken for ten minutes to emulsify the sample and frozen overnight at -40°C . The following day, the sample was thawed at room temperature and then centrifuged at 1500g for 15 minutes at 8°C . A 0.1ml aliquot of supernatant was inoculated into BK and MRC cell lines. These were examined the following day for evidence of viral CPE and the medium changed. Thereafter, the cells were examined twice

a week for three weeks before being discarded. Cells showing a characteristic viral CPE were identified by appropriate neutralisation tests.

In the next chapter, we will look at some literature about the agents of infection found in the studies. In chapter 4 we will discuss the data in more details and in chapter 5 we will discuss some of the statistical tests conducted on the data.

Chapter 3

Microbiological and Medical Background

This chapter contains a summary of some of the literature written about the microorganisms found during the laboratory analysis of the stool specimens of the children.

3.1 Viruses

3.1.1 Adenovirus

Adenoviruses derived their name from the fact that they were first isolated from adenoid tissues and have a certain affinity for lymph glands, where they may remain latent for years. There are seven subgenera with 41 species. Symptoms of infection and epidemiological patterns vary between subgenera and even for different species. *Adenoviruses* invade primarily the respiratory and gastrointestinal tracts and the conjunctiva. They may cause a variety of clinical manifestations ranging from pharyngitis to bronchitis, croup, pneumonia and diarrhoea. Some

species invade the urinary tract or central nervous system or cause systematic infection. *Adenovirus* infections are widely distributed and common. Most infections occur in childhood.

A recognised syndrome seen especially among military recruits is febrile acute respiratory disease. Other disease syndromes caused by certain specific species of *Adenovirus* are pharyngoconjunctival fever and epidemic keratoconjunctivitis. Several other disease syndromes that usually occur in children (hemorrhagic cystitis, pertussislike disease, skin rashes, gastroenteritis and intussusception) have been associated with *Adenoviruses*.

The viruses are unusually resistant to inactivation and degradation, allowing purification and molecular investigation. As much detailed structure and biochemistry of *Adenovirus* are known as for any microorganism. Although many problems remain concerning production and use of *Adenovirus* vaccines, both inactivated and live vaccines have been developed and found effective. Subunit and genetically engineered vaccines are possible.

In this study seven different serotypes of *Adenovirus* were isolated in virology testing. They are serotypes 1, 2, 3, 5, 7, 10 and 31 and one unidentified serotype, referred to here as *Adenovirus* 99 was isolated by electron microscopy.

Pathogenesis and Immunity

Adenoviruses infect epithelial cells lining the respiratory and enteric organs. Following local replication, viremia may occur with spread to visceral organs. This dissemination is more likely to occur in immunocompromised patients. Differences in viral fibre proteins result in different target cell specificity among *Adenovirus* serotypes. The virus has a

propensity to become latent in lymphoid tissues, such as adenoids and tonsils or Peyer's patches and can be reactivated by immunosuppression or infection with other agents. Although certain *Adenoviruses* can transform and are oncogenic in rodent cells, *Adenovirus* transformation of human cells has not been observed.

Epidemiology

Adenovirons are stable to drying, detergents, the gastrointestinal tract and even mild chlorine treatment. This allows their spread by the faecal-oral route and by fingers and fomites including towel and medical instruments and in poorly chlorinated swimming pools. *Adenovirus* is spread exclusively in human-to-human transmission and no apparent animal reservoirs for the virus. *Adenoviruses* are mainly spread by respiratory or faecal-oral contact. Virus spread is promoted by close interaction as in classroom and military barracks. *Adenovirus* may be shed intermittently and over long periods from the pharynx and especially in faeces. Most infections are asymptomatic, which greatly facilitates their spread in the community. *Adenoviruses* 1 through 7 are the most prevalent serotypes. From 5% to 10% of paediatric respiratory diseases can be attributed to *Adenoviruses* types 1, 2 and 5. *Adenoviruses* serotypes 4 and 7 seem specially able to spread among military recruits, since antibody prevalence to these serotypes is low in young children.

Clinical Syndromes

Adenoviruses cause infection primarily in children and less commonly in adults. Infections by reactivated virus occurs in immunocompromised children and adults. Several distinct clinical syndromes are associated

with *Adenovirus* infection.

Acute Febrile Pharyngitis and Pharyngoconjunctival Fever

Adenovirus causes pharyngitis, which is often accompanied by conjunctivitis (pharyngoconjunctival fever). Pharyngitis alone occurs in young children, practically, those younger than three years of age, and may mimic streptococcal infection. Pharyngoconjunctival fever occurs more often in outbreaks involving older children.

Acute Respiratory Disease

Acute respiratory disease is a syndrome of fever, cough, pharyngitis and cervical adenitis seen primarily in outbreaks among recruits because of serotypes 4 and 7.

Other respiratory Diseases

Adenoviruses are definite but infrequent causes of true viral pneumonia in both children and adults. Laryngitis croup and bronchiolitis may also occur. Pertussis-like illness with a prolonged clinical course has been associated with *Adenoviruses*.

Conjunctivitis and Epidemic Keratoconjunctivitis

Adenoviruses cause a follicular conjunctivitis in which the mucosa of the palpebral conjunctiva becomes pebbled or nodular and both conjunctivae become inflamed. Conjunctivitis may occur sporadically or in outbreaks that can be traced to a common source. Epidemic Keratoconjunctivitis may be an occupational hazard for industrial workers. Irritation of the eye by a foreign body, dust, debris etc. is a risk factor for this presentation.

Gastroenteritis and Diarrhoea

Adenovirus is a major cause of acute viral gastroenteritis, accounting for 15% of hospitalised cases. *Adenovirus* serotypes 40, 41 and 42 have been grouped as enteric *Adenoviruses* and appear to be responsible for episodes of diarrhoea in infants. These enteric *Adenoviruses* will not replicate in the same tissue culture cells as other *Adenoviruses* and rarely cause fever or respiratory symptoms.

Other Presentations

Adenovirus has also been associated with a pertussis-like illness, intussusception in young children, acute hemorrhagic cystitis with dysuria and hematuria in young boys, musculoskeletal disorder and genital and skin infections.

Systematic Infection in Immunocompromised Patients

Immunocompromised patients are at risk of serious *Adenovirus* infections. Diseases include pneumonia and hepatitis. Infection can be exogenous or endogenous sources.

3.1.2 Calicivirus (Norwalk agent) and other small rounded gastroenteritis viruses

Norwalk agent was discovered by electron microscopy in stools from adults suffering from gastroenteritis in Norwalk, Ohio. Norwalk agent and other small rounded gastroenteritis viruses include the *Caliciviruses*, *Astroviruses* and unclassified viruses bearing the names of the geographical locations where they were identified.

Pathogenesis

Jejunal biopsy of human volunteers infected with *Caliciviruses* reveals blunting of villi, cytoplasmic vacuolation and infiltration with mononuclear cells, but virus particles are not detected by electron microscopy of epithelial cells. The virus compromises the function of the brush border which prevents proper absorption of water and nutrients. Although no histological changes occur in gastric mucosa, gastric emptying may be delayed.

Epidemiology

Norwalk and related viruses typically cause outbreaks of gastroenteritis such as those that occur on cruise ships or in communities as a result of common source contamination. These viruses are transmitted by faecal-oral and possibly airborne routes.

Exposure to Norwalk agents in developed countries occurs later in life than in underdeveloped countries, presumably as a result of good sanitation. In developed countries outbreaks may occur year-round and have been described in schools, resorts, hospitals, nursing homes, restaurants and cruise ships. A measure of the importance of these agents is that almost 10% of all gastroenteritis outbreaks and almost 60% of those that nonbacterial are attributed to Norwalk viruses. Immunity is generally short-lived at best and may not be protective.

Approximately 3% of gastroenteritis in children in day-care in the United States can be attributed to *Caliciviruses*. As many as 70% of children have antibodies to *Astroviruses* by the age of seven.

Clinical Syndromes

Norwalk and related viruses cause symptoms similar to *Rotaviruses*. Infection causes diarrhoea with nausea and vomiting, especially in children. Fever may be present in as many as one third of patients. The incubation period is 24 to 28 hours, and the illness resolves within 12 to 60 hours without problems. *Astroviruses* cause diarrhoea without vomiting and are less likely to cause disease in adults.

Treatment, Prevention and Control

No specific treatment is available. Bismuth subsalicylate may reduce gastroenteritis symptoms. Outbreaks may be minimised by handling food carefully and maintaining the purity of the water supply.

3.1.3 Coronavirus

Coronaviruses isolated from man can be divided into at least three antigenic types by neutralisation assays. The strict host-host range specificities suggest that it is probably not a biological accident that the viruses initially isolated in organ cultures are immunologically distinct from those first detected in human embryo kidney cell cultures. The initial isolates of each group serve as the prototype strains for two immunological types. A third group of six strains obtained in human embryo tracheal organ cultures have a weak cross-reaction with the tissue culture viruses, but they are antigenically unrelated to the other organ culture isolates. The viruses from man appear antigenically unrelated to *Coronavirus* from other animals in neutralisation tests, but complement-fixation assays indicate a partial immunological relatedness of human types with mouse hepatitis virus.

Pathogenesis

Coronaviruses constitute another group of viruses responsible for common colds and pharyngitis. Patients from whom the viruses were recovered and volunteers who were inoculated intranasally to establish the etiological relation of *Coronavirus* to disease, exhibited signs and symptoms of acute upper respiratory infections; coryza, nasal congestion, sneezing and sore throat were common; less frequent symptoms were headache, cough, muscular or general aches, chills and fever. Pulmonary involvement has not been noted. In volunteers inoculated intranasally with the virus the average incubation period was 6 to 7 day. Although *Coronaviruses* have many characteristics that differentiate them from other viruses, *Coronaviruses* disease can not be distinguished clinically from common colds produced by *Rhinovirus* or even from those occasionally initiated by influenza viruses.

Epidemiology

Acute upper respiratory infections generally have their greatest incidence in children, particularly those between 4 and 10 years of age. It is therefore especially striking that *Coronaviruses* are more frequently associated with common colds in adults. Viruses have been rarely isolated from children, and complement-fixing antibodies are detected in only about 2% of them. Epidemiological studies using neutralisation titration, however, detected antibodies in 20 to 25% of children 5 to 9 years old, but infection was more common in older children and adults. In further contrast to the usual upper respiratory infections, it is a member who is 15 years or older who most frequently introduces infection into a family.

Coronavirus infections occur predominantly during winter and early spring months, and during a particular respiratory disease season only a single immunological type appears to be the causative agent. It is noteworthy that during the periods of prevalent *Coronaviruses* infections, rhinovirus infections were uncommon, during epidemics all segments of the population are affected, but virus tends to spread preferentially within families. Unfortunately, because of the technical difficulties comparatively few data on epidemiological incidence and spread of *Coronaviruses* are available.

Treatment, Prevention and Control

While an effective vaccine can probably be developed, additional data on duration of immunity following natural infection, antigenic structure and immunogenic potential of the virions, and incidence of infections are essential before the needs and feasibility of a vaccine are established.

3.1.4 Enteroviruses

The *Coxsackie viruses* and *Echoviruses* are described as *Enteroviruses* because they are all found in the intestine and extracted in faeces. New *Enteroviruses* continue to be isolated, but they are no longer subdivided into the three named groups. Instead they are called enteroviruses and numbered. All these viruses share the following features:

They all replicate in the intestine.

They commonly cause asymptomatic immunising infections which protect against future infections with the same virus.

They can give rise to viraemia.

They occasionally cause infection in the central nervous system.

They are commoner in children than in adults.

In temperate climates they cause infections in the summer and autumn.

Coxsackie Virus

This group are named after the place in New York state where the first member was isolated. In all there are 30 serotypes. Two broad groups of the viruses have been defined according to the histological nature and the sites of the lesions they produce in mice. *Coxsackie A virus* of which there are 24 serotypes and of *Coxsackie B virus* there are 6 serotypes. In our study four different serotypes of *Coxsackie B virus* namely 2, 3, 4 and 5 and only one serotype of *Coxsackie A virus* which is 9.

Clinical Syndromes

Coxsackie A viruses give rise to a number of different illnesses.

Aseptic meningitis, indistinguishable clinically from that caused by other enteroviruses, is caused by a number of types, e.g. 2, 4, 7, 9 or 23. Type A7 has given rise to epidemics of paralysis disease in Scotland and Russia. The paralysis is not as severe in its after-effects as that due to *Poliovirus*.

Herpangina is an acute feverish disease, usually in young children, characterised by lesions in the mouth consisting of papules on the anterior pillars of the fauces; these papules become vesicles and finally shallow ulcers with a grayish base and punched-out edge. There are usually 5 to 10 lesions and

the illness lasts about 4 days. Types 2, 4, 5, 6, 8, 10 and 23 have been isolated at various times from these infections. Outbreaks may be seen in nurseries and schools.

Hand, foot and mouth disease presents as a painful stomatitis with a vesicular rash on the hands and feet. Typically, it lasts about a week; most of the cases are seen in the summer in children aged 1-10 years and cases can occur in clusters and in families. The viruses usually implicated are types 5 and 16.

Coxsackievirus A21 has caused epidemics of colds in camps of military recruits.

Coxsackie B viruses give rise to the following illnesses:

Epidemic myalgia or Bornholm disease which is characterised by fever and the sudden onset of agonising stitch-like pains in the muscles of the chest, epigastrium or hypochondrium. Although the disease is most frequently recognised in its epidemic form, many sporadic cases occur. Epidemic myalgia may be complicated by pleurisy and pericarditis.

Myocarditis of which severe and fatal cases have been reported in newborn infants. The virus can be found in high concentrations in the myocardium at autopsy. Epidemics have occurred in nurseries when there is evidence of group B virus activity in a community. The baby is presumed to be infected by its mother and is highly susceptible to the virus. Myocarditis are pericarditis that can occur in children and adults, and the virus has been isolated from pericardial fluid.

Coxsackie B virus are a major cause of human myopericarditis.

Although severe in the neonate, it tends to follow a more benign course in adults. The initial symptoms are often of an upper respiratory or influenza-like illness followed 7-10 days later by clinical heart disease. Chest pain is a feature and electrocardiogram abnormalities have been found.

Coxsackie B viruses have been implicated in the condition known as the post-viral fatigue syndrome. This is a post-infectious complaint in which fatigue, profound muscle weakness and lack of concentration are cardinal features.

Coxsackie B strains have also been associated with influenza like illnesses, upper respiratory infections, bronchitis and pneumonia in young children.

Echovirus

Enteric Cytopathic Human Orphan viruses (*Echoviruses*) were first identified when cell culture techniques were applied to the isolation of *Polioviruses* from the faeces of patients suffering from paralytic and non-paralytic illness. There are currently over 30 serotypes of *Echoviruses* and there is still no clear association of all types with specific disease. The *Echovirus* serotypes isolated during this study were 5, 9, 11, 14, 18, 22, 30 and 31.

Clinical Syndrome

Most *Echovirus* infections cause few or no clinical symptoms, but some have been associated with clinical syndromes.

Infections can be widely spread in a community, although only a few suffer from clinical illness. Infections generally follow a short

incubation period of 3-5 days. Non-specific rashes of fleeting duration have been reported. The main syndromes associated with *Echovirus* are aseptic meningitis, paralysis, rash and respiratory diseases.

The onset of meningitis is abrupt, with severe headache and vomiting. Symptoms are usually self-limiting and after a variable convalescent period, a full recovery is made, although rare cases of paralysis have been reported. Serious infection can occur in neonates and fatal cases have occurred during outbreaks in special care nurseries. The more likely course is mild infection with simple fever, upper respiratory symptoms or diarrhoea.

Most of the *Echovirus* types have been associated with sporadic cases of aseptic meningitis or one of the other diseases mentioned earlier. In the eruptive fevers caused by types 9 and 16 the aetiological relationship between the rash and the infective agent is well established. A number of types, notably 4, 6, 9, 16, 20, 28 and 30, have considerable epidemic potential and the clinical features are very varied.

Treatment, Prevention and Control

No specific antiviral therapy is available for *Enterovirus* infections. Supportive therapy is extremely important for patients with paralytic disease to assist in their potential recovery.

No vaccines exist for *Coxsackie viruses* or *Echoviruses*. Transmission can presumably be reduced by improvements of hygiene and living conditions.

3.1.5 Reovirus

The term REOVIRUS (Respiratory Enteric Orphan VIRUS) infect both the respiratory and the intestinal tracts, usually without producing a disease. Though originally considered members of the *Echovirus* group (and classified as *Echovirus* 10) *Reoviruses* are larger and differ in producing characteristic cytoplasmic inclusion bodies.

Pathogenesis

Reoviruses have been isolated from faeces and respiratory secretions of healthy persons, as well as from patients with a variety of clinical illnesses, particularly minor upper respiratory and gastrointestinal diseases. Despite the frequent isolation of these viruses their relation to disease is not yet clear. Human transmission experiments have also failed to clarify the pathogenetic role of *Reoviruses*.

Epidemiology and Control

Although *Reovirus* infections do not seem to be of great importance, further studies are needed to define their pathogenetic potential. Both the respiratory and the gastrointestinal tracts may well be a source of their spread. Unrecognised *Reovirus* infections are common, for approximately 10% of children by 5 years of age and 65% of young adults in the United States have *Reovirus* in their sera. Antibodies are also available in various wild and domestic animals, but it is not known whether the animals serve as a reservoir for human infections.

3.1.6 Rotavirus

The *Rotaviruses* are a large group of gastroenteritis-causing viruses found in many different mammals and birds. Although viruses have long been considered as common cause of childhood diarrhoea, it was not until 1973 that a *Rotavirus* was first detected in human. Since then *Rotaviruses* have been implicated worldwide as common agent of infantile diarrhoea.

Human and animal *Rotaviruses* are divided into serotypes, groups and subgroups.

Epidemiology

Rotaviruses are found worldwide. Children are infected at an early age, as determined by the early acquisition of antibody. *Rotaviruses* are assumed to be passed from person to person by the faecal-oral route. Maximum shedding of virus occurs 2 to 5 days after initiation of diarrhoea but can occur without symptoms. Virus survives well on inanimate objects, such hands, furniture and toys because it can withstand drying. Although domestic animals are known to harbour serologically related *Rotaviruses*, they are not believed to be a common source of human infection. Outbreaks occur in preschool and day-care centres and among hospital infants.

Rotaviruses are one of the most common causes of serious diarrhoea in young children worldwide. Most children have been infected by *Rotavirus* by 4 years of age.

Clinical Syndromes

The incubation period for *Rotavirus* diarrhoeal illness is estimated to be approximately 48 hours. The major clinical findings in hospitalised pa-

tients include diarrhoea, vomiting, fever and dehydration. faecal leukocytes and blood in stool are not associated with *Rotavirus* diarrhoea. *Rotavirus* gastroenteritis is a self-limited disease. Recovery is generally complete and without sequel. In rare instances severe untreated dehydration can lead to death.

The role of *Rotaviruses* in respiratory infections is controversial. Human studies have detected virus secretions of children with respiratory symptoms, and animal studies have documented that aerosol transmission can occur.

Treatment, Prevention and Control

No specific antiviral therapy is available. Morbidity and mortality in patients with *Rotavirus* diarrhoea result from dehydration and electrolyte imbalance. The purpose of supportive therapy is to replace fluids so that blood volume electrolyte and acid-base imbalances are corrected.

Rotavirus vaccines are being developed to protect children, especially those in underdeveloped countries, from potentially fatal disease. Vaccines have been prepared from animal *Rotaviruses*. These vaccines share antigenic determinants with human *Rotaviruses*, do not cause infection in humans, and offer some protection against infection. Unfortunately they may not protect against all serotypes of *Rotavirus*.

Acquisition of *Rotaviruses* occur very early in life. The ubiquitous nature of these viruses make prevention of spread and infection unlikely. Once hospitalised, however, diseased patients must be identified and isolated to limit spread of infection to other susceptible patients.

3.2 Bacteria

The family Enterobacteriaceae is the largest, most heterogeneous collection of medically important gram-negative bacilli. At present, at least 27 genera and 102 species as well as eight enteric groups have been described.

Enterobacteriaceae are ubiquitous organisms that are found worldwide in soil, water, vegetation and are part of the normal intestinal flora of most animals and humans. Some members of the family (e.g. *Shigella*, *Salmonella* and *Yersinia pestis*) are always associated with disease when isolated from human, whereas others (e.g. *Escherichia coli*, *Klebsiella pneumonia* and *Proteus mirabilis*) are members of the normal commensal flora that can cause opportunistic infections. Infections can originate from an animal reservoir, from a human carrier or by endogenous spread of organisms in a susceptible patient involving virtually all body sites. The Enterobacteriaceae are responsible for 30% to 35% of all septicaemias, more than 70% of urinary tract infections and many intestinal infections.

Here we will discuss the members of the enterobacteriaceae related to our work.

3.2.1 Campylobacter

There are eleven species and seven subspecies of *Campylobacter* known now. *Campylobacter jejuni* is the most common cause of bacterial gastroenteritis in the United States. *Campylobacter coli* is responsible for 2% to 5% of *Campylobacter* gastroenteritis although it is reported to be more in under developed countries. *Campylobacter lari* and *Campylobacter upsaliensis* has also been associated with diarrhoeal diseases

in humans. These species are primarily restricted to the gastrointestinal tract with bacteremia observed in fewer than 1% of the infections. In contrast with other species, *Campylobacter fetus* subspecies *fetus* is most commonly responsible for systematic infections such as bacteremia, septic thrombophlebitis, arthritis, septic abortion and meningitis.

Pathogenesis and Immunity

The major factors with the development of disease are infectious dose of organisms and the level of specific immunity. Patients exposed to a large number of organisms or who lack gastric acids are more likely to develop disease. Patients with hypogammaglobulinemia have prolonged severe disease with *Campylobacter jejuni*.

The pathogenesis of *Campylobacter jejuni* gastrointestinal disease is not completely understood. Disease is characterised by destruction of the mucosal surfaces of the jejunum, ileum and colon. On gross examination, the mucosal surface appears edematous and bloody. Histologic examinations reveals ulceration of the mucosal surface, crypt abscesses in the epithelial glands, an infiltration into the lamina propria with neutrophils mononuclear cells and eosinophils. This inflammatory process is consistent with invasion of the organisms into the intestinal tissues. Enterotoxins, cytopathic toxins and endotoxin activity have been detected in *Campylobacter jejuni* isolates. However, the precise roles of these factors in the disease have not been defined.

Campylobacter fetus has a propensity to spread from the gastrointestinal tract to the bloodstream and distal foci. This is particularly common in debilitated and immunocompromised patients such as those with liver disease, diabetes mellitus, chronic alcoholism or malignancy. In vitro studies have demonstrated that *Campylobacter fetus* is res-

istant to complement-mediated and antibody mediated serum killing, where as *Campylobacter jejuni* is rapidly killed. *Campylobacter fetus* is covered with a protein that prevents complement-mediated killing in serum.

Epidemiology

Campylobacter are commensals of cattle, sheep, dogs, cats, rodents and fowl. Lifelong asymptomatic carriage in animal following an initial symptomatic phase, represents an important reservoir for human disease. Human infections result from consumption of contaminated food, milk or water. Contaminated poultry is responsible for more than half of the *Campylobacter* infections in developed countries. Food products that neutralise gastric acids effectively reduce the infectious dose. Faecal-oral transmission from person to person may also occur but transmission from food handlers is uncommon.

The actual incidence of *Campylobacter* is unknown because the disease is not systematically reported to public health officials. However, it has been estimated that more than 2 million *Campylobacter jejuni* infections occur annually and such infections are more common than *Salmonella* and *Shigella* infections combined. Disease is more common in the warm months of the year, but dose occur throughout the year. The peak incidence of disease is in young adults. *Campylobacter fetus* infections are relatively less common, with fewer than 250 cases reported. In contrast with *Campylobacter jejuni*, *Campylobacter fetus* infects immunocompromised elderly individuals.

Clinical Syndromes

Campylobacter jejuni infections are seen most commonly as acute en-

teritis with diarrhoea, malaise, fever and abdominal pain. Ten or more bowel movements per day can occur during the peak of disease, and grossly bloody stools may be present. The disease is generally self-limiting although symptoms may last for one week or longer. The range of clinical manifestations can include colitis, acute abdominal pain and bacteraemia. *Campylobacter fetus* infections are most commonly seen as septicemia with dissemination to multiple organs. Although the initial presentation may be referable to the gastrointestinal tract or abdomen. Endovascular localisation is reported.

Treatment, Prevention and Control

Campylobacter are susceptible to a wide variety of antibiotics. *Campylobacter* gastroenteritis is prevented by the proper preparation of food particularly poultry, consumption of pasteurised milk, and safeguards to prevent contamination of water supplies. Elimination of *Campylobacter* carriage in animal is unlikely.

3.2.2 Escherichia Coli

The genus *Escherichia* consists of five species, with *Escherichia coli* the most frequently isolated. Large numbers of *E. coli* are present in the gastrointestinal tract and are the Enterobacteriaceae most frequent associated with bacterial sepsis, neonatal meningitis, infections of the urinary tract and gastroenteritis in travellers to countries with poor hygiene. Most infections are endogenous in the sense that an individual's normal microbial flora is able to establish infection under conditions in which the host defences are compromised. Fourteen different serotypes of *E. coli* were isolated during the laboratory analysis of the stool samples of the children. They were O1, O18, O26, O44, O55, O86,

O111, O114, O119, O124, O125, O126, O127 and O128.

The antigenic composition of *E. coli* is complex, with a very large number of O, H, and K antigens described. The serologic classification of *E. coli* isolates is useful for epidemiologic purposes, and specific serotypes are associated with increased virulence.

Clinical Syndromes

septicaemia

E. coli is the most common gram-negative bacillus isolated from septic patients. The focus of infection from which the organisms spread into the bloodstream is commonly either the urinary tract or the gastrointestinal tract. The mortality associated with *E. coli* septicaemia is influenced by the source of infection and the underlying disease of the patient, with a significantly higher incidence of death in immunocompromised patients or with infections originating from intestinal perforation.

Urinary Tract Infections

E. coli is responsible for more than 80% of all community acquired urinary tract infections and the majority of hospital-acquired infections. Infecting strains originate from the gastrointestinal tract, with disease associated with specific serotypes, primarily O4, O6 and O75. The ability of these bacteria to resist killing in serum, produce hemolysins and bind to uroepithelial cells is associated with increased virulence.

Neonatal Meningitis

E. coli is one of the most common cause of neonatal meningitis. Although colonisation of infants with *E. coli* at the time of deliv-

ery is common, disease is relatively infrequent.

Gastroenteritis

Strains of *E. coli* that cause gastroenteritis are subdivided into five groups: enterotoxigenic (ETEC), enteroinvasive (EIEC), enteropathogenic (EPEC), enterohemorrhagic (EHEC) and enteroaggregative (EAaggEC).

ETEC are found in the small intestine. They are mediated by two types of enterotoxins, heat-labile and heat-stable enterotoxins. Secretory diarrhoea caused by ETEC follows a 1- to 2-day incubation period and persists for an average of 3 to 4 days. Symptoms are characteristically mild with cramps, nausea, vomiting and watery diarrhoea. Disease mediated by either toxin is indistinguishable.

EIEC are normally found in the large intestine. They are able to invade and destroy the colonic epithelium, producing a disease characterised by fever and cramps with blood and leukocytes in stool specimens. Disease has been specified with O serotypes of *E. coli*, however, serologic classification of isolates cannot reliably identify invasive strains.

EPEC, which live in the small intestine, are historically important agents of childhood diarrhoea particularly in impoverished countries. Although specific O serotypes have been associated with nursery outbreaks of EPEC diarrhoea, serotyping *E. coli* isolated in random or endemic disease is discouraged except in epidemiologic investigations. Disease is caused by the ability of the organism to adhere to the enterocyte plasma membrane and cause destruction of the adjacent

microvilli.

EHEC lives in the large intestine. It produces a toxin called verotoxin. This toxin has a cytopathic effect in Vero cell line of tissue culture cells. Two verotoxins have been described in EHEC: one homologous to *Shigella dysenteriae* toxin, the other toxin with a 60% homology to the *Shigella* toxin. The range of disease caused by EHEC varies from mild uncomplicated diarrhoea to hemorrhagic colitis with severe abdominal pain, bloody diarrhoea and little or no fever. Hemolytic Uremic Syndrome which is acute renal failure, thrombocytopenia and microangiopathic hemolytic anaemia is also associated with this organism. Serologic classification of isolates has limited usefulness; however approximately half of EHEC strains are serotype O157:H7. Disease is most prevalent in the warm months of the year, with the greatest incidence in children younger than 5 years of age. Most cases of epidemic and endemic disease have been attributed to consumption of undercooked ground beef and other beef products, as well as unpasteurised milk.

EAggEC live in the small intestine. They were originally called enteroadherent *E. coli*, have been implicated as a cause of persistent diarrhoea in infants in developing countries.

3.2.3 *Shigella*

Shigella is a member of the enterobacteriaceae. There are four species of *Shigella* consisting of approximately 38 O-antigen-based serotypes, *Shigella dysenteriae*, *Shigella flexneri*, *Shigella sonnei* and *Shigella boydii*.

Epidemiology

Shigella sonnei is the most common cause of shigellosis in the industrial world, and *Shigella flexneri* is the most common cause in under-developed countries. Shigellosis is primarily a paediatric disease, with most infections in children from 6 months to ten years of age. Endemic disease in adults is frequently due to contact with infected children. Infections in male homosexuals are also observed. Epidemic outbreaks of disease are associated with day-care centres, nurseries, and custodial institutions. Shigellosis is transmitted by the faecal-oral route, primarily by contaminated hands and less commonly in water or food. Bacilli can remain viable in water for as long as 6 months. Because as few as 200 bacilli can establish disease, shigellosis spreads rapidly in communities where sanitary standards and the level of personal hygiene are low.

Clinical Syndrome

Shigellosis is characterised by abdominal cramps, diarrhoea, fever and bloody stools. Clinical signs and symptoms appear 1 to 3 days after bacilli are ingested. The bacilli colonise the small intestine and begin to multiply within 12 hours. The initial sign of infection, profuse watery diarrhoea without histologic evidence of mucosal invasion, is mediated by enterotoxin. However, the cardinal feature of shigellosis is lower abdominal cramps and tenesmus, with abundant pus and blood in the stool. This results from invasion of the colonic mucosa by the bacilli, destruction of the superficial mucosa and production of mucosal ulceration. Abundant neutrophils, erythrocyte and mucus are found in the stool. The bacilli rarely penetrate beyond the mucosal layer and bacteraemia is uncommon. Infection is generally self-limited, although

antibiotic treatment is recommended to reduce the risk of secondary spread to family members and other contacts. Asymptomatic carriage of the organism in the colon develops in a small number of patients and can represent a persistent reservoir for infection in a community.

3.2.4 Yersinia

This organism is an animal parasite that, under certain conditions, is transmissible to man, either directly or through insect vectors. There are seven species of *Yersinia* and the most commonly known are *pestis*; *pseudotuberculosis* and *Enterocolitica*.

Pathogenesis

The transfer of *Yersinia pestis* from rats to man through the bites of infected fleas may occasionally result in a localised infection, known as pestis minor, with mild constitutional symptoms. More often the lymph nodes draining the area of the flea bite become affected and the adenitis caused results in painful swellings or buboes in the inguinal, axillary or cervical regions depending of the position of the flea bite. From these primary buboes the plague bacilli may spread to all parts of the body. In the absence of adequate antibiotic therapy administered early in the course of the disease, the case of fatality rate may be as high as 50%.

Pneumonic plague which develops as a result of droplets of animals infected with *Yersinia pestis* from person to person. A severe bronchopneumonia develops. The sputum becomes a thin and blood-stained. It contains numerous plague bacilli that are demonstrable by stained films or culture of the sputum. This type

of plague is highly contagious and is usually fatal, unless treated early.

Urbanic plague may occur as a primary infection or as a complication of bubonic plague. The plague bacilli spread throughout the body and the outcome is invariably fatal. *Yersinia pestis* is also the cause of this kind of plague.

Yersinia pseudotuberculosis may give rise to a subclinical infection or it may cause fatal septicaemia. Human infections occasionally result in a severe typhoid-like illness with fever, purpura and enlargement of liver and spleen which is usually fatal. More frequently, it causes mesenteric lymphadenitis, simulating acute or subacute appendicitis, with or without erythema nodosum. This mainly affects young males aged 5-15 years. They usually make an uneventful recovery.

Yersinia enterocolitica causes mild and severe gastroenteritis, mesenteric lymphadenitis, and septicaemia especially in the elderly. Secondary complications include erythema nodosum, polyarthrititis, Reiter's syndrome and meningitis. It differs from *Yersinia pseudotuberculosis* in that it affects children and adults of both sexes. In young children the infection may produce fever, diarrhoea, abdominal pain and vomiting. The symptoms may last for several weeks.

In animals, *Yersinia enterocolitica* has been isolated from pseudotuberculosis-like lesions as well as from apparently healthy animals.

3.2.5 Spirochetes

Spirochetes belong to the order *Spirochaetales* of bacteria. This order is subdivided into two families and seven genera, three of which are responsible of human diseases. The genus we have here, *Spirochetes* is not responsible for any human diseases.

3.3 Protozoa

3.3.1 Cryptosporidium

Cryptosporidium is a protozoa which is found within the brush border of the intestinal epithelium. It can be on either of two forms, a cyst or a trophozoite.

Epidemiology

Cryptosporidium species are distributed worldwide. Infection is reported in a wide variety of animals including mammals, reptiles and fish. Zoonotic spread from animal reservoirs to human, as well as person-to-person spread by faecal-oral and oral-anal routes are the most common means of infection. Veterinary personnel, animal handlers and homosexuals are at particularly high risk for infection. Many outbreaks have now been described in day-care centres where faecal-oral transmission is common.

Clinical Syndromes

As with other protozoan infections, exposure to *Cryptosporidium* may result in asymptomatic carriage. Disease in previously healthy individuals is usually a mild self-limiting enterocolitis that is characterised

by watery diarrhoea without blood. Spontaneous remission after an average of ten days is characteristic. In contrast, disease in immunocompromised patients characterised by 50 or more stools per day and a tremendous fluid loss, can be severe and last for months to years. In some AIDS patients disseminated *Cryptosporidium* infections have been reported.

Treatment, Prevention and Control

Unfortunately, no broadly effective therapy has been developed for managing *Cryptosporidium* infections in immunocompromised patients. No controlled studies have been published and all therapeutic information is based on isolated reports and inoculated information.

Because of the widespread distribution of this organism in humans and other animals, preventing infection is difficult. The same methods of improved personal hygiene and sanitation used for other intestinal protozoa should be maintained for this disease. Contaminated water supplies should be treated with both chlorination and filtration. In addition, avoidance of high risk sexual activities is critical.

3.3.2 *Giardia intestinalis*

Giardia intestinalis is a protozoa which colonise and infect the duodenum, small bowel and colon of human. It belongs to the flagellates group of protozoa which moves by the lashing of flagella that pull the organisms through fluid environments. Diseases produced by such organisms are primarily the result of mechanical irritation and inflammation. Both cysts and trophozoite forms of *Giardia intestinalis* are detected in faecal specimens from infected patients.

Epidemiology

Giardia has a worldwide distribution, and this flagellate has a sylvatic or "wilderness" distribution in streams, lakes and mountain resorts. Giardiasis is acquired by drinking inadequately treated contaminated water, ingestion of contaminated uncooked fruits or vegetables or person-to-person spread by the faecal-oral or oral-faecal routes. The cyst stage is resistant to chlorine concentrations used in most water treatment facilities. Therefore adequate water treatment should include chemical treatment combined with filtration. Risk factors associated with *Giardia* infections include poor sanitary conditions, travel to known endemic areas, drinking inadequately treated water, day-care centres, and oral-anal sexual practices. Infections may occur in outbreaks and endemic forms within day-care centres and other institutional settings and among family members of infected children. Scrupulous attention to handwashing and treatment of all infected individuals are important in controlling the spread of infection in these settings.

Clinical Syndromes

Giardia infection can result in either asymptomatic carriage or symptomatic disease ranging from mild diarrhoea to a severe malabsorption syndrome. The incubation period before symptomatic disease develops ranges from 1 to 4 weeks. The onset of disease is sudden, with foul-smelling watery diarrhoea, abdominal cramps, flatulence and steatorrhea. Blood and pus are rarely present in stool specimens, consistent with the absence of tissue destruction. Spontaneous recovery generally occurs after 10 to 14 days, although a more chronic disease with multiple relapses may develop. This is particularly a problem for patients with IgA deficiency or intestinal diverticuli.

Treatment, Prevention and Control

It is important to eradicate *Giardia* from both asymptomatic carriers and diseased patients. Prevention and control of giardiasis involves avoidance of contaminated water and food, especially by the traveller and the outdoors person. Protection is afforded by boiling drinking water from streams and lakes or in countries with a high incidence of endemic disease. Maintenance of properly functioning filtration systems in municipal water supplies is also required, because cysts are resistant to standard chlorination procedures. Public health efforts should be made to identify the reservoir of infection to prevent spread of disease. In addition, high-risk sexual behaviour should be avoided.

Chapter 4

Statistical Overview

The data we have here are faecal specimens collected for a follow up period of one year in each of the two nursery schools. Specimens collection started in nursery school 2 for the first year commencing in February 1985 and in school 1 for the following year commencing February 1986.

One hundred and seventy three children were enrolled in the two schools during the study period. Six parents did not consent to their child's participation in the study, which left us with one hundred and sixty seven children only. This gives a participation rate of 96.5%.

Two kinds of data were available for each child demographic data and the history of disease over the study period. The first kind contained information about the child's parents, siblings, health report (vaccination and immunological history), pets, if any, and whether the child was breast or bottle fed and the type of milk they currently drank both at home and at school.

The second section of the data contained information about the child at the nursery school. The number of fortnights the child was enrolled at school, the number of days he/she was present, the number of holiday and bank holiday days. It also contained information about whether the child gave a specimen on every fortnight of his/her enrolment and the laboratory results of the specimen analysis. If the child was ill during that fortnight, the record also stated the kind of illness the child suffered from, whether seen by a doctor, if medicine was prescribed and if any member of the family was affected. It also contained the date of onset of illness in the child and the date of onset of symptoms at home. If any gastro-enteritis symptoms were present at the child's family, a specimen from the ill relative was also taken. Figure 4.1 contains an example of one of the children's records. The name has been changed to presume anonymity.

The first line of figure 4.1 contains the child's name, date of birth, sex and the number of siblings the child had at the beginning of the study. Next, it contains a table of eight columns. The first column contains a list of fortnight numbers on which the child was enrolled in the study. Columns 2 to 5 contain the number of days the child was present at the nursery, absent, the number of days that were school holidays and the number of days that were bank holidays respectively.

The sixth column in the first part of figure 4.1 contains an "S" if the child gave a specimen at the specified fortnight or an "N" if he did not. If the child gave a specimen on any fortnight, column 7 would contain a list of agents names and serotypes that were found in the laboratory analysis. If no agent was found "NAD" would be written in this column. The last column contains any remarks from the nursery

Alison McDonald

24-JAN-85 F 0

1	9	0	0	1	S	REOVIRUS ADENO 99 SRV ADENO
2	9	0	0	1	S	SRV
3	10	0	0	0	S	SRV
4	10	0	0	0	S	NAD
5	10	0	0	0	S	NAD
6	9	0	0	1	S	ADENO 1
7	10	0	0	0	S	ADENO 1
8	10	0	0	0	S	ADENO 99
9	10	0	0	0	N	
10	5	5	0	0	S	COLI 0111 ADENO 3
11	4	6	0	0	S	ADENO 1
12	9	0	0	1	S	ECHOVIRUS 18 ADENO 1
13	10	0	0	0	S	COLI 0111 ECHOVIRUS 18
14	4	6	0	0	S	ECHOVIRUS 18
15	6	4	0	0	S	CSB 2
16	9	1	0	0	S	CSB 2 ADENO
17	8	2	0	0	S	CRYPTO
18	3	7	0	0	S	CRYPTO ECHOVIRUS 22
19	7	3	0	0	S	NAD
20	3	2	0	0	N	
21	7	0	3	0	N	
22	5	0	5	0	N	
23	10	0	0	0	N	
24	10	0	0	0	S	SRV
25	10	0	0	0	N	
26	10	0	0	0	N	
1	N					
2	N					
3	N					
4	N					
5	N					
6	N					
7	N					
8	N					
9	N					
10	N					
11	N					
12	N					
13	N					
14	Y	19-SEP-86	N	N	N	N VOM
15	N					
16	N					
17	Y	05-NOV-86	N	N	Y	Y LS STL 03-NOV-86
18	N					
19	N					
20	N					
21	N					
22	N					
23	N					
24	N					
25	N					
26	N					

END

Figure 4.1: An example of a child's record

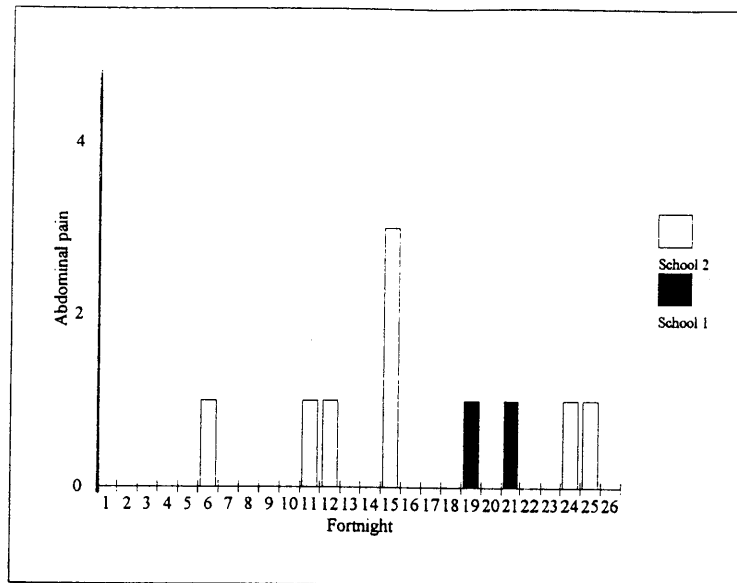


Figure 4.2: Abdominal pain by fortnight

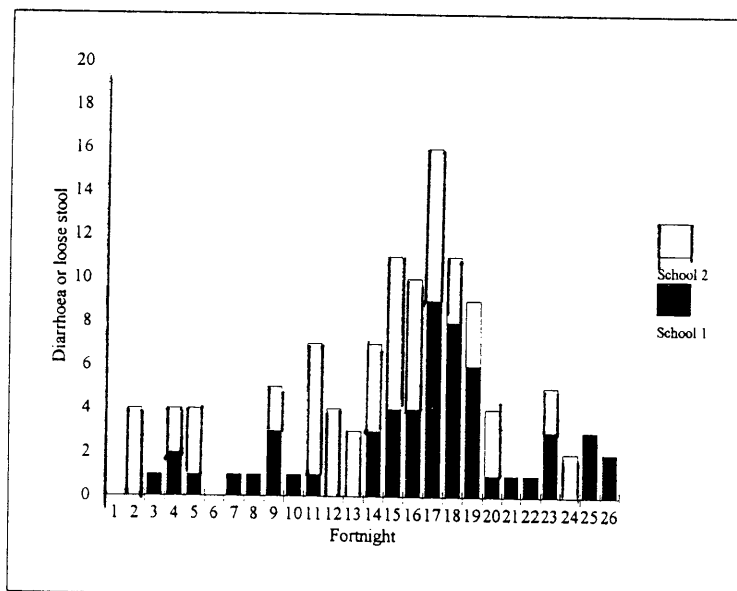


Figure 4.3: Diarrhoea or Loose Stool by fortnight

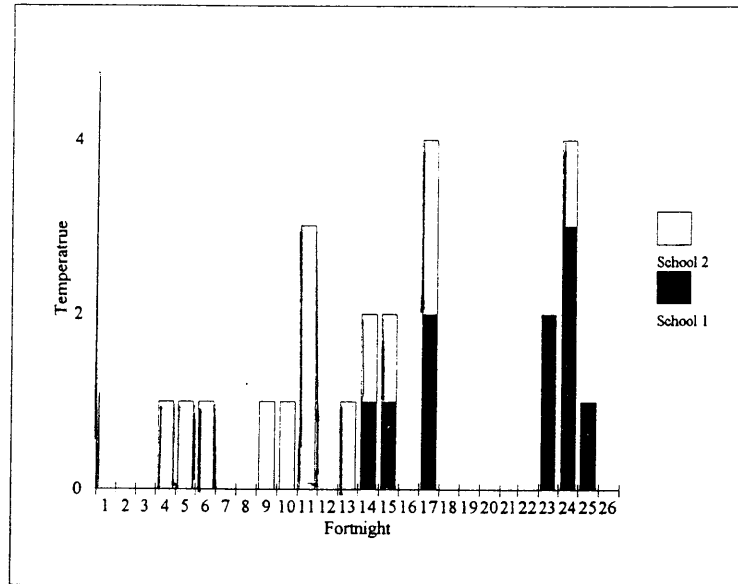


Figure 4.4: Temperature by fortnight

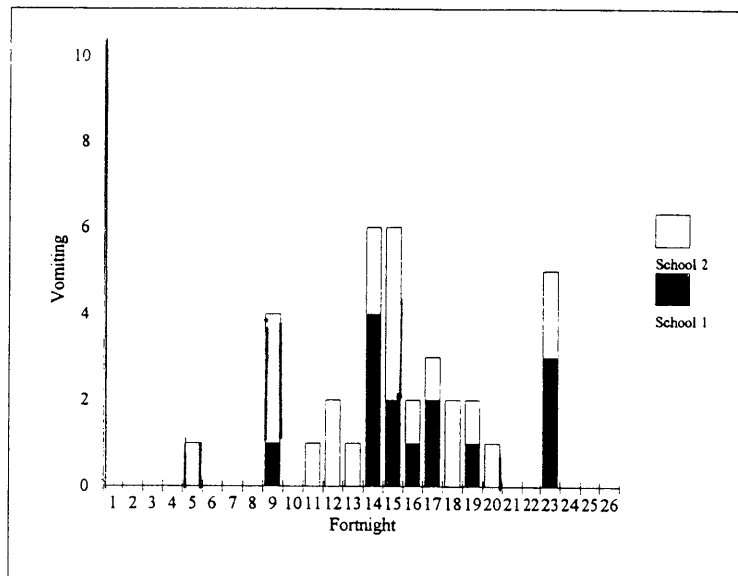


Figure 4.5: Vomiting by fortnight

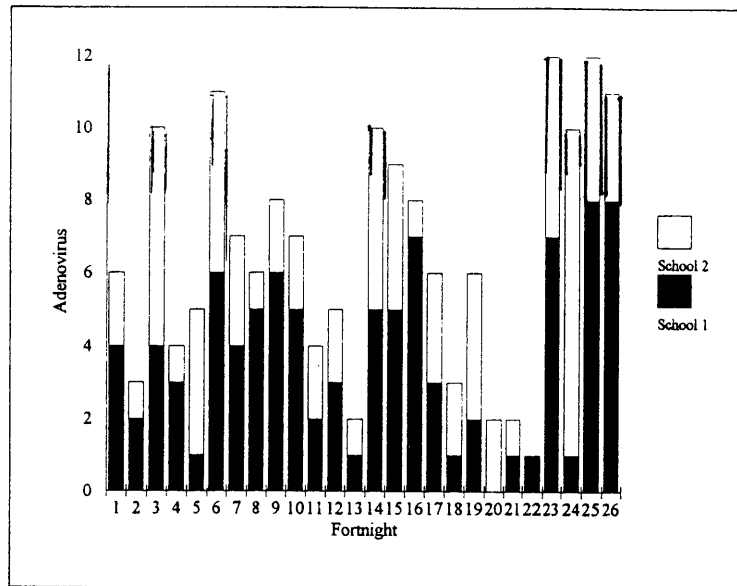


Figure 4.6: Adenovirus by fortnight

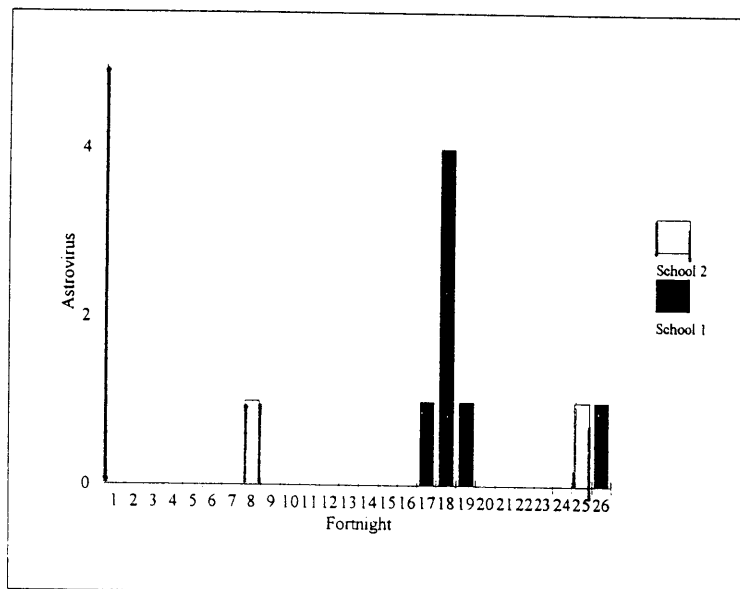


Figure 4.7: Astrovirus by fortnight

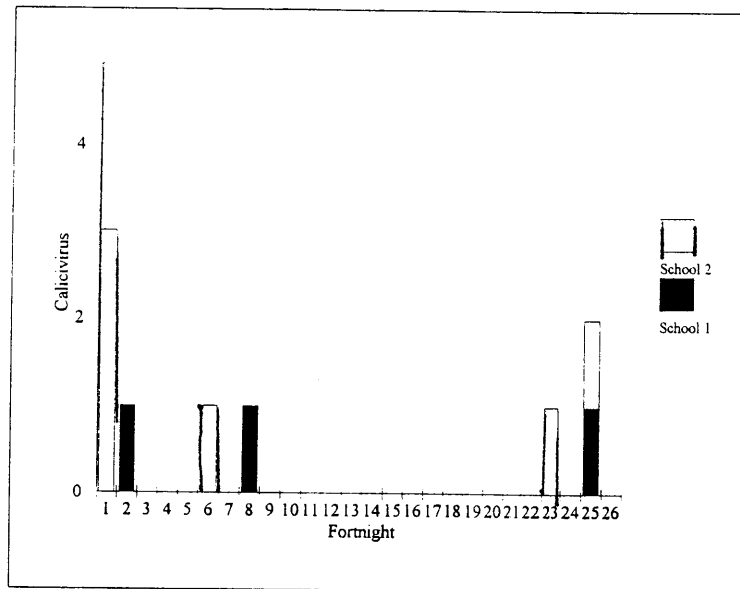


Figure 4.8: Calicivirus by fortnight

nurse. This column was discarded during the analysis.

The second part of figure 4.1 contains a list of the fortnights the child was enrolled in the study. The second column contains a list of Y's and N's stating whether the child showed any symptoms during the particular fortnight. If this column contains a "Y", the next column will contain the date of onset of symptoms followed by whether the child was examined by a doctor, given any medication, whether any of the child's family members had the same symptoms and whether a stool specimen was collected for the child's ill family member. The last column contains the date of which symptoms were noticed on the child's family member. This date is indicative of whether the child received the infection from home or not. Each child's record was ended with "END".

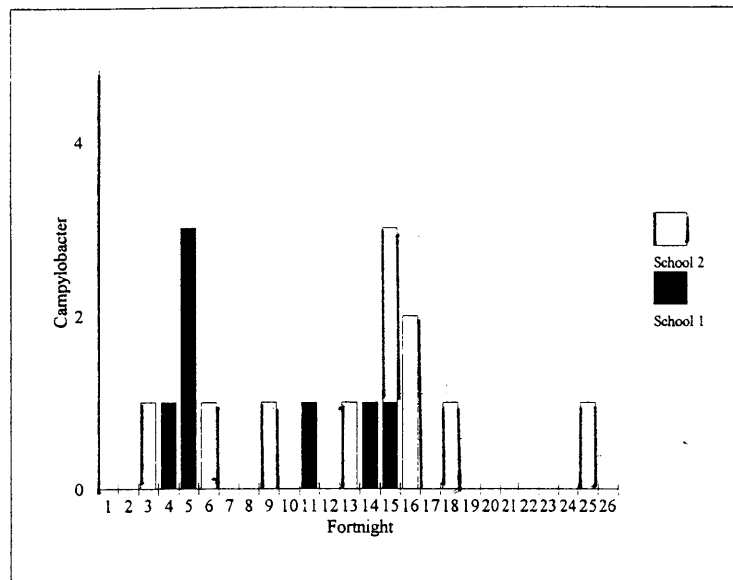


Figure 4.9: Campylobacter by fortnight

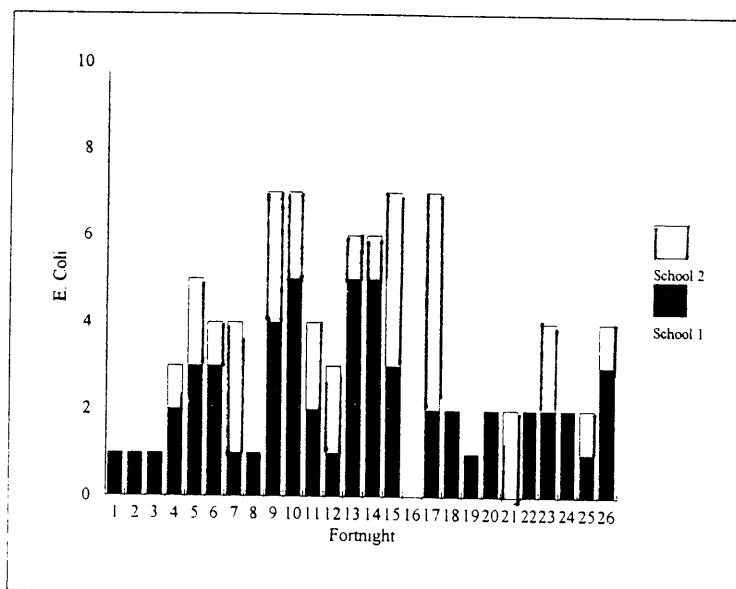


Figure 4.10: E. Coli by fortnight

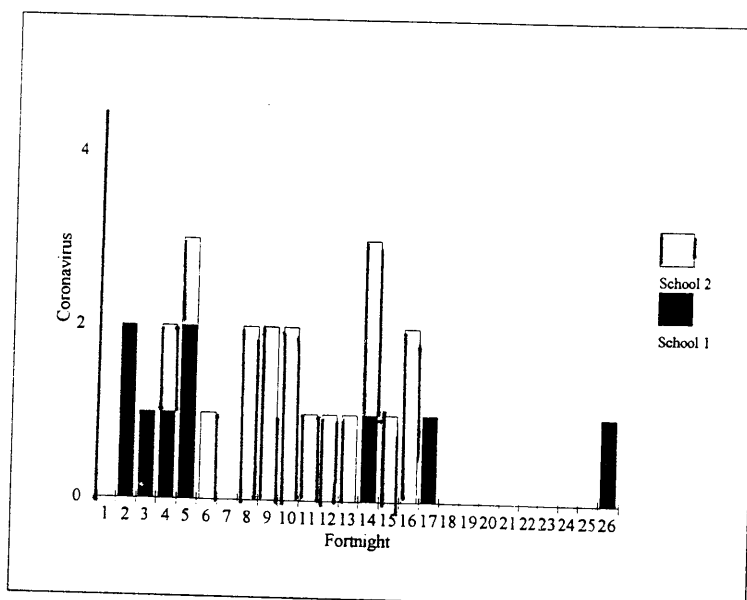


Figure 4.11: Cororna virus by fortnight

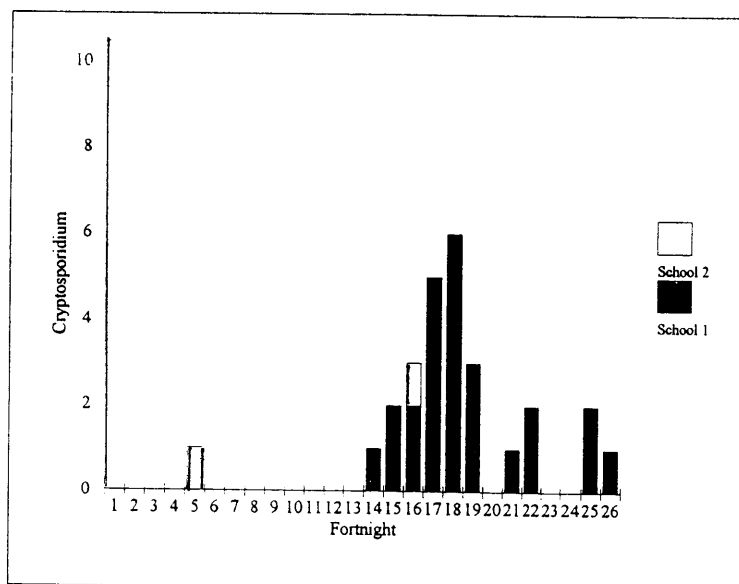


Figure 4.12: Cryptosporidium by fortnight

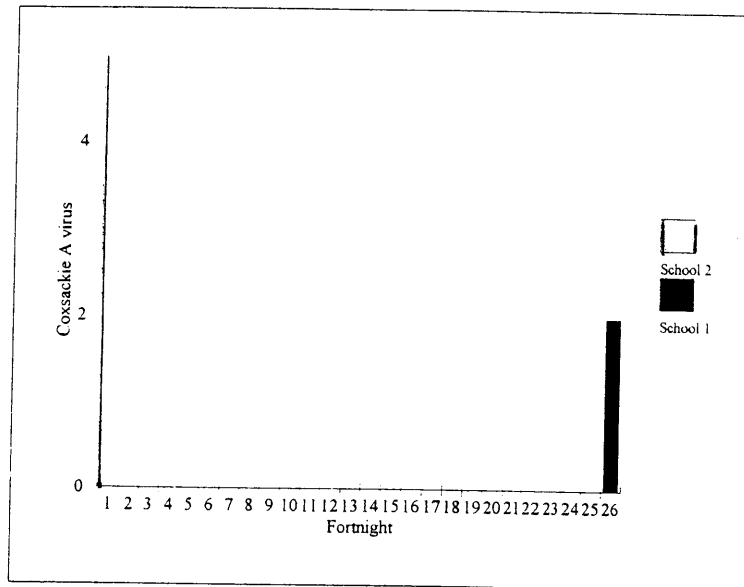


Figure 4.13: Cocksackie A virus by fortnight

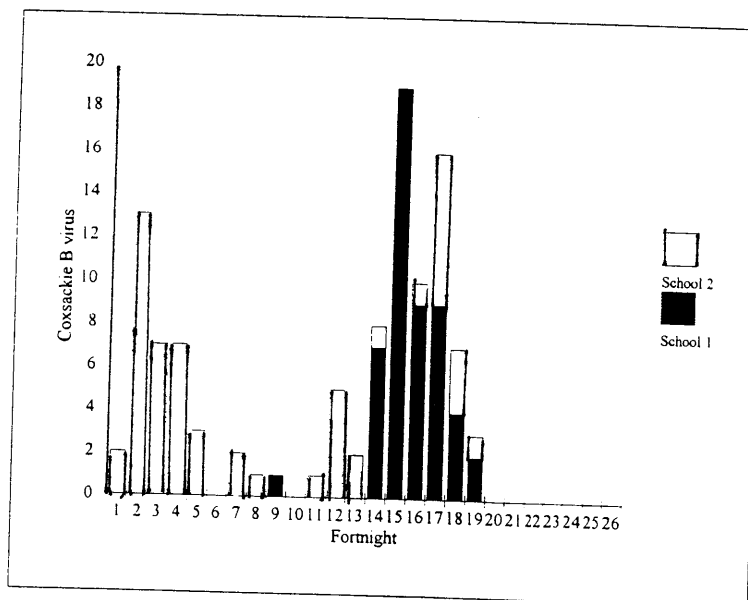


Figure 4.14: Cocksackie B virus by fortnight

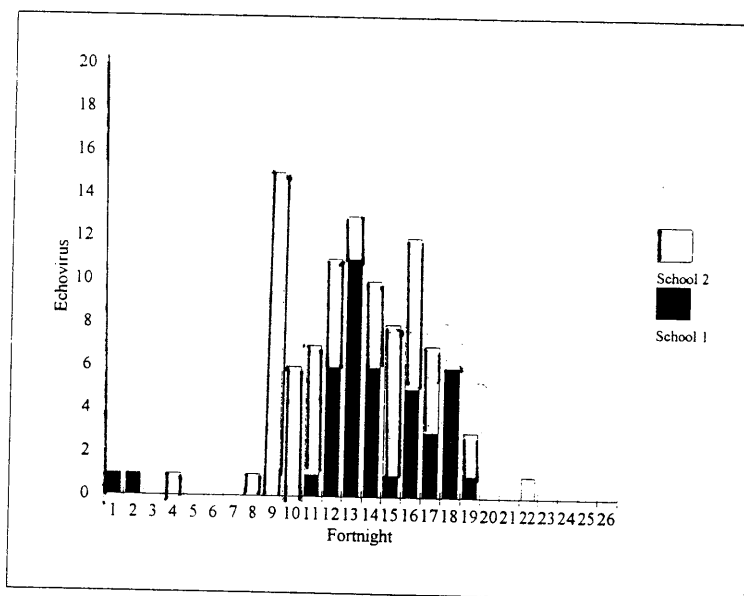


Figure 4.15: Echovirus by fortnight

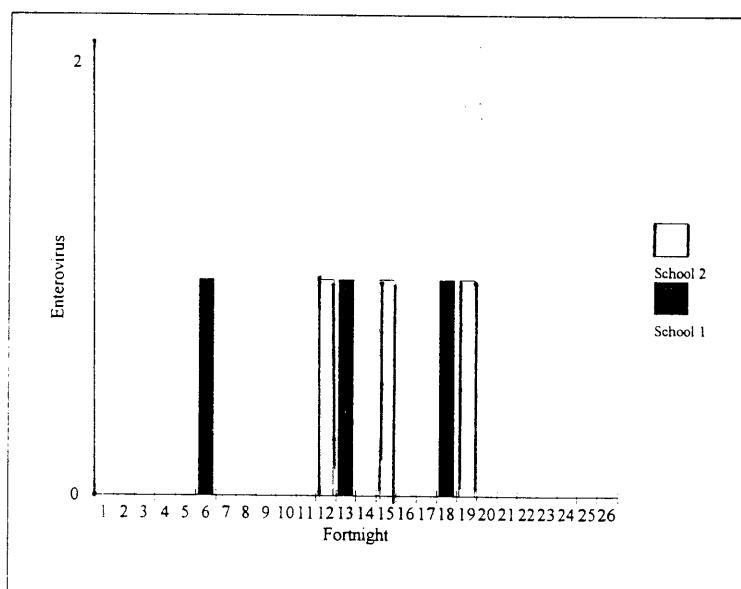


Figure 4.16: Enterovirus by fortnight

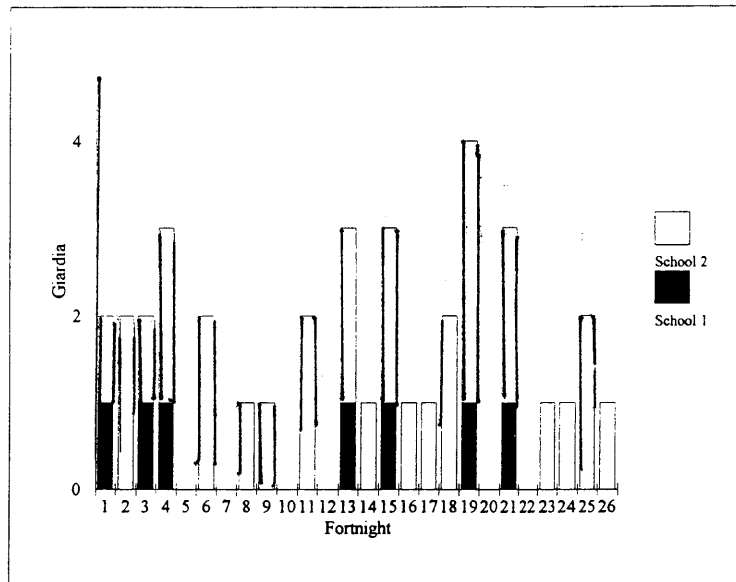


Figure 4.17: Giardia intestinalis by fortnight

Tables 4.1 to 4.4 contain summaries of the data. It can be seen from table 4.1 that 54% of the children were from school 1 and 46% were from school 2. 65% of the children from school 1 were males against 62% from school 2. Even though, the number of children in school 1 was larger, there were more specimens collected from school 2, more occurrences of illness but fewer children seen by a doctor. Occurrence of illness here includes any one or more symptoms occurring simultaneously to a child.

Frequencies of symptoms are all listed in table 4.2. It can be seen that diarrhoea or loose stool had the highest frequency of occurrence, followed by vomiting, then temperature. Symptoms like abdominal swelling, nausea, and swollen glands occurred only once during the two year study period. About 48% of the diarrhoeal cases occurred in school

1 and approximately 40% of the vomiting cases.

Table 4.3 displays the distribution of the different microorganisms that were found in stool specimens in the two schools. In this table different serotypes are not distinguished. It can be seen that *Adenovirus* was the most frequently isolated virus and *Escherichia Coli* had the highest frequency of bacterial isolation. *Giardia intestinalis*, *Echovirus*, *Campylobacter* and *Coronavirus* were found more in school 2 whereas *Cryptosporidium*, *Reovirus* and *Astrovirus* were isolated more from specimens of school 1.

Table 4.4 shows the frequencies of occurrences of the different symptoms related to gastro-enteritis with different agents that are thought, from microbiological literature, to be the cause of such symptoms. Entries are ordered by school in alphabetical order of agent's name. Here *Coxsackie B* virus, *Echovirus*, *Cryptosporidium* and *Adenovirus* have the highest incidence of occurrence associated with diarrhoea. *Shigella sonnei*, which is known to be a cause of diarrhoea, was associated with it only on three occasions. This is due to the lower number of isolations of this bacterium. *Coxsackie B* virus was also frequently associated with temperature and vomiting. *Adenovirus*, *Echovirus*, and *Campylobacter* were associated with all the four symptoms. This association was later confirmed using more sophisticated statistical methods.

On a few occasions children complained of more than one symptom simultaneously. Table 4.5 shows the frequencies of such symptoms. It can be seen from this table that in approximately 39% of diarrhoeal cases the child also vomited. Only on three occasions the child experienced both diarrhoea and abdominal pain.

	School 1	School 2	Total
No. of children	90	77	167
No. of Males	59	48	107
No. of Females	31	29	60
No. of Specimens	556	619	1175
No. of Occurrence of Illness	85	96	181
No. of Doctor Involvements	61	48	109
No. of Families Affected	24	12	36

Table 4.1: Summary of the data

Figures 4.2 to 4.5 display stacked histograms of the distribution of the reported symptoms per fortnight in the two schools together.

From figure 4.2, it can be seen that school 1 had fewer attacks of abdominal pain compared to school 2. Also the highest number of abdominal pain attacks was 3 in the fifteenth fortnight in school 1. As had been said earlier diarrhoea or loose stool had the maximum number of attacks among the other symptoms. It can be seen from figure 4.3 that children in school 1 had suffered from diarrhoea all year long with the highest intensity around fortnights 14 to 19; whereas in school 2, the highest frequency of diarrhoeal attacks among the children were between fortnights 11 and 20. Children in school 1 did not suffer from high temperature until fortnight 14 but in school 2, children had high temperature more frequently than school 1. This was concluded from figure 4.4. Figure 4.5 tells that in school 1 during the fortnights with high frequencies of reported diarrhoea some children also vomited. The same may also be said about school 2.

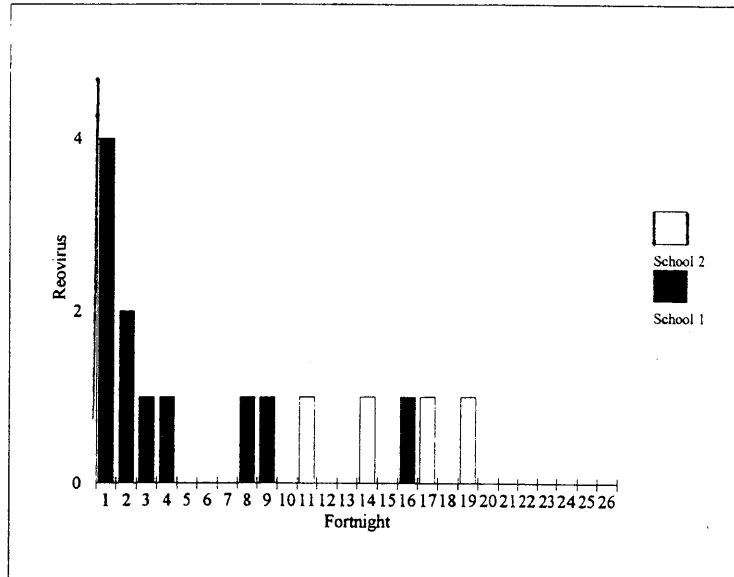


Figure 4.18: Reovirus by fortnight

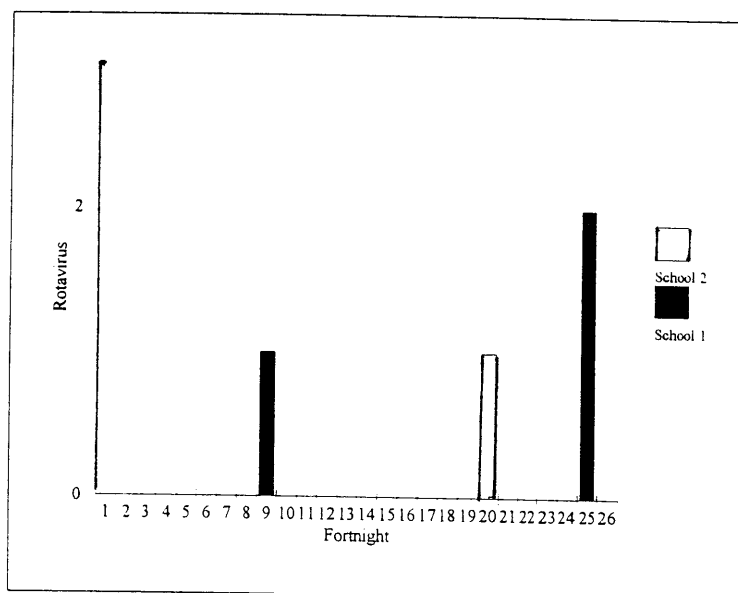


Figure 4.19: Rotavirus by fortnight

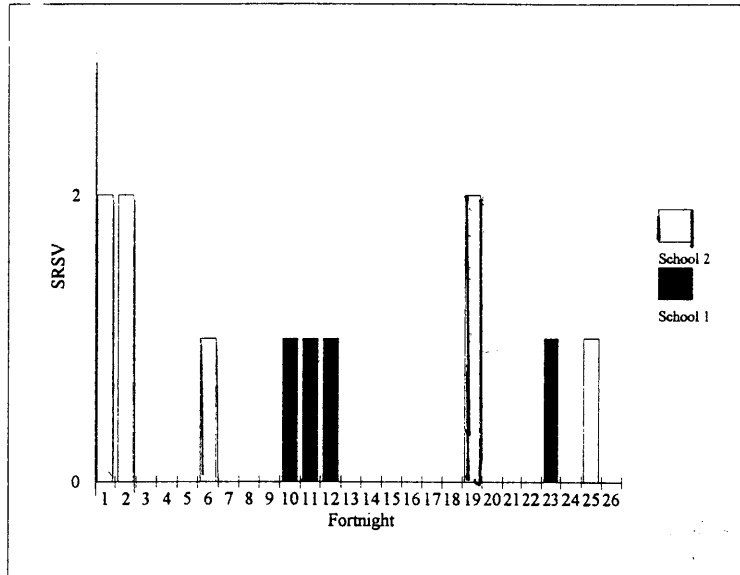


Figure 4.20: SRSV by fortnight

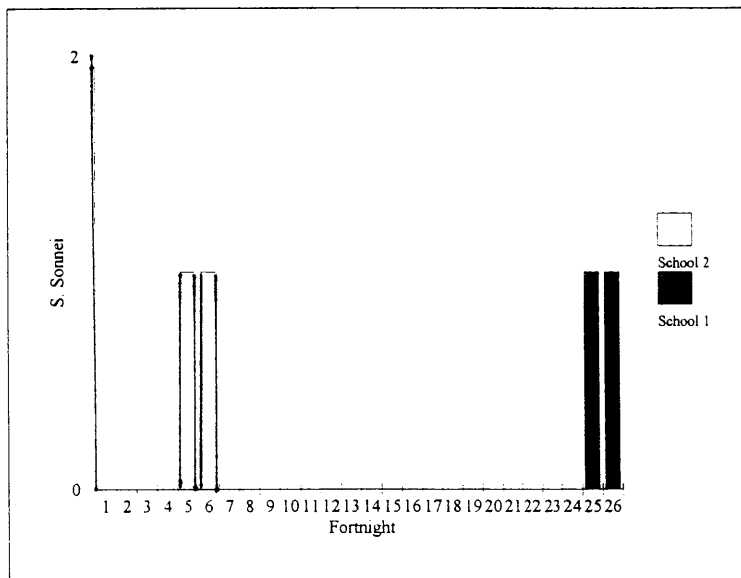


Figure 4.21: S. Sonnei by fortnight

Symptoms/School	School 1	School 2	Total
Diarrhoea or Loose Stool	56	61	117
Vomiting	14	22	36
Temperature	10	14	24
Abdominal Pain	2	8	10
Loss of Appetite	4	2	6
Whooping Cough	4	0	4
Convulsion	1	2	3
Lethargy	1	1	2
Loss of Colour	1	1	2
Bloody Stool	0	2	2
Pale Stool	0	2	2
Abdominal Swelling	0	1	1
Blood Infection	0	1	1
Ear Infection	1	0	1
Nausea	1	0	1
Swollen Glands	0	1	1

Table 4.2: Distribution of symptoms between the two schools

Agent/School	School 1	School 2	Total
Adenovirus	95	75	170
Coxsackie B	51	56	107
Echovirus	42	61	103
Escherichia Coli	55	33	88
Giardia intestinalis	7	31	38
Cryptosporidium	25	2	27
Coronavirus	9	17	26
Campylobacter	7	10	17
Reovirus	11	4	15
Small Rounded Structured Virus (SRSV)	4	8	12
Astrovirus	7	2	9
Calicivirus	3	6	9
Enterovirus	3	3	6
Rotavirus	3	1	4
Spirochetes	2	2	4
Shigella Sonnei	2	2	4
Coxsackie A	2	0	2
Cysts	1	1	2
Yersinia	0	1	1

Table 4.3: Occurrences of agents in the two schools

Symptom	Agent	School 1	School 2
Diarrhoea or Loose Stool	Adenovirus	8	7
	Astrovirus	3	0
	Campylobacter	2	2
	Escherichia Coli	5	3
	Coronavirus	3	1
	Cryptosporidium	14	1
	Coxsackie B	13	5
	Echovirus	4	12
	Enterovirus	0	1
	Giardia intestinalis	1	4
	Rotavirus	2	0
	SRSV	0	2
	Reovirus	0	1
	Spirochetes	1	0
	Shigella Sonnei	2	1
Temperature	Adenovirus	1	1
	Campylobacter	0	1
	Coxsackie B	2	1
	Echovirus	0	2
	Giardia intestinalis	0	1
	Shigella Sonnei	0	1
Vomiting	Adenovirus	3	3
	Campylobacter	0	1
	Escherichia Coli	4	1
	Cryptosporidium	4	1
	Coxsackie B	5	1
	Echovirus	2	2
Abdominal Pain	Adenovirus	0	2
	Cambpylobacter	0	2
	Echovirus	1	1
	Giardia intestinalis	0	1

Table 4.4: Frequencies of occurrences of symptoms and agents

Symptom	Is Present With	Frequency
Abdominal Pain	Diarrhoea or Loose Stool	3
	Temperature	2
	Vomiting	1
Diarrhoea or Loose Stool	Temperature	1
	Vomiting	16
Temperature	Vomiting	1

Table 4.5: Distribution of symptom occurred simultaneously

Parents	Number Employed			Total
	0	1	2	
Both Parents	37	29	16	82
Mother Only	27	38	0	65
Father Only	2	1	0	3
Mother and Grandparents	3	8	0	11
Others	2	4	0	6
Total	71	80	16	167

Table 4.6: Distribution of children by parents and parents employment rates

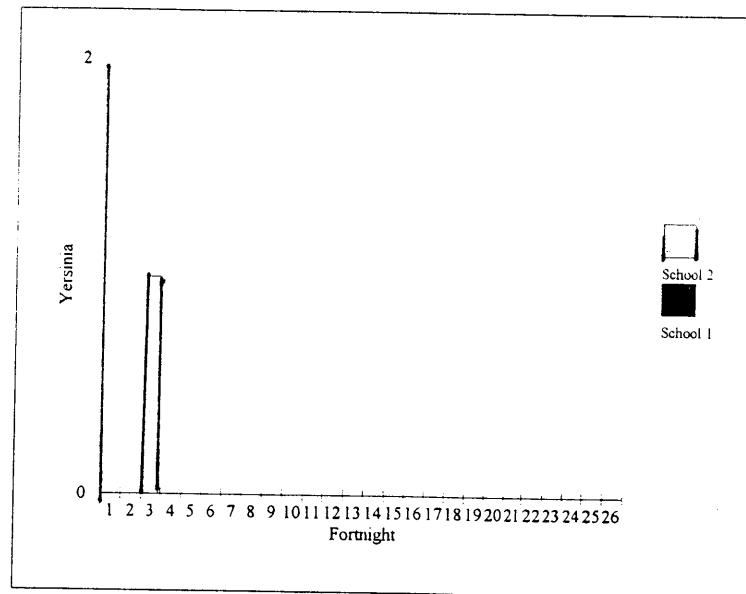


Figure 4.22: Yersinia by fortnight

Figures 4.6 to 4.22 display stacked histograms of the frequency of the different agents that were isolated from the children's faecal specimens during the laboratory analysis per fortnight. As can be seen from figure 4.6 *Adenovirus* was isolated from those specimens throughout the year, and so was *Escherichia coli*, from figure 4.10. *Echovirus* was mostly isolated from the two schools around the same time as from figure 4.15. In figure 4.17 it can be seen that *Giardia intestinalis* was mostly isolated from specimens from school 2. Specimens were tested positive for *Giardia intestinalis* for the whole year but not in school 1. In figure 4.14 *Coxsackie B* virus was only isolated between fortnights 14 to 19 in school 1. This may have a relation with the reported number of episodes of diarrhoea in the same school at the same time. The rest of figures 4.6 to 4.22 display the remaining microorganisms that were mentioned in table 4.3.

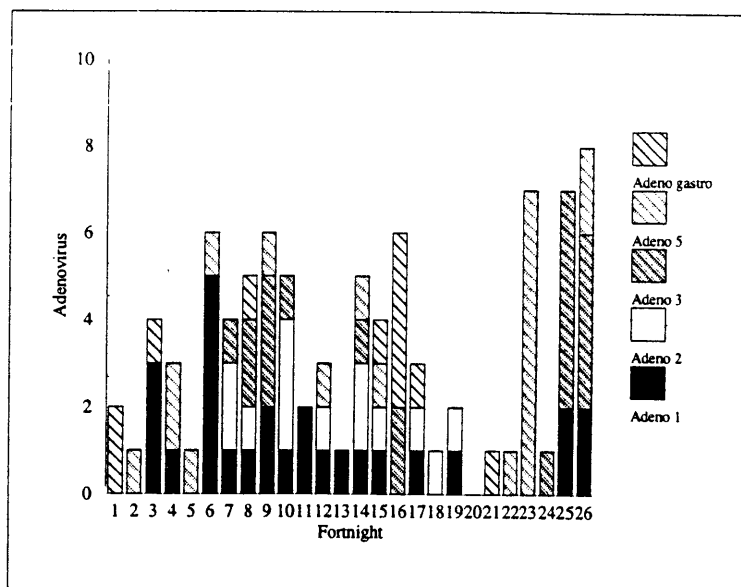


Figure 4.23: Adenovirus serotypes for school 1 by fortnight

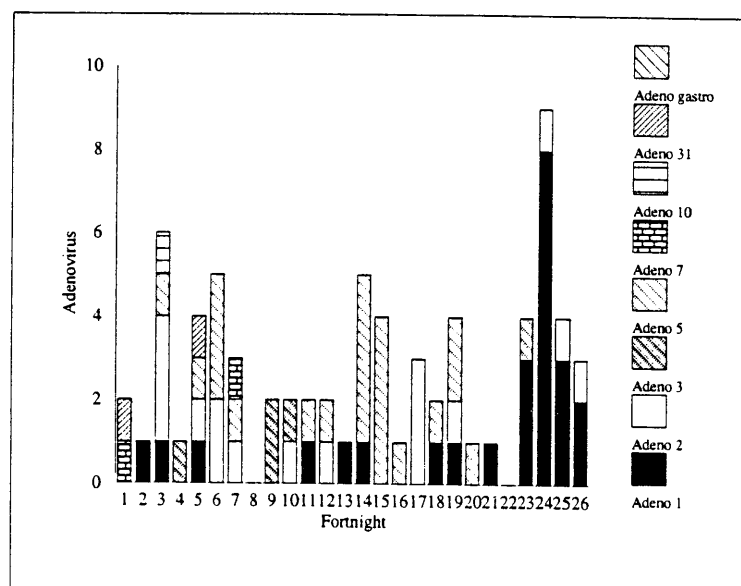


Figure 4.24: Adenovirus serotypes for school 2 by fortnight

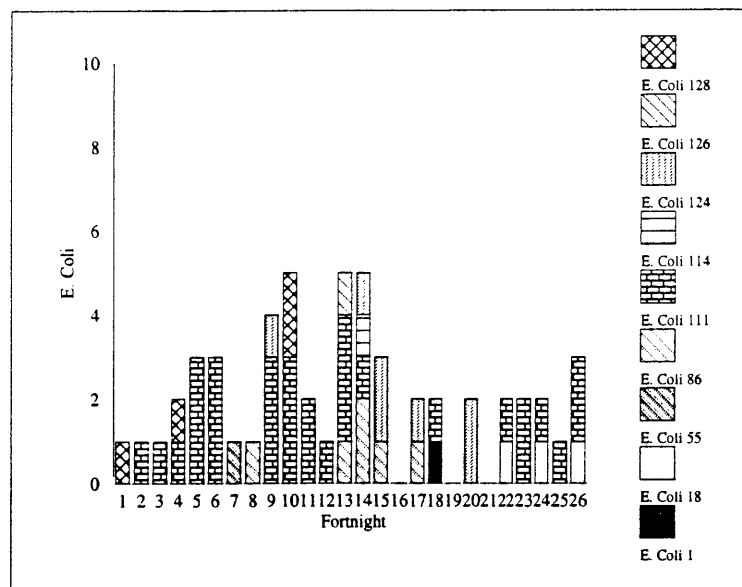


Figure 4.25: E. Coli serotypes for school 1 by fortnight

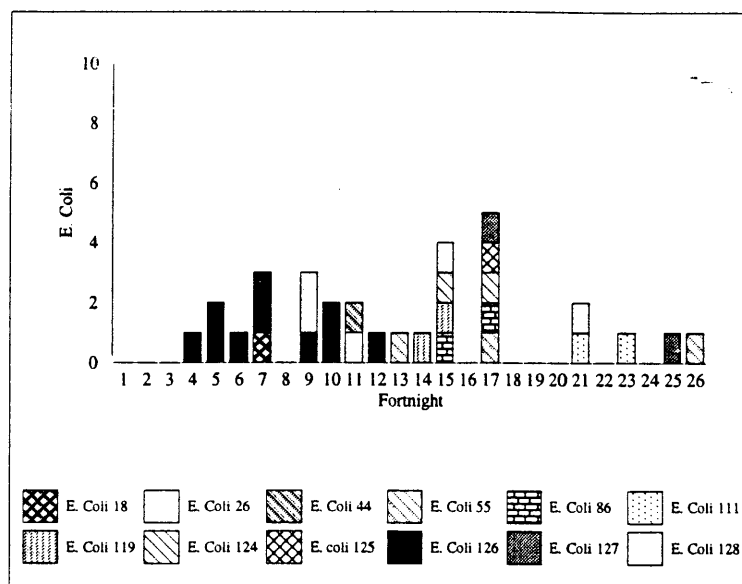


Figure 4.26: E. Coli serotypes for school 2 by fortnight

Figures 4.23 to 4.30 display stacked histograms of the frequencies of the number of isolates of different serotypes of each of *Adenovirus*, *Escherichia coli*, *Coxsackie B* virus and *Echovirus* in each school separately. From figures 4.23 and 4.24, four different serotypes of *Adenovirus* were isolated from school 1 and seven from school 2. *Adenovirus* serotypes 1, 2, 3 and 5 were isolated from both schools where 7, 10 and 31 were isolated from specimens from school 2 only. This could either mean that children from school 1 were immune against these serotypes or that such serotypes were not circulating in the community during specimen collection time in school 1, since it started a year later than at school 2. These three serotypes were isolated fewer times than the other ones. The same can be said about *Escherichia coli*, more serotypes being isolated from school 2 as can be seen from figures 4.25 and 4.26. In school 1 *Escherichia coli* serotype O111 had the highest number of occurrences in comparison with school 2 in which serotype O126 occurred more often.

From figures 4.27 and 4.28, there were four different serotypes of *Coxsackie B* virus isolated. The strange thing about these serotypes is that each two were isolated from a different school. *Coxsackie B* virus serotypes 2 and 3 were isolated from specimens collected from school 1 and serotypes 4 and 5 were from school 2. The only explanation for this is that either serotypes 2 and 3 were not circulating in the community during the specimen collection time in school 2 or the children had acquired immunity against these serotypes prior to our study. The same here applies to serotypes 2 and 3. It is also apparent from these figures that both serotypes isolated from school 1 were found around the same time. The case in school 2 is a bit different, *Coxsackie B*

virus serotype 4 was isolated earlier in the year than serotype 5. This suggests that there might be an epidemic of these serotypes around the time of their isolation.

It can be noticed from figures 4.29 and 4.30 that there were seven different serotypes of *Echovirus* isolated from the two schools. The number of different serotypes isolated from school 2 was more than those from school 1. *Echovirus* serotypes 14 and 31 were only isolated from specimens in school 2, but the frequency of these isolations is very low compared to the other serotypes. *Echovirus* serotype 18 had the highest number of occurrences in school 1 but was isolated only once in school 2. *Echovirus* serotype 9 was not found in specimens from school 2 but was isolated only once in school 1.

Table 4.6 displays the number of children who were single-parented or lived with both parents broken down by the number of parents employed. It can be seen that only 49.1% of the children lived with both parents. 6.6% of the children lived with their mother and grandparents. Out of the remaining children, 38.9% lived with their mothers only. 42.5% of the children lived with unemployed parents. Twenty-seven single mothers were unemployed.

It was also noticed from the records that 85% of the children lived in council accommodation and 12.6% lived in private houses only.

It was stated in an earlier chapter, that some of the children had enrolled after data collection started in their nursery school. Also some of the children dropped out before it finished. Table 4.7 shows the total number of child-fortnights broken down by age and sex. In this

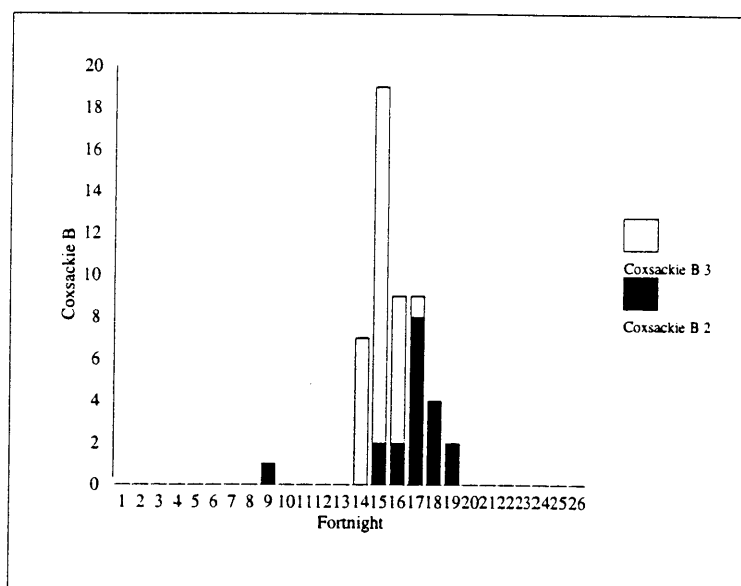


Figure 4.27: Cocksackie B virus serotypes for school 1 by fortnight

table, fortnights attended means the number of times the children were present at school for the whole fortnights, i.e. the number of days the child was present equals 10 per fortnight. 89.1% of the male children under 2 years of age attended for the whole fortnight and 96.2% of the females in the same age group. Specimens were collected from 42.9% of the males who were two years or older and from 51.9% of the females in the same age group compared to 57.1% of the males and 59.2% of females who were younger than 2 years.

Table 4.8 shows the incidence density of reported gastro-enteritis broken down by age and sex. There were 39 cases of gastro-enteritis reported in males under two years of age. The number of episodes per child-year is the result of dividing the total number of fortnights enrolled by 26 (the number of fortnights per year) and dividing the

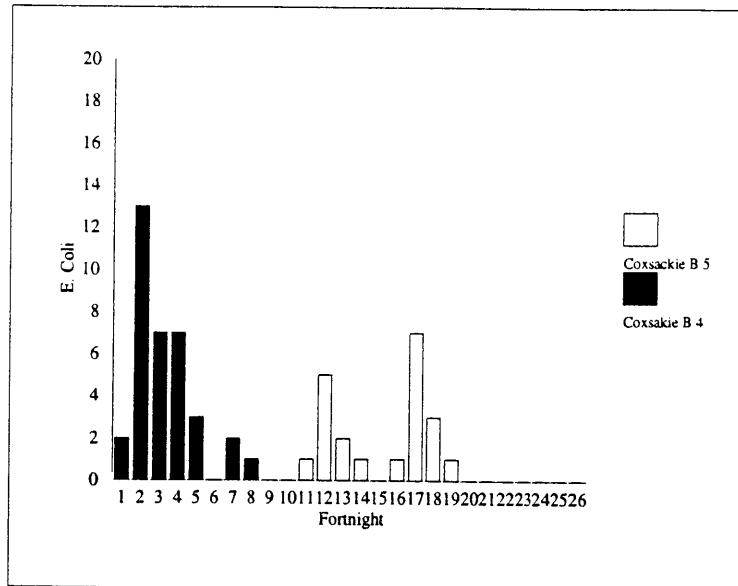


Figure 4.28: Coxsackie B serotypes for school 2 by fortnight

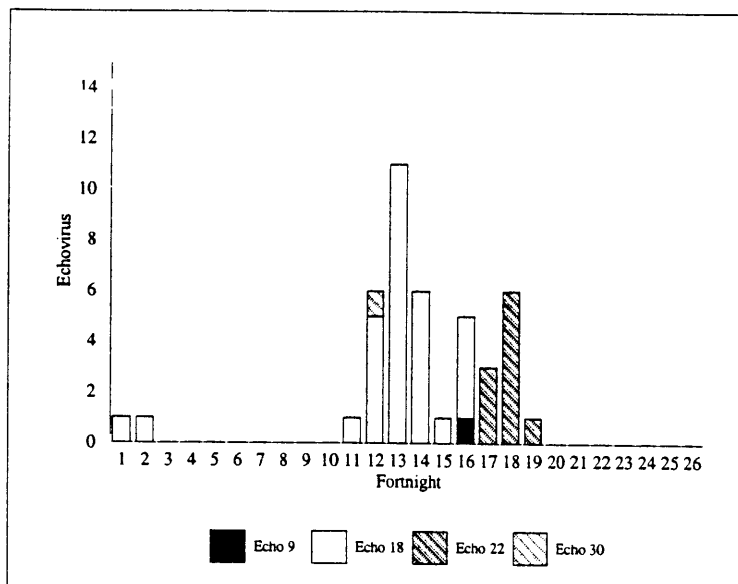


Figure 4.29: Echovirus serotypes for school 1 by fortnight

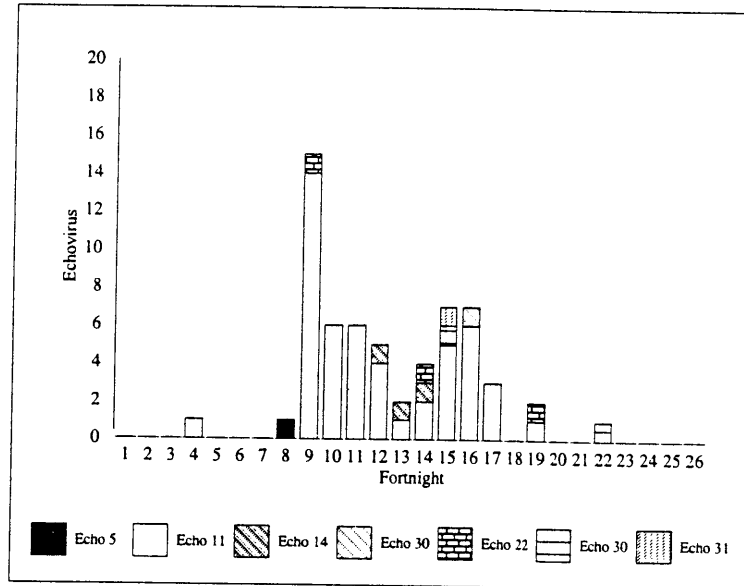


Figure 4.30: Echovirus serotypes for school 2 by fortnight

quotient by the number of fortnights with reported gastro-enteritis. It can be concluded from this table that this number is higher in younger children.

Figure 4.30 displays a frequency histogram of the number of specimens collected from the children each fortnight by child's gender. From this figure, the mean number of specimens collected from males was 25.42 per fortnight and from females was 19.77 with standard deviations of 4.97 and 5.30 respectively. Figure 4.31 shows a histogram of the frequency distribution of the number of fortnights enrolled by each child. It can be seen from this histogram that the mean number of fortnights enrolled by a child is 7.26 and the standard deviation is

	Aged 2 years or less		More than 2 years		Total
	Male	Female	Male	Female	
fortnights enrolled	393	432	1077	544	2446
fortnights attended	350	407	1039	518	2314
faecal specimens collected	200	241	446	264	1151
percentage attendance	89.1%	96.2%	96.5%	95.2%	94.6%
percentage specimen collection	57.1%	59.2%	42.9%	51.0%	47.1%

Table 4.7: Enrolment, attendance and specimen collection.

	Aged 2 years or less		More than 2 years		Total
	Male	Female	Male	Female	
fortnights with reported G-E	39	33	36	29	137
fortnights enrolled	350	407	1039	518	2446
episodes of G-E per child-year	2.05	1.48	0.27	0.81	1.17
95% CI	1.56 to 2.68	1.02 to 2.06	0.12 to 0.53	0.44 to 1.38	0.96 to 1.40

Table 4.8: Incidence density of reported gastro-enteritis.

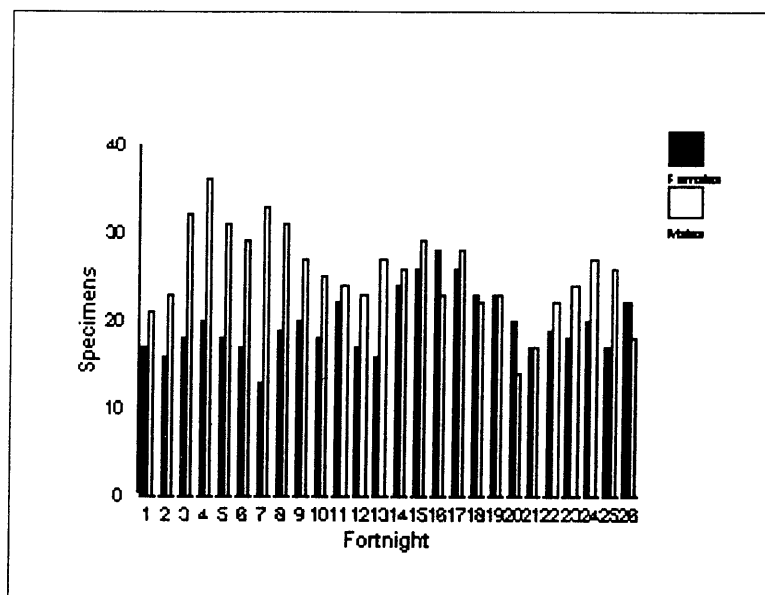


Figure 4.31: Distribution of specimens collected per fortnight

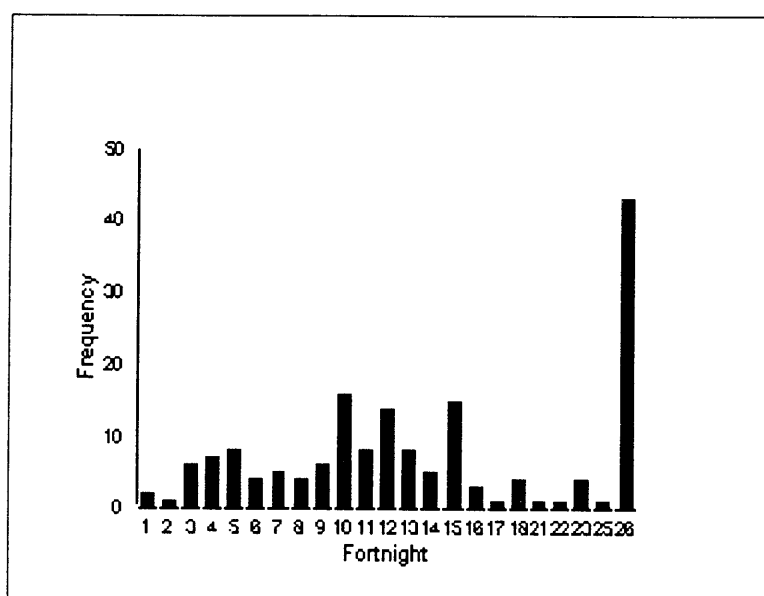


Figure 4.32: Distribution of number fortnights enrolled

Age	< 1	1	2	3	4	Total
M	5 (4.7)	28 (26.7)	36 (34.3)	28 (26.6)	8 (7.6)	105 (64.0)
F	2 (3.4)	23 (39.0)	20 (33.9)	10 (16.9)	4 (6.8)	59 (36.0)
Total	7 (4.2)	51 (31.1)	56 (34.1)	38 (23.2)	12 (7.3)	164 (100.0)

Table 4.9: The study population. Age at entry to the study, years (%). Ages of 2 males and 1 female are missing.

8.74. Around 25% of the children were enrolled for the whole year.

Table 4.9 shows the distribution of the ages of the children at the beginning of their enrolment in our study broken down by age and sex. The highest percentage of children is among the two year olds which is the ideal age for nursery schools followed by the one year olds.

Table 4.10 shows the distribution of the number of siblings each child had. The average number of siblings was 1.27. 31.1% of the children were the only child and approximately 9% of the children came from a family with four or more children. These figures are consistent with national statistics which state that 24% of the families in 1986 had two or less children and 22% had three or more children (Birth Statistics)(1991)

Siblings	0	1	2	3	4	5	6	Total
Number of subjects	52	61	32	7	9	5	1	167
Percentage	(31.1)	(36.5)	(19.2)	(4.2)	(5.4)	(3.0)	(0.6)	(100)

Table 4.10: The study population. Percentage of siblings.

Organism	SOLE ISOLATE		WITH OTHER ORGANISMS		TOTALS	
	Symp	No Symp	Symp	No Symp	Symp	No Symp
Adenovirus	13	104	8	45	21	149
Echovirus	14	56	8	25	22	81
Coxsackie B virus	11	52	14	30	25	82
E coli	4	46	8	30	12	76
Giardia	3	16	3	16	6	32
Coronavirus	2	15	2	7	4	22
Campylobacter	6	6	1	4	7	10
Cryptosporidium	5	7	11	4	16	11
SRSV	0	2	3	7	3	9
Reovirus	1	6	0	8	1	14
Calicivirus	1	4	0	4	1	8
Astrovirus	0	5	3	1	3	6
Spirochaetes	0	2	1	1	1	3
Cysts	0	0	0	2	0	2
Rotavirus	1	2	1	0	2	2
Shigella sonnei	1	1	2	0	3	1
Yersinia	0	0	1	0	1	0

Table 4.11: Association between symptoms and isolation of specific organisms during the same fortnight.

Finally, table 4.11 shows the frequencies of isolation of the organisms either alone or with others and the frequencies of their occurrence with any symptoms reported. Symptoms here are taken to be either diarrhoea or loose stool, abdominal pain, vomiting or temperature. It can be noticed here that 87.6% of the time that *Adenovirus* was isolated, there were no reported symptoms. On 62% of the times when specimens were positive for *Adenovirus* and symptoms were present, no other organisms were isolated with it. On almost 79% of the times that *Echovirus* was isolated there were no reported symptoms and on 69% of these occasions, it was the sole isolate. At 9 of 12 times when SRSV was isolated there were no reported illnesses. *Escherichia coli* was associated with illness only 12 out of 88 times it was found in the stool samples and on 33% of these occasions it was isolated alone. Out of the three times *Rotavirus* was isolated alone, it was associated with symptoms on two times. It can also be concluded that 41.2% of the time *Campylobacter* was associated with symptoms and *Giardia intestinalis* was only associated with reported symptoms only on 14% of the time.

In the next chapter, we will look at these proportions in more details.

Chapter 5

Elementary Statistical Analysis

5.1 Introduction

This chapter will discuss the elementary statistical analyses that were conducted on the Manchester data. First, it will briefly describe the data. Next, it will look at the means and standard deviations of the number of occurrences of agents and symptoms. It will also look at the correlation coefficients of the rates of occurrences of the different agents and symptoms per child with some social and economic factors. Then it will also discuss logistic regression models that were fitted to the data. It will also look at the proportions of asymptomatic carriage of each agent of infection.

5.2 The Data

The data were all typed into a text file "PROJECT.DAT" as shown in figure 5.1. A FORTRAN 90 program was written to reformat the

data and convert it to another text file of 0's and 1's corresponding to the presence or absence of each agent of infection and each symptom. This made it easier to convert this file into any statistical package data file. In this study SPSS for windows version 6.1 was used to do most of the statistical analysis. The later data file was then converted to an SPSS data file "TRANS90.SAV". This file contains records of 26 rows for each child. The variables in each row are listed in table 5.1 with a dummy value for each variable. In other words, the contents of table 5.1 are a list of the variables and one row of values in TRANS90.SAV. A new binary variable called "missing" was added to the data. It assumes the following values

$$missing = \begin{cases} 0 & \text{if child was enrolled in our study} \\ 1 & \text{otherwise.} \end{cases}$$

The data file was regularly updated and new variables were added whenever necessary. Such variables include the child's birth order and age and were sometimes derived by defining newer names for symptoms such as "diarrhoea or loose stool" by adding both diarrhoea values and loose stool or "gastro-interitis" by adding the values for diarrhoea, loose stool and vomiting.

5.3 Elementary Statistics

This section will look at the means and standard deviation of the number of attacks of each agent and the number of episodes of each symptoms per child. It will also discuss the correlation coefficients between the number of episodes of symptoms and the number of attacks of each agent with child's sex, number of siblings the child had, number of parents the child lived with, the number of parents employed, child's

Variable	child	birth day	birth month	birth year
Dummy Value	1	31	1	85
Variable	siblings	school	sex	fortnight
Dummy Value	0	1	F	1
Variable	missing	specimen	days present	days absent
Dummy Value	0	N	10	0
Variable	days bank holiday	days holiday	ill day	ill month
Dummy Value	0	0	0	0
Variable	ill year	onset day	onset month	onset year
Dummy Value	0	0	0	0
Variable	ill	doctor	family	nad
Dummy Value	0	0	0	0
Variable	adeno	sadeno	calici	srv
Dummy Value	0	0	0	0
Variable	csb	scsb	giardia	polio
Dummy Value	0	0	0	0
Variable	spolio	astro	camp	crypto
Dummy Value	0	0	0	0
Variable	srsv	nvr	coli	scoli
Dummy Value	0	0	0	0
Variable	ent	echo	secho	csa
Dummy Value	0	0	0	0
Variable	scsa	reo	corona	spiro
Dummy Value	0	0	0	0
Variable	cysts	rota	cdtx	ssonnei
Dummy Value	0	0	0	0
Variable	pus cell	spc	yersinia	red blood cells
Dummy Value	0	0	0	0
Variable	srbc	stm	diarrhoea	loose stool
Dummy Value	0	0	0	0
Variable	vomiting	temperature	ear infection	lethargy
Dummy Value	0	0	0	0
Variable	nausea	abdominal pain	pale stool	loss of appetite
Dummy Value	0	0	0	0
Variable	blood infection	swoll	convu	loss of colour
Dummy Value	0	0	0	0
Variable	abs	whooping cough	bloody stool	
Dummy Value	0	0	0	

Table 5.1: One row of the Data as it appears in the SPSS data file.

age at entry at the study, the child's family size and the number of fortnights the each child was enrolled in the study. These results are listed in tables 5.2 to 5.7.

Table 5.2 displays the means and standard deviations of the number of attacks of each agent of infection per child. It can be noticed here that *Adenovirus* had the highest mean of occurrence. Respiratory *Adenovirus* had a mean of 0.9521 isolation per child per year. *Yersinia* had the lowest average and standard deviation. This is because it was isolated only once during the whole study period.

The mean number of diarrhoeal episodes was 0.6946 per child per year as can be seen from table 5.3. This shows that the average number of diarrhoeal episodes in this study population was still lower than the average number of diarrhoeal episodes in children 4 years old and younger in the third world which was found to be 2.6 per child year (Bern 1992).

From table 5.4, it can be seen that the older the child was the less frequently he/she would get infected with most of the infectious agents found in this study as can be seen from the negative correlations between these agents and age at entry in the study. This is thought to be due to the fact that children will acquire immunity against these agents with age and will be less likely to be reinfected at an older age.

Agent	Mean	Standard Deviation
<i>Adenovirus</i> (all serotypes)	1.0180	1.5188
<i>Adenovirus</i> (gastro-intestinal)	0.0659	0.2719
<i>Astrovirus</i>	0.0539	0.2265
<i>Calicivirus</i>	0.0539	0.2265
<i>Campylobacter</i>	0.1018	0.3225
<i>E. coli</i>	0.5269	0.9171
<i>Coronavirus</i>	0.1557	0.6764
<i>Cryptosporidium</i>	0.1617	0.4948
<i>Coxsackie A virus</i>	0.0120	0.1091
<i>Coxsackie B virus</i>	0.6407	1.0426
<i>Echovirus</i>	0.6168	1.0223
<i>Enterovirus</i>	0.0359	0.1867
<i>Giardia intestinalis</i>	0.2275	1.0100
<i>Rotavirus</i>	0.0240	0.1886
SRSV	0.0719	0.2590
<i>S. sonnei</i>	0.0240	0.2182
<i>Yersinia</i>	0.0060	0.0774

Table 5.2: Mean and standard deviation of the number of fortnights of occurrence of the different agents of infection per child

Symptom	Mean	Standard Deviation
Diarrhoea or Loose Stool	0.6946	1.0280
Temperature	0.1437	0.4942
Vomiting	0.2156	0.4923
Abdominal Pain	0.5990	0.2621

Table 5.3: Mean and standard deviation of the number of episodes of symptoms per child

Agent	Age at entry	Birth Order	Number of Parents Employed	Number of Parents the child lived with
<i>Adenovirus</i> (All serotypes)	-0.2623	-0.0679	0.0923	-0.0865
<i>Adenovirus</i> (gastrointestinal)	-0.1255	0.0181	-0.131	0.0779
All the agents together	-0.2825	-0.0002	0.0658	-0.0578
<i>Astrovirus</i>	-0.1896	0.0617	0.1224	-0.0514
<i>Calicivirus</i>	-0.0420	0.0177	0.0398	-0.0514
<i>Campylobacter</i>	-0.0087	-0.0073	0.1333	-0.0572
<i>E. Coli</i>	-0.1666	0.0808	0.0304	-0.0237
<i>Coronavirus</i>	0.0162	0.0172	-0.1858	-0.0131
<i>Cryptosporidium</i>	-0.1732	-0.0761	-0.0020	-0.0062
<i>Corsackie A</i>	-0.0346	-0.0374	-0.1150	0.0087
<i>Corsackie B</i>	-0.2193	0.0029	0.1008	0.0133
<i>Echovirus</i>	-0.2035	0.0158	0.1091	-0.0818
<i>Enterovirus</i>	-0.0582	0.0943	-0.0012	0.0722
<i>G. Intestinalis</i>	-0.0898	-0.0226	-0.0415	0.0810
<i>Reovirus</i>	-0.0189	0.0723	-0.1084	-0.0687
<i>Rotavirus</i>	0.0550	0.0095	-0.0338	0.0101
SRSV	-0.2316	0.0015	0.1427	-0.2239
<i>S. Sonnei</i>	0.0586	0.0538	-0.0293	0.0087
<i>Yersinia</i>	0.0225	-0.0585	0.0398	-0.0625

Table 5.4: Pearson correlation coefficients (1)

Symptom	Age at entry	Birth Order	Number of Parents Employed	Number of Parents the child lived with
Abdominal Pain	-0.0573	-0.0399	0.1532	-0.1034
Diarrhoea or Loose Stool	-0.1864	-0.0657	0.1292	-0.0185
Temperature	-0.0790	0.1022	0.1874	0.0876
Vomiting	-0.1087	-0.0481	0.1682	-0.0083
Gastro-enteritis	-0.2019	-0.0345	0.2350	-0.0046

Table 5.5: Pearson correlation coefficients (2)

It can be seen from table 5.5 that age at entry in the study had a negative correlation with all the symptoms. This can be interpreted as meaning that the older the child in the study was, the less frequently he/she would induce symptoms. Most symptoms also have negative correlations with the child's birth order as well. This can be referred to the fact that the younger the child among his/her siblings, the less frequently he/she will show symptoms of illness because his/her older siblings had already got infected at an earlier time. Hence the child would be less likely to get infected from home. The same applies to birth order.

The correlations between the number of parents the child lived with and the rate of occurrence of symptoms were negative for most of the symptoms. This means that children who lived with both parents were less likely to get ill than single parented children. But children with both parents employed were more likely to get ill as can be seen from the positive correlations between the number of individual symptoms and the number of parents employed.

Table 5.7 shows that the correlation between the number of fortnights for which each child was enrolled in the study and the number of

Agent	Number of Fortnights the child was enrolled for	Family Size	Sex	Number of Siblings
<i>Adenovirus</i>	0.3229	-0.1159	-0.2219	-0.0963
<i>Adenovirus</i> (gastrointestinal)	0.0243	-0.0075	-0.1863	0.0006
All the agents together	0.4445	-0.0537	-0.1377	-0.0370
<i>Astrovirus</i>	-0.0160	-0.0092	-0.0976	0.0117
<i>Calicivirus</i>	0.1815	-0.0444	-0.0424	-0.0289
<i>Campylobacter</i>	0.2543	-0.0410	-0.2287	-0.0225
<i>E. Coli</i>	0.2543	0.0575	0.0084	0.0767
<i>Coronavirus</i>	-0.0097	-0.0167	-0.1047	-0.0136
<i>Cryptosporidium</i>	0.1788	-0.0931	-0.1594	-0.1048
<i>Coxsackie A virus</i>	0.0868	-0.0529	-0.0323	-0.0649
<i>Coxsackie B virus</i>	0.3973	0.0041	-0.1268	-0.0566
<i>Echovirus</i>	0.2880	-0.0375	0.0001	-0.0079
<i>Enterovirus</i>	0.1238	0.0780	0.0104	0.0587
<i>Giardia intestinalis</i>	0.1414	0.0295	0.1072	-0.0011
<i>Reovirus</i>	0.1644	0.0468	0.0685	0.0838
<i>Rotavirus</i>	0.1206	0.0233	0.0290	0.0225
SRSV	0.1843	-0.0876	-0.0816	-0.0041
<i>Shigella Sonnei</i>	0.0521	0.0566	0.0824	0.0616
<i>Yersinia</i>	-0.0172	-0.0373	0.0581	-0.0160

Table 5.6: Pearson correlation coefficients (3)

Symptom	Number of Fortnights the child was enrolled for	Family Size	Sex	Number of Siblings
Abdominal Pain	0.1084	-0.0798	-0.2105	-0.0473
Diarrhoea or Loose Stool	0.2032	-0.0545	-0.1135	-0.0548
Temperature	0.1533	0.0614	0.0664	0.0329
Vomiting	0.2462	-0.0411	-0.1288	-0.0439
Gastro-enteritis	0.2929	-0.0451	-0.1366	-0.0500

Table 5.7: Pearson correlation coefficients (4)

occurrences of symptoms was positive. This is due to the fact the children who were enrolled for a longer time were observed for longer and hence were seen to be ill more often than children with fewer fortnights of enrolment.

Sex was coded here as 1 for females and 2 for males. It can also be seen from table 5.7 that females were at higher risk of getting abdominal pain, diarrhoea or loose stool, or vomiting, while males were at higher risk of getting a high temperature than females.

5.4 Logistic Regression

5.4.1 Microbiological Factors

The first thing to be done was to find a statistical relation between the presence of the infectious agents in the children's stool specimens and the occurrences of symptoms. Each variable in the SPSS data file assumes one of two values, 1 if the agent/symptom was present and 0 otherwise. The most proper statistical test here is to fit logistic regression models. Symptoms were taken to be the dependent variables

and infectious agents were the covariates. SPSS syntax was written to fit four different logistic regression models which were fitted taking each of abdominal pain, diarrhoea or loose stool, temperature and vomiting as the independent variables and all of the agents as the covariates for the whole data set. This was repeated for each school and each gender. The logistic regression equation used was

$$\pi_j = \frac{e^{\beta_0 + \sum \beta_i x_{ij}}}{1 + e^{\beta_0 + \sum \beta_i x_{ij}}} \quad (5.1)$$

where x_{ij} denotes presence or absence of the i th infectious agent in the j th child fortnight; and

π_j is the conditional probability that the child induced a particular symptom in the j th child fortnight given the infectious agents that were present during that fortnight. The result of fitting Abdominal Pain to the data in school 1 is shown in figure 5.1.

Total number of cases: 4342 (Unweighted)
 Number of selected cases: 2340
 Number of unselected cases: 2002

Number of selected cases: 2340
 Number rejected because of missing data: 0
 Number of cases included in the analysis: 2340

The variable YERSINIA is constant for all selected cases.
 Since a constant was requested in the model,
 it will be removed from the analysis.

Dependent Variable Encoding:

Original	Internal		
Value	Value	Dependent Variable..	ABDOMINAL PAIN
0	0		
1	1		

Beginning Block Number 0. Initial Log Likelihood Function

-2 Log Likelihood 32.257326

* Constant is included in the model.

Beginning Block Number 1. Method: Enter

Variable(s) Entered on Step Number

1.. *Adenovirus*
Astrovirus
Calicivirus
Campylobacter
E. coli
Coronavirus
Cryptosporidium
Coxsackie A virus
Coxsackie B virus
Cysts
Echovirus
Enterovirus
Reovirus
Rotavirus
Spirochetes
 SRSV
S. sonnei
Giardai intestinalis

Estimation terminated at iteration number 12 because
 Log Likelihood decreased by less than .01 percent.

Figure 5.1: Logistic regression output.

-2 Log Likelihood 26.047
 Goodness of Fit 2107.001
 Chi-Square df Significance
 Model Chi-Square 6.210 18 0.9953
 Improvement 6.210 18 0.9953
 Classification Table for ABDOMINAL PAIN
 Selected cases SCHOOL EQ 1

		Predicted		Percent Correct
		0	1	
Observed	0	2338	0	100.00%
	1	2	0	0.00%
Overall				99.91%

Classification Table for ABDOMINAL PAIN
 Unselected cases SCHOOL NE 1

		Predicted		Percent Correct
		0	1	
Observed	0	1994	0	100.00%
	1	8	0	0.00%
Overall				99.60%

Variables in the Equation							
Variable	B	S.E.	Wald	df	Sig	R	Exp(B)
<i>Adenovirus</i>	-6.0834	65.5046	0.0086	1	0.9260	0.0000	0.0023
<i>Astrovirus</i>	-8.3115	240.4374	0.0012	1	0.9724	0.0000	0.0002
<i>Calicivirus</i>	-5.5647	425.0235	0.0002	1	0.9896	0.0000	0.0038
<i>Campylobacter</i>	-5.0723	257.3762	0.0004	1	0.9843	0.0000	0.0063
<i>E. coli</i>	-6.6976	82.9084	0.0065	1	0.9356	0.0000	0.0012
<i>Coronavirus</i>	-5.2182	233.9837	0.0005	1	0.9822	0.0000	0.0054
<i>Cryptosporidium</i>	-7.0731	126.0397	0.0031	1	0.9552	0.0000	0.0008
<i>Coxsackie A virus</i>	-4.5923	452.7063	0.0001	1	0.9919	0.0000	0.0101
<i>Coxsackie B virus</i>	-5.9703	88.5939	0.0045	1	0.9463	0.0000	0.0026
<i>Cysts</i>	7.1033	746.0932	0.0001	1	0.9924	0.0000	1215.954
<i>Echovirus</i>	4.2709	1.4264	8.9654	1	0.0028	.4647	71.5861
<i>Enterovirus</i>	-4.2086	373.2272	0.0001	1	0.9910	0.0000	0.0149
<i>Reovirus</i>	-7.0966	180.9387	0.0015	1	0.9687	0.0000	0.0008
<i>Rotavirus</i>	-5.0463	401.5728	0.0002	1	0.9900	.0000	0.0064
<i>Spirochetes</i>	-4.6878	474.3242	0.0001	1	0.9921	0.0000	0.0092
SRSV	-8.1805	317.3817	0.0007	1	0.9794	0.0000	0.0003
<i>S. sonnei</i>	2.3852	490.7844	0.0000	1	0.9961	0.0000	10.8618
<i>Giardia intestinalis</i>	-4.8500	257.2632	0.0004	1	0.9850	0.0000	0.0078
Constant	-7.6382	1.0002	58.3199	1	0.0000		

Figure 5.1 (continued): Logistic regression output.

The first part of figure 5.1 is self explanatory. It shows the total number of cases. The number of selected cases in this case is the number of children in school 1 multiplied by the total number of fortnights (26) and the number of unselected cases is the number of children in school 2 multiplied by 26. Since none of the rows contains missing values, the number of selected cases equals the number of cases included in the analysis. The output also gives the initial log likelihood function and a list of all the variables included in the model. The analysis iterates until the difference between the old and the new log likelihood reaches a threshold. In this case the threshold equals 0.01.

The final part of the output is the most important. It contains a table of all the variables in the equation along with their coefficients (column labelled B), their standard error (S.E.), Wald statistic, degrees of freedom for each variable, p value (sig.), correlation coefficients with the dependent variable and the expected value of the coefficient.

Taking the p value to be 0.05 in the results of fitting logistic regression models to the whole data, the logistic regression coefficients which are statistically significant are listed in table 5.8. The other covariates which have p value greater than 0.05 are discarded.

It can be seen from this table that the higher the coefficient of the agent, the more likely that the child will get the symptom if the infectious agent was present in his stool. For instance, the child will definitely induce diarrhoea if his stool specimen tested positive for *S. sonnei* but he may not produce any diarrhoeal symptoms if *Adenovirus* was present in his stool, as the latter might be a respiratory tract infection. This is in accordance with the microbiological literature.

After fitting logistic regression models to the data, the next step here was to calculate the expected number of cases of symptoms due

Symptom	Agent	Females	Males	School 1	School 2	All
Gastro-enteritis	<i>Adenovirus</i>	1.2903	0.6441	0.8801	1.4905	0.9731
	<i>Astrovirus</i>	1.3954	3.4948	2.5599	-3.3526	2.2706
	<i>Campylobacter</i>	2.7664	2.1379	1.5502	3.2158	2.7681
	<i>E. coli</i>	0.7511	0.7550	1.3579	-0.1374	0.8153
	<i>Cryptosporidium</i>	2.5294	5.2112	3.8212	3.8878	3.5211
	<i>Corsackie B</i>	1.3318	2.6111	2.5681	1.3558	1.9959
	<i>Echovirus</i>	1.8660	2.6081	1.8251	2.5180	2.2437
	<i>Rotavirus</i>	1.9816	3.3743	4.7607	-3.3826	3.8784
	<i>S.sonnei</i>	0.0	4.3547	8.4570	3.8191	4.1520
Diarrhoea Or Loose Stool	<i>Adenovirus</i>	1.2597	.7452	.8488	1.4592	0.9152
	<i>Astrovirus</i>	1.7452	3.6360	2.9547	-3.1940	2.4989
	<i>Campylobacter</i>	2.4920	2.3933	2.0515	2.8877	2.5943
	<i>Coronavirus</i>	1.3676	0.8372	3.0272	-0.4883	1.3979
	<i>Cryptosporidium</i>	2.4418	5.3623	3.7945	3.9629	3.5454
	<i>Corsackie B</i>	1.4237	2.3115	2.3861	1.3298	1.836
	<i>Echovirus</i>	1.6188	2.4984	1.3328	2.5212	2.0811
	<i>Rotavirus</i>	12.1678	3.5324	5.0178	-3.1940	4.0915
	<i>S. Sonnei</i>	-	4.4898	8.6094	4.0077	4.3086
Abdominal Pain	<i>G. Intestinalis</i>	1.8027	1.1423	1.9320	0.5239	1.2339
	<i>Adenovirus</i>	1.8759	-.5998	-6.0834	2.1595	1.8573
	<i>Campylobacter</i>	4.2148	6.7186	-5.0723	5.4797	5.2223
	<i>Echovirus</i>	2.4814	-6.1263	4.2709	1.9074	2.4981
	SRSV	3.7703	0.3179	-8.1805	3.6779	3.3619
	<i>G. Intestinalis</i>	4.0571	-5.6957	-4.8500	1.3949	1.6208
Temperature	<i>Campylobacter</i>	-6.0471	3.4243	-7.0317	3.2213	2.2483
	<i>Corsackie B</i>	-6.3523	2.1911	2.5511	1.0157	1.7380
	<i>S. Sonnei</i>	-	5.4612	1.1386	5.3430	5.4890
Vomiting	<i>Adenovirus</i>	1.7870	-5.6396	1.4678	1.9726	1.4009
	<i>Campylobacter</i>	2.3274	3.3221	-5.8945	3.5757	2.7595
	<i>E. coli</i>	1.8523	1.4255	2.3190	-0.2556	1.5518
	<i>Cryptosporidium</i>	3.0333	2.4166	2.4852	4.9941	2.7833
	<i>Corsackie B</i>	0.1223	2.3967	2.5471	0.5090	1.5657
	<i>Echovirus</i>	1.6820	1.6590	2.2592	1.6294	1.7344

Table 5.8: Logistic regression Coefficients

to all the agents together and to each agent individually. This was done by substituting for the coefficients in equation 5.1 to calculate the respective probabilities π_j . In this calculation, only the covariates from table 5.8 were used. The others were discarded since they were non-significant. SPSS code was written to calculate these probabilities. The expected number of cases of symptoms due to all the significant agents was calculated as the sum of the conditional probabilities. In order to calculate the expected number of cases due to individual agents, the coefficient of such agent was set to zero and the conditional probabilities were calculated. Then the sum of the new $\pi(x_i)$'s with the effect of agent i removed was subtracted from the original sum of $\pi(x_i)$'s. These values are listed in table 5.9.

In this table the expected number of cases of abdominal pain cases is equal to 12.56. This is calculated as follows:

$$\sum_{j=1}^{4342} \pi_j = \sum_{j=1}^{4342} \frac{1}{1 + e^{g(agent)}} \quad (5.2)$$

where

$$g(agent) = -(1.8578 \text{ Adenovirus} + 5.2223 \text{ Campylobacter} + 2.4891 \text{ Echovirus} \\ + 3.3619 \text{ SRSV} + 1.6208 \text{ Giardia intestinalis} - 6.8317).$$

Then, to calculate the expected number of cases of abdominal pain due to *Adenovirus* for example the following equation was evaluated:

$$\sum_{j=1}^{4342} \pi(x_j) = \sum_{j=1}^{4342} \frac{1}{1 + e^{g_1(agent)}} \quad (5.3)$$

where

$$g_1(agent) = -(5.2223 \text{ Campylobacter} + 2.4891 \text{ Echovirus} \\ + 3.3619 \text{ SRSV} + 1.6208 \text{ Giardia intestinalis} - 6.8317).$$

and the resulting sum was subtracted from 12.56. The difference is the expected number of abdominal pain cases due to *Adenovirus*. The rest of the figures, for the whole set of data, in table 5.9 were calculated in the same manner. The expected number of cases for each gender or school were calculated as the partial sums of the whole data.

It can be noticed from table 5.9 that the expected number of cases of diarrhoea or loose stool due to all the agents simultaneously was 116.71 where the observed number of cases was 117. The expected number of cases of diarrhoea or loose stool due to *Adenovirus* alone was 6.66. It can also be seen that the expected number of cases of vomiting was 17.31 due to all agents, of which 3.17 cases were due to *Cryptosporidium* only.

5.4.2 Social Factors

Logistic regression models were also used to determine the association between social and economical factors and occurrence of symptoms and infection with the different agents in children. These factors were child's age at acquiring the infection, child's birth order, the number of parents the child lived with, the number of parents that were employed during the time of the study, the child's family size during the period of the experiment, child's sex and the number of siblings the child had during the study period. The last factor was made redundant by the SPSS processor because it was a linear function of the child's number of parents and family size. Child's sex was coded here as 1 for females and 2 for males.

In order to determine the relation between the above factors and the infection with one or more different agents, all the agents were added together to form a new variable called "agents". Then two logistic

Symptom	Agent	Females	Males	School 1	School 2	All
Gastro-enteritis	All agents	848.01	1475.78	1248.41	1075.38	2323.79
	<i>Adenovirus</i>	17.44	17.53	18.89	16.08	34.79
	<i>Astrovirus</i>	1.39	1.27	1.85	0.81	2.66
	<i>Campylobacter</i>	4.22	1.87	2.36	3.73	6.09
	<i>E. coli</i>	4.38	8.67	8.17	4.88	13.05
	<i>Cryptosporidium</i>	4.46	3.35	7.04	0.77	7.81
	<i>Coxsackie B</i>	15.32	19.93	14.85	20.40	35.25
	<i>Echovirus</i>	13.57	24.72	14.51	23.78	38.29
	<i>Rotavirus</i>	0.48	1.26	1.26	0.48	1.74
	<i>S.sonnei</i>	0.0	1.01	0.04	0.97	1.01
Diarrhoea Or Loose Stool	All agents	48.56	68.15	65.29	51.42	116.71
	<i>Adenovirus</i>	3.96	2.70	3.67	2.99	6.66
	<i>Astrovirus</i>	1.37	1.16	2.26	0.27	2.53
	<i>Campylobacter</i>	3.08	1.06	1.87	2.27	4.14
	<i>Coronavirus</i>	1.47	0.92	0.84	1.55	2.39
	<i>Cryptosporidium</i>	8.05	4.46	11.74	0.77	12.51
	<i>Coxsackie B</i>	6.52	6.11	7.28	5.35	12.63
	<i>Echovirus</i>	5.17	7.79	5.97	6.99	12.96
	<i>Rotavirus</i>	0.45	1.48	1.48	0.45	1.93
	<i>S. Sonnei</i>	0	2.08	1.08	1.00	2.08
Abdominal Pain	<i>G. Intestinalis</i>	0.43	2.76	0.44	2.75	3.19
	All agents	6.02	6.54	5.78	6.78	12.56
	<i>Adenovirus</i>	1.66	1.02	1.23	1.45	2.68
	<i>Campylobacter</i>	2.76	0.84	1.54	2.06	3.60
	<i>Echovirus</i>	0.61	1.38	0.84	1.15	1.99
	SRSV	0.70	1.06	0.75	1.01	1.76
Temperature	<i>Giardia intestinalis</i>	0.35	1.19	0.36	1.18	1.54
	All agents	9.18	16.81	13.51	12.48	25.99
	<i>Campylobacter</i>	0.78	0.31	0.59	0.50	1.09
	<i>Coxsackie B virus</i>	1.35	1.40	1.43	1.32	2.75
Vomiting	<i>S. sonnei</i>	0	2.15	1.16	0.99	2.15
	All agents	17.31	21.49	22.72	16.08	38.80
	<i>Adenovirus</i>	3.06	1.80	3.31	1.55	4.86
	<i>Campylobacter</i>	1.76	0.51	1.09	1.18	2.27
	<i>E. coli</i>	1.68	2.23	2.35	1.56	3.91
	<i>Cryptosporidium</i>	3.17	1.49	4.33	0.32	4.65
	<i>Coxsackie B</i>	2.74	1.88	3.21	1.41	4.62
	<i>Echovirus</i>	1.39	2.14	1.76	1.77	3.53

Table 5.9: Expected numbers of episodes of symptoms attributable to various agents

regression models were fitted using the newly defined variable as the dependent variable and the above mentioned factors as covariates. The first of these contained an interaction term between the child's sex and age at infection. At the 5% level of significance, it was found that both age and sex were statistically significant but the interaction term was not. Also, the number of parents the child lived with, and the number of parents employed were significant at the 5% level. The second logistic regression model was fitted using the same dependent variable and covariates but without the interaction term. It was found that both birth order and family size were not significant at the 5% level and all the other covariates were. The correlation coefficients between agents and number of parents the child lived with, child's sex and age at infection were all negative. This means that female children were more susceptible to infection than males and the older the child was the less infections he/she would get. It also indicates the children who lived with both parents were at lower risk of being infected.

All the gastro-enteritis symptoms were added together to form another new variable called "gastro". Then two logistic regression models were fitted using the same covariates mentioned above. The first model, also, contained an interaction term between age at appearance of symptoms and child's sex and the second model was without the interaction term. The results of fitting these models showed that only the number of parents employed was statistically significant at the 5% level in both models, and in the second model, child's age at appearance of symptoms as well as the number of parents employed were both significant at the 1% level. In the first model, the correlation coefficient between the number of episodes of gastro-enteritis and the number of parents employed was positive while the other coefficients were all zero. However,

Age	Females	Males	Total
Less than one year	3	7	10
One year	23	38	51
Two years	20	36	56
Three years	10	28	38
Four years	4	8	12

Table 5.10: Distribution of age by sex

in the second model, the correlation coefficient between "gastro" and child's sex and age at induction of symptoms were both negative but the correlation with the number of parents employed was positive.

In conclusion, child's age, sex and number of parents employed appeared to be significantly related to the amount of illness the child endured where the other social factors such as family size, child's birth order and number of parents the child lived with did not show any significance.

5.5 Proportions of asymptomatic carriage

This section will discuss the proportions of asymptomatic carriage of each agent of infection. Asymptomatic carriage occurs when the agent is isolated from the child's stool specimen but no gastro-intestinal symptoms are reported. These symptoms were taken to be diarrhoea or loose stool, abdominal pain, vomiting or high temperature. Also here, children's ages were divided into five categories; under one year old, one, two, three and four years old. The children with missing ages were considered as under one year old here. The distribution of children's age by sex is displayed in table 5.10.

Agent/Age	Less than 1	1	2	3	4
<i>Adenovirus</i>	8	50	22	7	1
<i>Adenovirus</i> (respiratory)	8	42	19	5	1
<i>Adenovirus</i> (gastro-intestinal)	0	8	3	2	0
<i>Astrovirus</i>	1	3	1	0	0
<i>Caliciviruses</i>	0	0	3	0	1
<i>Campylobacter</i>	1	4	4	2	1
<i>E. coli</i>	3	14	11	2	1
<i>Coronavirus</i>	0	2	13	0	1
<i>Cryptosporidium</i>	3	6	6	1	0
<i>Coxsackie A virus</i>	0	1	0	0	0
<i>Coxsackie B virus</i>	6	23	17	2	1
<i>Echovirus</i>	5	22	9	0	2
<i>Enterovirus</i>	0	2	0	0	0
<i>Giardia intestinalis</i>	0	3	2	0	0
<i>Reovirus</i>	0	2	1	0	0
<i>Rotavirus</i>	0	0	0	0	1
SRSV	1	5	0	0	0
<i>S. sonnei</i>	0	0	0	0	0
<i>Yersinia</i>	0	0	0	0	0

Table 5.11: Distribution of agents by age for females

The total number of times each agent was isolated for each age group and each sex was calculated separately. These totals are displayed in table 5.11 for females and 5.12 for males. Also the total number of times each agent was present and no symptoms were reported were calculated for each age group and sex. Tables 5.13 and 5.14 show these totals for females and males respectively. The proportions of

Agent/Age	Less than 1	1	2	3	4
<i>Adenovirus</i>	5	29	26	19	3
<i>Adenovirus</i> (respiratory)	4	26	24	18	3
<i>Adenovirus</i> (gastro-intestinal)	1	3	2	1	0
<i>Astrovirus</i>	0	3	1	0	0
<i>Caliciviruses</i>	0	3	2	0	0
<i>Campylobacter</i>	0	1	1	3	0
<i>E. coli</i>	4	18	21	11	3
<i>Coronavirus</i>	0	3	5	2	0
<i>Cryptosporidium</i>	1	4	4	2	0
<i>Coxsackie A virus</i>	0	0	1	0	0
<i>Coxsackie B virus</i>	8	21	10	16	3
<i>Echovirus</i>	10	10	20	12	1
<i>Enterovirus</i>	0	1	1	2	0
<i>Giardia intestinalis</i>	10	2	14	6	1
<i>Reovirus</i>	0	6	2	1	3
<i>Rotavirus</i>	0	2	0	1	0
SRSV	1	3	2	0	0
<i>S. sonnei</i>	0	0	2	2	0
<i>Yersinia</i>	0	0	1	0	0

Table 5.12: Distribution of agents by age for males

asymptomatic carriage for each agent were then calculated by dividing the numbers from the second set of tables by corresponding cells in the first set of tables. These proportions are listed in tables 5.15 for females and 5.16 for males. The "-" in these tables means dividing by zero, i.e. the particular agent was not isolated at all from specimens in this age group where the "0.0" means that it was isolated and symptoms were

Agent/Age	Less than 1	1	2	3	4
<i>Adenovirus</i>	1	7	3	1	0
<i>Adenovirus</i> (respiratory)	1	5	2	1	0
<i>Adenovirus</i> (gastro-intestinal)	0	2	1	0	0
<i>Astrovirus</i>	1	0	0	0	0
<i>Caliciviruses</i>	0	0	0	0	0
<i>Campylobacter</i>	0	2	1	1	1
<i>E. coli</i>	1	0	2	1	0
<i>Coronavirus</i>	0	1	2	0	0
<i>Cryptosporidium</i>	3	2	2	0	0
<i>Coxsackie A virus</i>	0	0	0	0	0
<i>Coxsackie B virus</i>	3	3	3	0	0
<i>Echovirus</i>	2	5	0	0	0
<i>Enterovirus</i>	0	0	0	0	0
<i>Giardia intestinalis</i>	0	1	0	0	0
<i>Reovirus</i>	0	0	0	0	0
<i>Rotavirus</i>	0	0	0	0	1
SRSV	0	0	0	0	0
<i>S. sonnei</i>	0	0	0	0	0
<i>Yersinia</i>	0	0	0	0	0

Table 5.13: Distribution of agents by age for females when at least one symptom was present

Agent/Age	Less than 1	1	2	3	4
<i>Adenovirus</i>	1	0	2	1	0
<i>Adenovirus</i> (respiratory)	1	0	0	1	0
<i>Adenovirus</i> (gastro-intestinal)	0	0	2	0	0
<i>Astrovirus</i>	0	2	0	0	0
<i>Caliciviruses</i>	0	1	0	0	0
<i>Campylobacter</i>	0	0	0	1	0
<i>E. coli</i>	0	4	1	2	0
<i>Coronavirus</i>	0	1	0	0	0
<i>Cryptosporidium</i>	1	2	4	2	0
<i>Coxsackie A virus</i>	0	0	0	0	0
<i>Coxsackie B virus</i>	3	7	3	0	0
<i>Echovirus</i>	3	1	3	3	0
<i>Enterovirus</i>	0	1	0	0	0
<i>Giardia intestinalis</i>	1	2	1	0	0
<i>Reovirus</i>	0	1	0	0	0
<i>Rotavirus</i>	0	1	0	0	0
SRSV	0	2	0	0	0
<i>S. sonnei</i>	0	0	1	2	0
<i>Yersinia</i>	0	0	0	0	0

Table 5.14: Distribution of agents by age for males when at least one symptom was present

reported. It can be seen from these two tables that most of the isolates were in the age groups 1 and 2. This is because the majority of the study population was in that age, 64% of the children. Another thing that can be noticed from these table is that in only a few cases these proportions were less than 50%. This means that the frequency of illness is very low in the study population compared to other populations in other studies as can be seen in section C.1.

In the next chapter, we will discuss the theoretical estimation of the basic reproductive rate R_0 and the survival models.

Agent/Age	Less than 1	1	2	3	4
<i>Adenovirus</i>	87.5	86.0	86.4	85.7	100.0
<i>Adenovirus</i> (respiratory)	80.0	87.5	100.0	100.0	-
<i>Adenovirus</i> (gastro-intestinal)	-	75.0	66.7	100.0	-
<i>Astrovirus</i>	0.0	100.0	100.0	-	-
<i>Caliciviruses</i>	-	-	100.0	-	100.0
<i>Campylobacter</i>	100.0	50.0	75.0	50.0	0.0
<i>E. coli</i>	66.7	100.0	81.8	50.0	100.0
<i>Coronavirus</i>	-	50.0	84.6	-	100.0
<i>Cryptosporidium</i>	0.0	66.7	66.7	100.0	-
<i>Coxsackie A virus</i>	-	100.0	-	-	-
<i>Coxsackie B virus</i>	50.0	87.0	82.4	100.0	100.0
<i>Echovirus</i>	60.0	77.3	100.0	-	100.0
<i>Enterovirus</i>	-	100.0	-	-	-
<i>Giardia intestinalis</i>	-	66.7	100.0	-	-
<i>Reovirus</i>	-	100.0	100.0	-	-
<i>Rotavirus</i>	-	-	-	-	100.0
SRSV	100.0	100.0	-	-	-
<i>S. sonnei</i>	-	-	-	-	-
<i>Yersinia</i>	-	-	-	-	-

Table 5.15: Proportion of asymptomatic carriage by age for females

Agent/Age	Less than 1	1	2	3	4
<i>Adenovirus</i>	80.0	100.0	92.3	94.7	66.7
<i>Adenovirus</i> (respiratory)	75.0	100.0	91.7	94.4	66.7
<i>Adenovirus</i> (gastro-intestinal)	100.0	100.0	100.0	100.0	-
<i>Astrovirus</i>	-	33.3	100.0	-	-
<i>Caliciviruses</i>	-	-	66.7	100.0	-
<i>Campylobacter</i>	-	100.0	100.0	66.7	-
<i>E. Coli</i>	100.0	77.8	95.2	81.8	100.0
<i>Coronavirus</i>	-	66.7	100.0	100.0	-
<i>Cryptosporidium</i>	0.0	50.0	0.0	0.0	-
<i>Coxsackie A virus</i>	-	-	100.0	-	-
<i>Coxsackie B virus</i>	62.5	66.7	70.0	100.0	100.0
<i>Echovirus</i>	70.0	77.8	85.0	81.3	100.0
<i>Enterovirus</i>	-	0.0	100.0	100.0	-
<i>Giardia intestinalis</i>	90.0	0.0	92.9	100.0	100.0
<i>Reovirus</i>	-	83.3	100.0	100.0	100.0
<i>Rotavirus</i>	-	50.0	-	100.0	-
SRSV	100.0	33.3	100.0	-	-
<i>S. sonnei</i>	-	-	50.0	100.0	-
<i>Yersinia</i>	-	-	100.0	-	-

Table 5.16: Proportion of asymptomatic carriage by age for males

Chapter 6

Bias Correction

6.1 Introduction

Besides the exploratory /simple analysis previously discussed, this chapter attempts to quantify the force of infection, best summarised as the basic reproductive rate R_0 .

This is done to address the problem of making such inference from data on common infections that might feasibly be collected, rather than to compute accurate results. Clearly many modelling assumptions must be made and the value of R_0 produced will be somewhat questionable. However, no previous estimates exist and it seems worthwhile to calculate R_0 as best as can be done, noting the inferential problems encountered along the way.

The strategy adopted is to fit a survival model to the first observation of an episode of infection for each serotype of an agent. This is possible by assuming that infection with a given serotype occurs only once for each individual, and yields the force of infection for each serotype. The basic reproductive rate R_0 is then calculated from the equilibrium condition that $R_0x = 1$, where x is the proportion of unin-

fectured individuals in the total population.

In fitting the survival model, the estimated probability p that an episode of infection by an agent will be detected if it occurs can readily be included. This will correct the estimated value of R_0 for a large measurement bias.

This chapter addresses the estimation of the probability p , and the following chapter proceeds to the estimation of R_0 .

6.2 Correction for cases missed due to low sensitivity

It is straightforward to calculate s , the sensitivity of the test, and to calculate r , where r is the probability that a specimen is not missing. If missing samples occur randomly, the ‘true sensitivity’ or probability of detecting a positive sample if the agent is present is $q = rs$. An obvious estimate of r is

$$\begin{aligned} & \frac{\text{total number of specimens collected}}{\text{total number of fortnights enrolled}} \\ &= \frac{1175}{2465} = 0.477. \end{aligned}$$

To calculate s , for each agent or serotype of a serotyped agent, the total number of fortnights between the first and last isolation of the agent were counted for each child. Then, the test sensitivity for the particular serotype/agent was then taken to be the proportion of times the agent was positively isolated to the total number of specimens collected between the first and last isolation *exclusive*. To illustrate this consider *Adenovirus* serotype 1. The occurrence of this agent was as follows:

1 1 0 - 0 1 1
1 1 0 1

where the 1's represent positive isolations, the 0's represent negative isolations and the -'s means no specimens were collected from the particular child . The test sensitivity for the particular agent for each child is $\frac{2}{4}$ and $\frac{1}{2}$ respectively. The total test sensitivity for *Adenovirus* serotype 1 was calculated to be

$$\frac{2+1}{4+2} = \frac{3}{6}.$$

For the rest of the agents the sensitivity was calculated in the same manner and was found to be as follows:

<i>Adenovirus</i> 1	$\frac{3}{6}$
<i>Adenovirus</i> 2	$\frac{4}{15}$
<i>Adenovirus</i> 5	$\frac{4}{14}$
<i>Corsackie virus B</i> 4	$\frac{1}{2}$
<i>E. Coli</i> O111	$\frac{2}{4}$
<i>E. Coli</i> O126	$\frac{2}{5}$
<i>Echovirus</i> 11	$\frac{3}{10}$
<i>Coronavirus</i>	$\frac{5}{8}$
<i>Giardia Intestinalis</i>	$\frac{16}{18}$

From this, the overall test sensitivity is

$$\frac{\text{the sum of numerators}}{\text{the sum of denominators}} = \frac{40}{82} = 0.4878.$$

For each agent we can calculate $p = 1 - Q_0$, the probability of detecting an episode of infection by a serotype. The observed incidence of infection I should then be corrected to $I/(1 - Q_0)$.

This section describes the calculation of the correction factor. For simplicity we first neglect edge effects due to infections starting or ending outside our ‘window’ of observation. Finally the full calculation is given, which includes edge effects.

The resulting correction factor and its standard error were computed in a FORTRAN program. To do this the following were required:

1. the frequency distribution of periods of infection;
2. the frequency distribution of number of fortnights enrolled;
3. estimate of the sensitivity and probability of missing samples.

The first and second requirements are obtained from chapter 4 and the third was calculated above. The calculation of Q_0 and its standard error was evaluated.

6.3 Correction factor ignoring edge effects

Let the probability that an infection lasts for n fortnights be P_n , and the observed probability of an m fortnight infection be R_m . This is obtained as the number of fortnights from the first to the last positive sample, inclusive, or as 1 if there is only one positive sample. It is the distribution of observed infection periods *conditional on the infection being observed*, and so is related to the probability of observing an infection of period m , Q_m , by $R_m = Q_m/(1 - Q_0)$. Define also $S_m = P_m/(1 - Q_0)$.

Then

$$Q_m = \sum_{n=m}^{\infty} f_{mn} P_n,$$

where f_{mn} is the probability of observing m fortnights of infection if the true infection lasted n fortnights, and

$$R_m = \sum_{n=m}^{\infty} f_{mn} S_n.$$

We assume that successful detection of an infection occurs randomly with probability q with each fortnight constituting an independent Bernoulli trial. Then

$$f_{mn} = \begin{cases} (1-q)^n & \text{if } m = 0 \\ nq(1-q)^{n-1} & \text{if } m = 1 \\ (n-m+1)q^2(1-q)^{n-m} & \text{if } 2 \leq m \leq n \\ 0 & \text{otherwise.} \end{cases} \quad (6.1)$$

This follows because the $n-m$ leading and trailing samples that were not found positive (with probability $1-q$ each) can be added front or rear of the observed positive samples in $n-m+1$ ways, and the first and last positive samples will be detected with probability q^2 . What comes sandwiched in between does not matter and so is observed with probability 1. It can be verified that the probabilities sum to one as they must,

$$\sum_{m=0}^n f_{mn} = 1 \quad (6.2)$$

i.e. we are certain to see an infection of *some* number of weeks (including 0) not exceeding the true number.

Proof:

The proof of

$$(1-q)^n + nq(1-q)^{n-1} + q^2 \sum_{m=2}^n (n-m+1)(1-q)^{n-m} = 1.$$

is very simple and is given as follows:

The series is evaluated as the differential of a geometric series. Then,

the left hand side of 6.2 is equal to

$$(1-q)^n + nq(1-q)^{n-1} + q^2 \sum_{m=2}^n (n-m+1)(1-q)^{n-m}$$

Consider the series

$$\sum_{m=2}^n (1-q)^{n-m+1}$$

This is a geometric series which sum is equal to $\frac{(1-q)^n - (1-q)}{(1-q)-1}$.

By differentiating both sides, we obtain

$$\begin{aligned} & \sum_{m=2}^n (n-m+1)(1-q)^{n-m}(-1) \\ &= \frac{q(-1 - n(1-q)^{n-1}(-1)) - ((1-q) - (1-q)^n)(1)}{q^2} \\ &= \frac{q(n(1-q)^{n-1} - 1) + (1-q)^n - (1-q)}{q^2} \\ &= \frac{nq(1-q)^{n-1} - q + (1-q)^n - (1-q)}{q^2} \\ &= \frac{nq(1-q)^{n-1} - q + (1-q)^n - 1 + q}{q^2} \\ &= \frac{nq(1-q)^{n-1} + (1-q)^n - 1}{q^2} \end{aligned}$$

$$q^2 \sum_{m=2}^n (n-m+1)(1-q)^{n-m} = (-1)(nq(1-q)^{n-1} + (1-q)^n - 1).$$

Therefore

$$\begin{aligned} \sum_{m=0}^n f_{mn} &= (1-q)^n + nq(1-q)^{n-1} - (1-q)^n - nq(1-q)^{n-1} + 1 \\ &= 1. \end{aligned}$$

With this choice for f_{mn} the equations

$$R_m = \sum_{n=m}^{\infty} f_{mn} S_n \tag{6.3}$$

must be solved for the S_n , where $n > 0$, and then Q_0 found from

$$Q_0 = \sum_{n=1}^{\infty} f_{0n} P_n,$$

$$Q_0/(1 - Q_0) = \sum_{n=1}^{\infty} f_{0n} S_n.$$

It is also possible to find the standard errors of the estimate Q_0 . This can be done analytically, but is complicated, because both q (and hence f_{mn}) and R_m are random variables. Bootstrap resampling would probably be the easiest method computationally, and would yield confidence intervals on the correction factor.

Let the longest infection observed last for L fortnights. Then $Q_m = 0$ if $m > L$, and equation 6.3 is satisfied if $Q_n = 0$ for all $n > L$.

Equation 6.3 becomes

$$R_m = \sum_{n=m}^L f_{mn} S_n, \quad (6.4)$$

which comprises L linear equations, already in the form for solution by back-substitution. Thus we can solve at once for S_L , plug that in and solve for S_{L-1} , and so on down to S_1 . Finally the truncated equation

$$Q_0/(1 - Q_0) = \sum_{n=1}^L f_{0n} S_n \quad (6.5)$$

is used to estimate Q_0 by substituting in the S_n .

This calculation can be done manually when L is small. When $L = 1$, $Q_0 = 1 - q$ as it must (the single infection is seen with probability q).

The estimator P_n is also interesting. It is trivial to use it to calculate the mean and standard deviation of the infected period for each agent.

6.4 Correction factor including edge effects

The above analysis is not quite correct, because it assumes that the full infection period always lies within the ‘window’ of observation, which we now take to be from fortnight 1 to fortnight N .

We are more likely to fail to see an infection than the above analysis would suggest, because some of the infection period may lie outside our window. On the other hand, we may see the end of an infection that started before the observation window opened, so that we record more infections than started during the period $1 \dots N$.

Let an infection start at time j . An infection of period n to be seen must end after our window opens, so that the end $j + n - 1 \geq 1$ or $j \geq 2 - n$. Let g_{mnj} be the probability that an infection of period n , starting at time j , is recorded as having period m . Assume that the distribution of infection start times is uniform over a long interval, with probability of starting in any given fortnight of η . Let $\xi_{mnN} = \sum_{j=2-n}^N g_{mnj}$. The probability of observing an infection of length m is then

$$\eta \sum_{n=m}^{\infty} P_n \xi_{mnN}.$$

If the distribution of enrolment periods has probability h_N for N fortnights’ enrolment, this becomes (for a random child):

$$\eta \sum_{N=1}^{26} h_N \sum_{n=m}^{\infty} P_n \xi_{mnN}.$$

Normalising this to unity gives the distribution of infection lengths:

$$Q_m = \frac{\sum_{N=1}^{26} h_N \sum_{n=m}^{\infty} P_n \xi_{mnN}}{\bar{N} + \mu_P - 1}, \quad (6.6)$$

where μ_P is the mean of the P distribution, and $\bar{N}$ the average enrolment in fortnights. The denominator follows on reversing the order of the sums over n and m , and recalling that $\sum_{m=0}^n \xi_{mnN} = (N + n - 1)$.

The calculation of P from the observed Q is best done by fitting a parametric form for P by maximum likelihood.

The correction factor C is the true probability of an infection,

$$\eta \sum_{N=1}^{26} h_n N = \eta \bar{N},$$

divided by the observed probability, so that

$$C = \frac{\bar{N}}{\sum_{N=1}^{26} h_N \sum_{n=1}^{\infty} P_m \sum_{m=1}^n \xi_{mnN}},$$

where the sum over m in the denominator runs from 1. This can be calculated exactly once ξ has been calculated. It is convenient to write $\sum_{m=1}^n \xi_{mnN} = (N + n - 1) - \xi_{0nN}$.

We then have

$$C = \frac{\bar{N}}{(\bar{N} + \mu_P - 1)(1 - Q_0)}. \quad (6.7)$$

For large $\bar{N}$, $Q_0 = G(1 - q)$, where G is the probability generating function of distribution P .

It only remains to calculate ξ_{mnN} . To calculate the g_{mnj} four cases can be distinguished:

1. the infection starts before the observation window opens and ends before it closes ($j < 1, j + n - 1 \leq N$);
2. the infection starts before the observation window opens and ends after it closes ($j < 1, j + n - 1 > N$);
3. the infection starts after the observation window opens and ends before it closes ($j \geq 1, j + n - 1 \leq N$);
4. the infection starts after the observation window opens and ends after it closes ($j \geq 1, j + n - 1 > N$).

However, it is possible to combine these four into one algorithm. We define u as the time when the infection could be last seen within the window, so that $u = \min(n + j - 1, N)$. Correspondingly, define l as the time when the infection could be first seen, so that $l = \max(j, 1)$. Then the maximum period of observable infection is $n' = u - l + 1$. The formula for g_{mnj} is simply the old formula for f_{mn} , with n' replacing n , i.e.

$$g_{mnj} = \begin{cases} (1 - q)^{n'} & \text{if } m = 0 \\ n'q(1 - q)^{n'-1} & \text{if } m = 1 \\ (n' - m + 1)q^2(1 - q)^{n'-m} & \text{if } 2 \leq m \leq n' \\ 0 & \text{otherwise.} \end{cases}$$

Note that $\sum_{m=0}^n \xi_{mnN} = (N + n - 1)$. The quantities ξ_{mnN} can be found analytically, because the sums are either geometric series or differentials of them. The two cases are

1. $n \leq N$;
2. $n > N$.

For case 1, n' increases from 1 to $n - 1$, then remains at n for $N - n + 1$ fortnights, then decreases from $n - 1$ to 1.

For case 2, n' increases from 1 to $N - 1$, remains at N for $n - N + 1$ periods, then decreases from $N - 1$ to 1. These facts enable the sums to be evaluated.

Hence for $n \leq N$

$$\sum g_{0nj} = 2(1 - q - (1 - q)^n)/q + (N - n + 1)(1 - q)^n,$$

and for $n > N$

$$\sum g_{0nj} = 2(1 - q - (1 - q)^N)/q + (n - N + 1)(1 - q)^N,$$

where n and N have changed roles.

The differential of $\sum_{k=0}^n x^k = \frac{x^{n+1}-1}{x-1}$ is $\sum_{k=1}^n kx^{k-1}$. This enables the identity

$$\begin{aligned} \sum_{k=1}^n kx^k &= xd \sum_{k=0}^n x^k / dx, \\ &= x \left\{ \frac{(n+1)x^n}{x-1} - \frac{x^{n+1}-1}{(x-1)^2} \right\} \end{aligned}$$

to be derived.

Applying this, for $n \leq N$

$$\begin{aligned} \sum g_{1nj} &= \frac{2(1 - (1-q)^n)}{q} - 2n(1-q)^{n-1} \\ &\quad + (N - n + 1)nq(1-q)^{n-1}, \end{aligned}$$

and for $n > N$

$$\begin{aligned} \sum g_{1nj} &= \frac{2(1 - (1-q)^N)}{q} - 2N(1-q)^{N-1} \\ &\quad + (n - N + 1)Nq(1-q)^{N-1}. \end{aligned}$$

For $m \geq 2$,

$$\begin{aligned} \sum g_{mnj} &= 2(1 - (1-q)^{n-m+1}) - 2(n-m+1)q(1-q)^{n-m} \\ &\quad + (N - n + 1)(n-m+1)q^2(1-q)^{n-m}, \end{aligned}$$

and if $n > N$,

$$\begin{aligned} \sum g_{mnj} &= 2(1 - (1-q)^{N-m+1}) - 2(N-m+1)q(1-q)^{N-m} \\ &\quad + (n - N + 1)(N-m+1)q^2(1-q)^{N-m}, \end{aligned}$$

where again n and N have exchanged roles.

The computation of the correction factor proceeds as follows:

1. compute the matrix ξ_{mnN} for all m, n of interest;

2. solve equation 6.6 by back-substitution (or just by using a NAG linear equation solver) to get P_n ;
3. calculate the correction factor C using equation 6.7.

There is a problem with this method, in that the probabilities P_n may be estimated as negative or greater than unity (they are unphysical). This is unsatisfactory. An alternative is to choose the probabilities P_n to maximise the likelihood of observing the probabilities R_m . We may parameterise the distribution with one parameter per probability, or simply fit a distributional form such as the Poisson or Negative Binomial. This 2-parameter distribution includes the Poisson as a special case.

The likelihood function is readily written down. Conditioning on the total number N_c of episodes of infection, the likelihood of observing $n_1 \dots n_L$ infections of lengths 1 to L is

$$\mathcal{L} = \prod_{i=1}^L V_i^{n_i},$$

a multivariate probability where the initial factorial term has been omitted. The probabilities V_i are given as

$$V_i = \sum_{n=i}^{\infty} f_{in} P_n / (1 - \sum_{n=1}^{\infty} f_{0n} P_n),$$

where the P_n are the true probabilities of infection periods, parameterised as Negative Binomial. The log-likelihood is then (on dividing through by N_c)

$$\ell = \sum_{m=1}^L R_m \log \left(\sum_{n=m}^{\infty} f_{mn} P_n \right) - \log \left(1 - \sum_{n=1}^{\infty} f_{0n} P_n \right),$$

where the infinite sum can be replaced by a sum to 26 weeks.

6.5 The distribution of observed infection period

This distribution was derived earlier in 6.1, and has $Prob(X = m) = f_{mn}$. It is the distribution of observed periods or lengths m given random deletion of elements from a true length of n .

It seems to be new, as it not in Johnson and Kotz (1992). It is a 2-parameter distribution related to the Binomial. It may therefore be of some interest in its own right. The probability generating function $G(z) = \sum_{k=0}^n f_{kn} z^k$ can be calculated to be

$$G(z) = (1-q)^n \{1 + nq(1-q)\omega + q^2 \left\{ \frac{\omega(\omega^{n+1} - 1)}{(\omega - 1)^2} - (n+1)\omega/(\omega - 1) \right\}\},$$

where $\omega = z/(1-q)$. From this the mean and variance can be found.

There is also the related truncated distribution of observed probability of infection of period m , given that an infection is observed. The probabilities are simply $Prob(X = m) = f_{mn}/(1 - f_{0n})$.

The mean is

$$\mu = (n + 2 - 2/q)(1 - (1-q)^n) + 2n(1-q)^n,$$

and the mean of the truncated distribution (no $m = 0$ category) is

$$\mu' = n + 2 - 2/q + \frac{2n(1-q)^n}{1 - (1-q)^n}.$$

Note that μ' has the correct intuitive properties: as $q \rightarrow 1$, $\mu' \rightarrow n$, and as $q \rightarrow 0$, $\mu' \rightarrow 1$. This last statement is proved by making a Taylor expansion of $(1-q)^n$. When $n = 1$, $\mu' = 1$ for all q .

The variance can be found but is complicated.

As $n \rightarrow \infty$ the mean and variance increase without limit. It is not clear whether or not the standardised distribution of X then tends to normality or not.

The next chapter will discuss the numerical results of fitting the Weibull survival models to the data. It will also give the estimated basic reproductive rates for the four highly isolated agents of infection from the data.

Chapter 7

Survival Analysis and R_0 Calculation

This chapter covers the estimation of the basic reproductive rate R_0 for several agents. The basic reproductive rate R_0 may be calculated in many ways, some of these methods are as follows:

1. Direct measure: as stated earlier this is not feasible ethically because a person has to be infected and introduced to a group of people to observe the infection pattern. It is also not feasible because it is difficult to distinguish the second from the third generation of infection. This method may be feasible if the infection is mild and the study is intensive.
2. From use of a model. The model leads to an expression for R_0 as a function of model parameters that can be estimated. This method is widely used and appendix C describes some of these models.
3. From the equilibrium infection rate using

$$R_0 x = 1 \tag{7.1}$$

where x is the proportion of the general population that is not infected. This method is used here to estimate R_0 for four of the agents isolated in the study.

We calculate x by fitting a survival model to the data for infection with each serotype. This gives the force of infection (hazard function for infection), and from the resulting survival function, x can be calculated and hence R_0 .

The existence of serotypes makes this programme of inference possible, because although infection with a given serotype can occur only once for many agents, at least to a good approximation, infection with the agent can occur as many times as there are serotypes. Without knowledge of serotypes we could not then fit a survival model (which postulates an event that can occur only once).

However, the existence of serotypes is also a complication. Serotypes of an agent are thought in general to behave similarly, but not identically. *E. coli* 0157 is a recent example of a serotype that is much more pathogenic than other *E. coli* serotypes. Different serotypes will have different R_0 values, and so some will occur more frequently than others. However, the transmission of ‘waves’ of serotypes around the UK will result in a 1 or 2-year study seeing only some serotypes in abundance, whilst others are not seen at all. We would thus naively estimate a large R_0 for a few serotypes, and a tiny or zero R_0 for many others. In addition, we do not feel able to quote an R_0 for each serotype, but would rather quote an average R_0 for an agent.

A solution to these problems is an Empirical Bayes approach in which the force of infection due to each serotype is modelled as proportional to a random variable from (say) a gamma distribution with pdf $f(x|\mu, \sigma)$. This model allows some serotypes to be much more infectious

than others, as observed.

The resulting model is similar to the Littlewood model of software reliability, in which the software contains a number of ‘bugs’, whose associated hazards of being detected and causing a software failure are random variates from a gamma distribution, Musa (1987).

On fitting this model to the ‘survival’ data, we expect the estimated variance of the forces of infection $\hat{\sigma}^2$ to be much larger than the true variance between serotypes, because of the ‘lumpiness’ of serotype abundance, and because some serotype waves may happen to clip the catchment region of the day-nursery. The distribution of R_0 values will thus have a larger variance than the true distribution. However, the mean value of R_0 will be correct if the frequency of infection observed was typical, and will be an average R_0 value for the serotypes of that agent.

Many serotypes of an agent are not detected at all. If they are truly absent, we have no information about their infectivity, and we therefore calculate R_0 only from observed serotypes, which are taken as a representative random sample of all the serotypes for that agent.

What distinguishes survival analysis from other fields of statistics is censoring. Censoring occurs when enough information about the observation is not available. If the data is known to be censored, whether with type I or type II censoring, the easiest and straight forward methods of survival analysis are Kaplan-Meier or Cox regression models.

Since in the case of the data studied here, it is not known whether the subjects of the experiment encountered their first infection with each agent prior to the study period or whether what is recorded is their first infection. Another problem is when infection was not observed at all during our window of observation i.e. there are some children in the

study that did not show any positive isolation of any agent. It is not known whether these children were infected prior to the study period and hence gained immunity against these agents and therefore these agents were not observed in their stool specimens, or whether they acquired the infections after the study period was over. Hence such non parametric tests are not feasible and it is better to fit parametric survival models such as the Weibull distribution.

The mathematics of all this is now given:

Assume a Weibull survival function $S(t|y)$ for age t at infection with a particular serotype of hazard scaling factor y , i.e.

$$S(t) = \exp(-(Yt)^\beta), \quad (7.2)$$

where Y is a random variable obeying the gamma distribution with probability density function (pdf)

$$f(y) = \frac{\alpha(\alpha y)^{\gamma-1} \exp(-\alpha y)}{\Gamma(\gamma)}. \quad (7.3)$$

Given n serotypes with the following assumptions:

1. each serotype infects only once,
2. serotypes are independent, i.e. infections from different serotypes are independent,

then the full likelihood is

$$\mathcal{L}(\alpha, \beta, \gamma) = \prod_{i=1}^n \mathcal{L}_i(\alpha, \beta, \gamma)$$

where $\mathcal{L}_i$ is the likelihood function of observing the pattern of infection for infections with the i th serotype.

Regarding $\mathcal{L}_i$, we first need to obtain $\mathcal{L}_i$ *conditional* on $Y = y$. This is done in two stages:

- a) if infection is seen in the age interval (a_{j1}, a_{j2}) , then the j th child contributes a factor hS , where S is as defined above and h is the hazard function for the Weibull distribution. Then the probability of seeing this infection is phS , where p is the probability of detecting an infection if it exists,
- b) if the infection is not seen during (a_{j1}, a_{j2}) , then there are two possibilities :
1. either it is seen later with probability $S(a_{j2})$,
 2. or it is seen earlier with probability $1 - S(a_{j1})$.

Therefore the total probability is

$$\begin{aligned} & S(a_{j2}) + 1 - S(a_{j1}) \\ &= 1 - (S(a_{j1}) - S(a_{j2})). \end{aligned}$$

Another possibility is that the infection was in our window of observation but was not detected due to any reason, then the probability of such an event is

$$(1 - p)(S(a_{j1}) - S(a_{j2})).$$

Therefore the probability of not seeing an infection is

$$\begin{aligned} & 1 - (S(a_{j1}) - S(a_{j2})) + (1 - p)(S(a_{j1}) - S(a_{j2})) \\ &= 1 - p(S(a_{j1}) - S(a_{j2})). \end{aligned}$$

Therefore the likelihood function is:

$$\mathcal{L}_i(\alpha, \beta, \gamma) = \int_0^\infty \prod_{j=1}^N \{ph(a_j|y)S(a_j|y)\}^{\delta_j} (1 - p(S(a_{j1}|y) - S(a_{j2}|y)))^{1-\delta_j} f(y) dy, \quad (7.4)$$

where $\delta_j = 1$ if the j th of N children sustains an infection at age a_j .

The likelihood function must be evaluated numerically. This was done by integrating it numerically using a 10-point Gauss-Laguerre quadrature.

From the numerical maximisation of the likelihood function, estimates $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\gamma}$ are obtained.

To assess model fit and compare the model with data, expected and predicted numbers of children infected in 6-month age-bands were compared visually. The predicted proportion of children in age band (a_1, a_2) infected (by any serotype) is

$$\mathcal{P} = np \int_0^\infty (S(a_1|y) - S(a_2|y))f(y)dy.$$

It was found that with suitable parameter values, the model could reproduce the observed proportions of infection well. However, using the maximum-likelihood estimates (MLEs) the model tended to under-predict numbers infected. This is thought to be due to the fact that in the Empirical Bayes approach, the estimates of individual serotype infectivities get pulled towards the mode of the distribution $f(x)$. Large infectivities Y are decreased and small ones are increased. The net effect is to decrease the number of infections predicted. The bias would disappear for large samples given large numbers of infections.

The calculation of R_0 from equation 7.1 is now described. We assume a steady-state population. Then if $r(t)$ is pdf of lifespan (age at death) with $R(t)$ the corresponding survival function, then the pdf for the age of an individual is

$$g(t) = \frac{R(t)}{L},$$

where L is the average lifespan, i.e. $L = \int_0^\infty R(t)dt$.

The proportion of not yet infected (susceptible) individuals x is then given by

$$x = \int_0^\infty g(t)S(t)dt,$$

from which $R_0 = 1/x$ can be calculated.

It may seem curious that average lifespan L can affect the value of R_0 for a childhood disease. It does so because L is used, in conjunction with the steady-state assumption, to determine the age-structure of the whole population, which includes children. An observed pattern of infection denotes a more infectious disease (larger R_0) if the population contains a smaller proportion of children.

Hence assuming that everybody lives to exactly the average age L and then dies,

$$R(t) = \begin{cases} 1 & \text{if } t < L \\ 0 & \text{if } t > L \end{cases}$$

and so

$$g(t) = \begin{cases} 1/L & \text{if } t < L \\ 0 & \text{if } t > L \end{cases}$$

Then $x = \int_0^L S(t)dt/L$.

For a given value of Y , R_0 can be estimated as

$$R_0(y) = 1/x = L / \int_0^L S(t|y)dt,$$

and if S is small at age L , the upper limit of the integral can be set to infinity. The integral then becomes

$$\int_0^\infty S(t|y)dt = \Gamma(1/\beta)/(\beta y).$$

The mean value of R_0 is then

$$R_0 = L\beta \int_0^\infty f(y)ydy/\Gamma(1/\beta),$$

which evaluates to

$$R_0 = \frac{\beta\gamma L}{\alpha\Gamma(1/\beta)} = \mu\beta L/\Gamma(1/\beta), \quad (7.5)$$

where $\mu = \alpha/\gamma$ is the mean of the gamma distribution.

Since boys tend to have a greater hazard of infection than girls, it is possible to treat each sex separately and calculate two R_0 values, R_{0m} and R_{0f} from the data. Assuming equal numbers of men and women in the general population, with homogeneous mixing and that both sexes are equally infective, a single first-generation infected male denoted by $(1, 0)$ gives a vector $(R_{0m}/2, R_{0f}/2)$ in the second generation, and a single infected female $(0, 1)$ also gives rise to a vector $(R_{0m}/2, R_{0f}/2)$. The resulting transition matrix that carries first into second generation infections has an eigenvalue $(R_{0m} + R_{0f})/2$ and an associated eigenvector (R_{0m}, R_{0f}) , and this eigenvalue is taken as R_0 : the first-generation state (R_{0m}, R_{0f}) is scaled up by this R_0 to become a second-generation state with the same ratio of infected numbers between the sexes. At equilibrium, applying the equilibrium condition with uninfected proportions (x, y) leads to the conclusion that

$$R_0 = \frac{x + y - 2xy}{xy(2 - x - y)}. \quad (7.6)$$

This analysis was done for the highly isolated agents from table 4.3. These agents are *Adenovirus*, *Coxsackie B virus*, *E. coli* and *Echovirus*. Each of these agents has more than one serotype. From chapter 4 it is noticed that there are 41 serotypes of *Adenovirus*, but there are only 9 of them isolated in this study. There are only six serotypes of *Coxsackie B virus* out of which four were seen here. In the case of *E. coli* there are

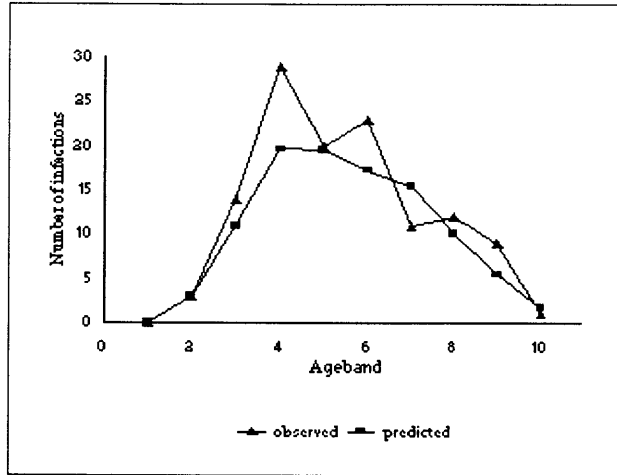


Figure 7.1: Observed versus predicted number of infections of *Adenovirus* for both sexes by age band

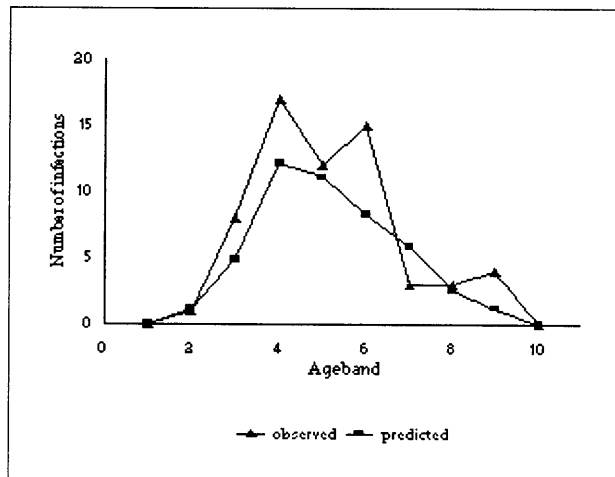


Figure 7.2: Observed versus predicted number of infections of *Adenovirus* for females by age band

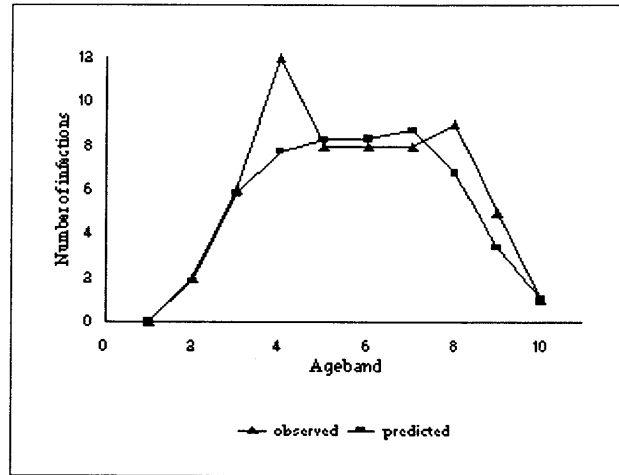


Figure 7.3: Observed versus predicted number of infections of *Adenovirus* for males by age band

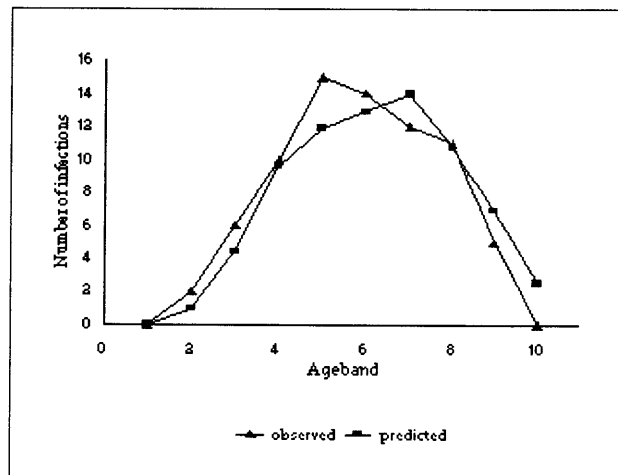


Figure 7.4: Observed versus predicted number of infections of *E. coli* for both sexes by age band

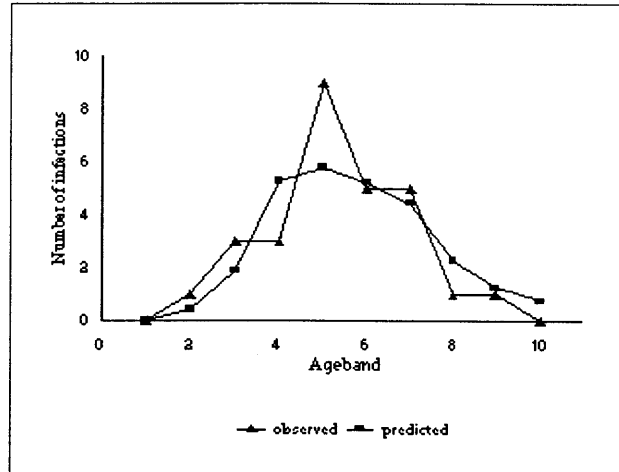


Figure 7.5: Observed versus predicted number of infections of *E. coli* for females by age band

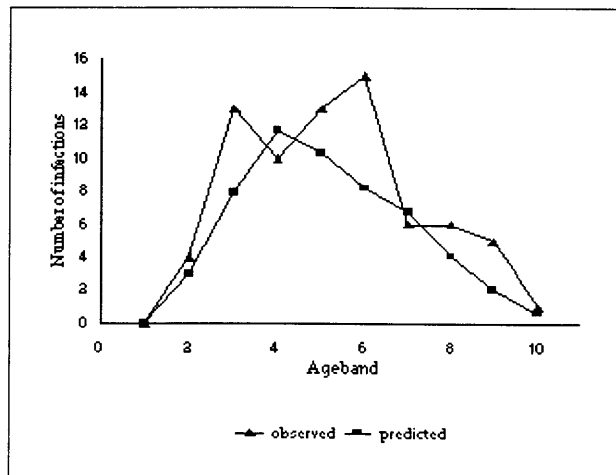


Figure 7.6: Observed versus predicted number of infections of *E. coli* for males by age band

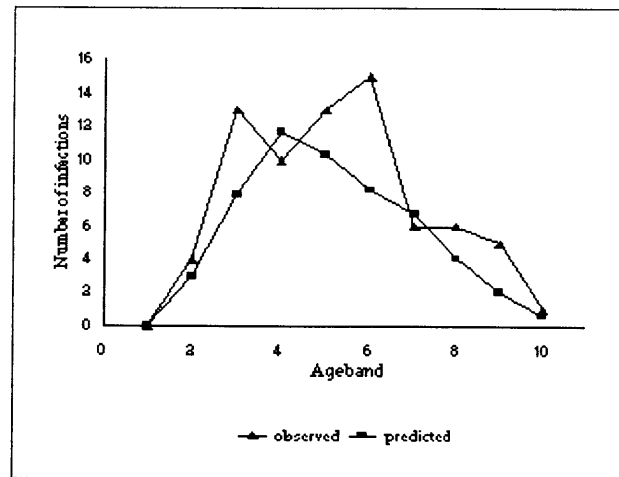


Figure 7.7: Observed versus predicted number of infections of *Coxsackie B* for both sexes by age band

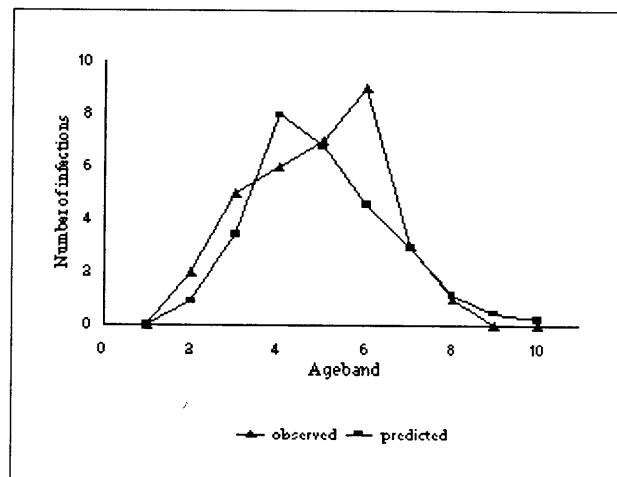


Figure 7.8: Observed versus predicted number of infections of *Coxsackie B* for females by age band

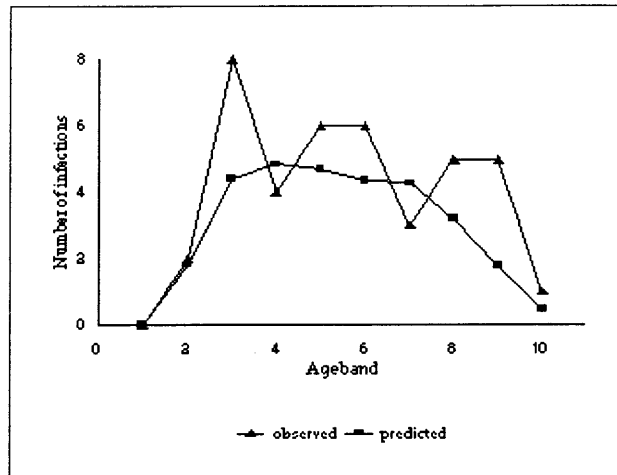


Figure 7.9: Observed versus predicted number of infections of *Coxsackie B* for males by age band

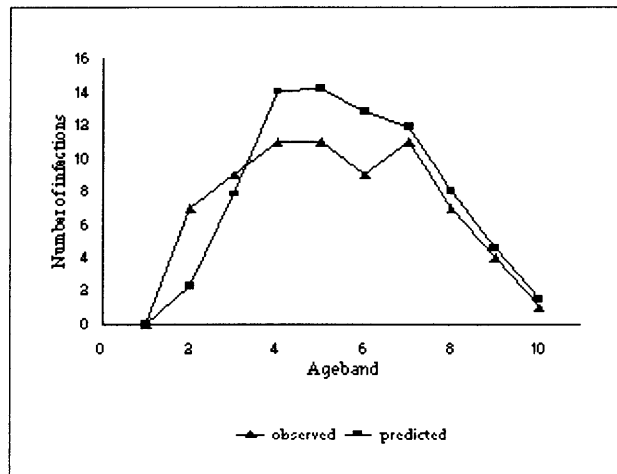


Figure 7.10: Observed versus predicted number of infections of *Echovirus* for both sexes by age band

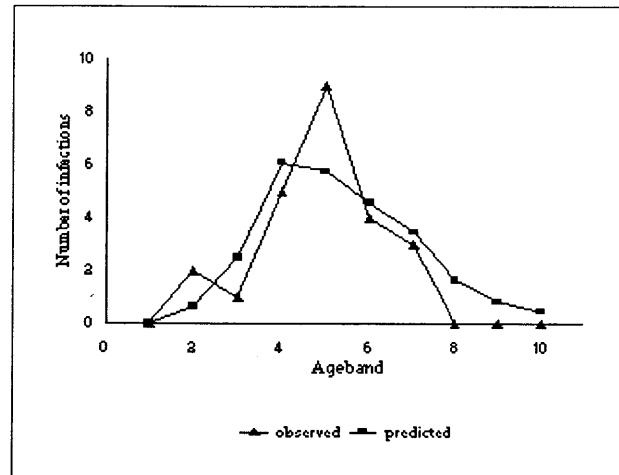


Figure 7.11: Observed versus predicted number of infections of *Echovirus* for females by age band

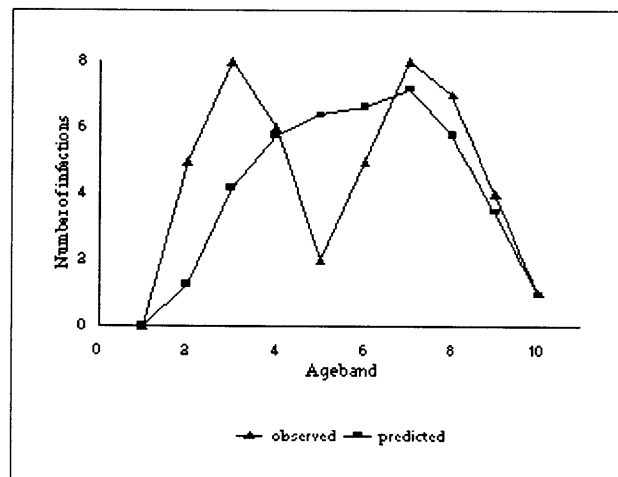


Figure 7.12: Observed versus predicted number of infections of *Echovirus* for males by age band

actually more than forty serotypes and fourteen of them were detected in this study. There are more than thirty serotypes of *Echovirus* and only about seven of them were isolated in this study.

For each of the above agents the above analysis was done for each sex individually and for the two sexes together. Estimates of α, β and γ for these agents were obtained and are displayed in table 7.1. These values were then substituted in equations 7.2 and 7.3 to calculate the expected numbers of infections for each agent and sex in each age band. The expected number of infections versus the observed by age band are plotted in figures 7.1 to 7.12. χ^2 goodness of fit tests were performed for each sex alone and for both sexes together. The results of the χ^2 values are displayed in 7.2. Under the null hypothesis that this fit is a reasonable one, this hypothesis must be rejected at the $\alpha = .01$ significance level if the value of the calculated χ^2 exceeds the table value of χ^2 with n degrees of freedom. Table 7.3 displays the different degrees of freedom and the table values of χ^2 . Only one of the χ^2 values in table 7.2 is significant at the 1% level. However, several others are close to being significant. In general, the model fits the data well in some cases and gives a fair fit in others.

The basic reproductive rate, R_0 was then estimated for each agent using the equilibrium equation defined earlier, under the assumption that life expectancy for men is 73.41 years and for women is 78.91 years (English Life Tables (1996)). Values for R_0 were estimated for each of the above agents for both sexes simultaneously and for each sex individually. These values are listed in table 7.4. The values of R_0 estimated here are high. This means that these agents are infectious and can spread easily in the community. This is expected in a childhood disease since childhood infectious diseases by definition catch people at

Agent	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\gamma}$
<i>Adenovirus</i> for all children	72.67	1.65	1.96
<i>Adenovirus</i> for females	68.93	1.97	2.06
<i>Adenovirus</i> for males	63.94	1.41	1.59
<i>E. coli</i> for all children	325.20	1.51	2.47
<i>E. coli</i> for females	189.53	1.36	1.33
<i>E. coli</i> for males	16.18	1.61	0.41
<i>Coxsackie B virus</i> for all children	59.54	5.38	2.49
<i>Coxsackie B virus</i> for females	200.0	1.63	7.96
<i>Coxsackie B virus</i> for males	38.01	1.84	1.46
<i>Echovirus</i> for all children	49.99	1.50	1.03
<i>Echovirus</i> for females	36.75	1.63	0.65
<i>Echovirus</i> for males	54.89	1.43	0.93

Table 7.1: Coefficients estimates

Agent	Females	Males	Both
<i>Adenovirus</i>	8.767	4.264	9.736
<i>E. coli</i>	2.33	5.175	4.548
<i>Coxsackie B virus</i>	3.564	8.370	14.897
<i>Echovirus</i>	3.438	16.558	12.776

Table 7.2: χ^2 values

a young age, Anderson and May (1991).

Agent	Females		Males		Both	
	df	χ^2	df	χ^2	df	χ^2
<i>Adenovirus</i>	3	11.345	4	13.227	5	15.086
<i>E. coli</i>	2	9.210	3	11.345	3	11.345
<i>Coxsackie B virus</i>	2	9.210	3	11.345	5	15.086
<i>Echovirus</i>	1	6.635	5	15.086	6	16.812

Table 7.3: Table χ^2 and degrees of freedom values

Agent	Females	Males	Both
<i>Adenovirus</i>	32.06	24.10	27.15
<i>E. coli</i>	7.32	23.65	7.60
<i>Coxsackie B virus</i>	21.44	37.33	41.48
<i>Echovirus</i>	18.59	17.36	20.60

Table 7.4: R_0 estimates

Error on R_0

The standard error of the estimate of R_0 as calculated from equation 7.5 could be found by several methods. The simplest but probably the least accurate would proceed using the covariance matrix M_{lm} of fitted parameter values estimated from equation 7.4 using the formula

$$M_{lm} = \partial^2 \log(\mathcal{L}_i)(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3) / \partial \theta_l \partial \theta_m,$$

where the three parameters α, β, γ have been renamed as $\theta_1, \theta_2, \theta_3$ respectively. The derivative is taken at the maximum of $\log(\mathcal{L})$.

The delta-method is then used to find the error of R_0 , and leads to

$$\delta R_0(\theta_1, \theta_2, \theta_3) = \sum_{l=1}^3 (\partial R_0 / \partial \theta_l) \delta \theta_l,$$

$$\hat{\sigma}_{R_0}^2 = \sum_{l=1}^3 \sum_{m=1}^3 (\partial R_0 / \partial \theta_l) (\partial R_0 / \partial \theta_m) M_{lm},$$

where $\hat{\sigma}_{R_0}^2$ is the estimated variance of the estimate of R_0 .

The covariance matrix M_{lm} can be found analytically from equation 7.4, or numerically. Once this is known, it is also possible to estimate $\hat{\sigma}_{R_0}^2$ by generating random variates from the trivariate Normal distribution with covariance matrix M_{ij} , recalculating R_0 (say) 1000 times from these randomised values of the parameters, and calculating the sample variance of the 1000 values of R_0 so obtained.

A variation on this theme is to reparameterise the model so that the model parameters are α, β and R_0 . The parameter γ is then found by making it the subject of equation 7.5. It is then possible to calculate the profile likelihood as a function of R_0 , by maximising the likelihood function for each value of R_0 with respect to the other two parameters. From the profile likelihood, the standard deviation of R_0 can be read off as the change in R_0 from its MLE (maximum-likelihood estimate) that causes the log-likelihood to decrease by $1/2$. This method avoids the need to calculate the covariance matrix, or to differentiate equation 7.5. The 95% confidence interval is the set of values of R_0 where the log-likelihood decreases by less than 0.5×1.96^2 from its maximum value.

The bootstrap method would give a more accurate estimate of $\hat{\sigma}_{R_0}^2$. Here we would draw 167 samples from the 167 children's data, sampling with replacement. Each of these resampled datasets would be used to recalculate R_0 , and $\hat{\sigma}_{R_0}^2$ would be found as before, as the sample variance of (say) 100 such resamples.

This would also be the easiest way of calculating standard errors on the probability p for each agent.

In practice, time did not permit these calculations to be carried out. Clearly, the various assumptions of the method of calculating R_0 will also give rise to an error in the final estimate, so that the statistical

error would be very much a lower bound on the true error.

Chapter 8

Conclusions and Further Work

The data studied here were a follow up study of two nursery schools in Salford. The data were collected between 1985 and 1987. One hundred and sixty-seven children were enrolled in the study. Faecal specimens were collected from the children fortnightly. These specimens were tested for certain agents of infection. Twenty-one such agents were isolated divided into three groups of micro-organisms, viruses, bacteria and protozoa. Episodes of gastro-intestinal disease were also reported. Demographic information about the subjects of the study were also obtained. The relationship between the isolation of infectious agents and episodes of illness were investigated. The relationship between the isolation of agents and the occurrence of illness and the social factors was also investigated using logistic regression. By fitting logistic regression models, it was found that certain agents of infection such as *Adenovirus*, *Coxsackie B virus*, *Echovirus* and *E. coli* had the highest contribution to the occurrence of gastro-intestinal illness in childhood. Social factors such as the number of parents the child lived with and the

number of parents employed had some contribution to the frequency of occurrence of childhood gastro-intestinal illness.

The basic reproductive rates, R_0 of the most frequently isolated agents were estimated from this data set. These agents are *Adenovirus*, *Coxsackie B virus*, *Echovirus* and *E. coli*. R_0 s were estimated for each agent for each sex alone and for the two sexes simultaneously. It was found that, for females the basic reproductive rate of *Adenovirus* is 32.06, for *E. coli* it is 7.32, for *Coxsackie B virus* it is 21.44 and for *Echovirus* it is 18.59. Estimates of R_0 for males differ from those for females. These are 24.10, 23.65, 37.33 and 17.36 for the above agents respectively. When R_0 's for the same agents were estimated for the whole data, estimates were 27.15, 4.60, 41.48, and 20.60 for *Adenovirus*, *Coxsackie B virus*, *Echovirus* and *E. coli* respectively. These estimates were expected to be high because these agents are infectious and this is expected in childhood diseases.

The Weibull survival model was fitted to the data. Twelve such models were fitted for each of the above agents for each sex alone and for the whole data. χ^2 goodness of fit tests were performed to test the fits. All but one fit were significantly good at $\alpha = 0.01$ level.

This study may be continued by deriving the mean and variance of the distribution defined in chapter 6. Two papers may be extracted from this study along with the one ready to be published, given in appendix B. One will be about the probability distribution of chapter 6 and the other about the estimates of the basic reproductive rates calculated in this thesis.

Appendix A

Glossary

This appendix contains a glossary of the medical and microbiological terms mentioned in the thesis. Most of these are taken from Stedman's Medical Dictionary, Stedman (1996).

adenitis :

Inflammation of a lymph node or of a gland.

adenoids :

1. A normal collection of unencapsulated lymphoid tissue in the nasopharynx. Also called pharyngeal tonsils.
2. Common terminology for the large (normal) pharyngeal tonsils of children.

aetiology (etiological) :

1. The science and study of the causes of disease and their mode of operation.
2. The science of causes, causality; in common usage, cause.

anterior :

1. In human anatomy, denoting the front surface of the body;

often used to indicate the position of one structure relative to another, i.e., situated nearer the front part of the body.

2. Near the head or rostral end of certain embryos.

3. Undesirable and confusing substitute for cranial in quadrupeds.

In veterinary anatomy, anterior is restricted to parts of the eye and inner ear.

4. Before, in relation to time or space.

antigen (Ag) (antigenic) :

Any substance that, as a result of coming in contact with appropriate cells, induces a state of sensitivity and/or immune responsiveness after a latent period (days to weeks) and which reacts in a demonstrable way with antibodies and/or immune cells of the sensitised subject in vivo or in vitro. Modern usage tends to retain the broad meaning of antigen, employing the terms "antigenic determinant" or "determinant group" for the particular chemical group of a molecule that confers antigenic specificity.

axilla, gen. and pl. axillae (axillary) :

The space below the shoulder joint, bounded by the pectoralis major anteriorly, the latissimus dorsi posteriorly, the serratus anterior medially, and the humerus laterally; it has a superior opening between the clavicle, scapula, and first rib (cervicoaxillary canal), and an inferior opening covered by the axillary fascia; it contains the axillary artery and vein, the infraclavicular part of the brachial plexus, axillary lymph nodes and vessels, and areolar tissue

bacillus, pl. bacilli :

1. A vernacular term used to refer to any member of the genus

Bacillus.

2. Term formerly used to refer to any rod-shaped bacterium.

bacteraemia :

The presence of viable bacteria in the circulating blood; may be transient following trauma such as dental or other iatrogenic manipulation or may be persistent or recurrent as a result of infection.

bacterium, pl. bacteria :

A unicellular prokaryotic microorganism that usually multiplies by cell division and has a cell wall that provides a constancy of form; they may be aerobic or anaerobic, motile or nonmotile, and free-living, saprophytic, parasitic, or pathogenic.

colon :

The division of the large intestine extending from the cecum to the rectum.

conjunctiva, pl. conjunctivae :

The mucous membrane investing the anterior surface of the eyeball and the posterior surface of the lids.

cyst :

1. A bladder.
2. An abnormal sac containing gas, fluid, or a semisolid material, with a membranous lining.

cytopathic effect :

A change or abnormality in cells due to viral infection.

diverticulum, pl. diverticula (diverticuli) :

A pouch or sac opening from a tubular or saccular organ, such as the gut or bladder.

duodenum, pl. duodena, gen. duodeni :

The first division of the small intestine, about 25 cm or 12 finger-breadths (hence the name) in length, extending from the pylorus to the junction with the jejunum at the level of the first or second lumbar vertebra on the left side. It is divided into the superior part, the first part of which is the duodenal cap, the descending part, into which the bile and pancreatic ducts open, the horizontal (inferior) part and the ascending part, terminating at the duodenojejunal junction.

dysuria :

Difficulty or pain in urination.

edema (edematous, oedema) :

An accumulation of an excessive amount of watery fluid in cells, tissues, or serous cavities.

electrolyte :

Any compound that, in solution, conducts electricity and is decomposed (electrolysed) by it; an ionisable substance in solution.

endemic disease :

The constant presence of a disease or infectious agent within a given geographic area or population group, may also refer to the usual prevalence of a given disease within such area.

endogenous infection :

Infection by an organism which is part of the normal body mi-

croflora.

enterotoxin :

A cytotoxin specific for the cells of the intestinal mucosa.

epidemic :

The occurrence in a community or region of cases of an illness, specific health related behaviour or other health-related events clearly in excess of normal expectancy. The community or region and the period in which the cases occur are specified precisely. The number of cases indicating the presence of an epidemic varies according to the agent, size and type of population exposed, previous experience or lack of exposure to the disease and the time and place of occurrence. Epidemicity is thus relative to usual frequency of the disease in the same area among the specified population, at the same season of the year. A single case of a communicable disease long absent from a population or first invasion by a disease not previously recognised in that area is required immediate reporting and full field investigation; two cases of such a disease associated in time and place may be sufficient evidence to be considered an epidemic. (Recent epidemic initially identified following the occurrence of small numbers of cases include the epidemic of vaginal cancer in daughters of women who took diethylstilbrestol during the pregnancy and the pandemic of AIDS that was heralded by a report of cases of *Pneumocystis carinii* pneumoonia among gay men in Los Angeles in 1981). The purpose of surviellence systems such as the EPIDEMIC INTELLIGENCE SERVICE is to identify epidemics as early as possible so that effective control measures can be put in place. This re-

mains the most important use for epidemiology. The word may be also used to describe outbreaks of disease in animal or plant populations.

epithelium, pl. epithelia (epithelial) :

The purely cellular avascular layer covering all the free surfaces, cutaneous, mucous, and serous, including the glands and other structures derived therefrom.

eradication :

termination of all transmission of infection by extermination of the infectious agent through surveillance and containment. Eradication as in the instance of smallpox, was based on the joint activities of control and surveillance. Regional eradication has been successful with malaria and in some countries appears close to succeeding for measles. The term "elimination" is sometimes used to describe eradication of disease such as measles from large geographic region or political jurisdiction. In 1992, the WHO put it this way: Eradication is defined as achievement of a status whereby no further cases of a disease can occur anywhere and continued control measures are unnecessary. Smallpox was eradicated in 1977, an eradication based on joint control and surveillance. Current candidates for eradication are poliomyelitis and dracunculosis.

erythema :

Redness of the skin due to infection or inflammation.

exogenous :

Originating or produced outside of the organism. For example

exogenous *Rotavirus* is a *Rotavirus* that infects a cell from a source outside the cell.

fauces, gen. faucium :

The space between the cavity of the mouth and the pharynx, bounded by the soft palate and the base of the tongue.

flagellum, pl. flagella :

A whip-like locomotory organelle of constant structural arrangement consisting of nine double peripheral microtubules and two single central microtubules; it arises from a deeply staining basal granule, often connected to the nucleus by a fibre, the rhizoplast. Though characteristic of the protozoan class Mastigophora, comparable structures are commonly found in many other groups, e.g., in spermatozoa.

follicle :

1. A more or less spherical mass of cells usually containing a cavity.
2. A crypt or minute cul-de-sac or lacuna, such as the depression in the skin from which the hair emerges.

fomes, pl. fomites :

Objects, such as clothing, towels, and utensils that possibly harbour a disease agent and are capable of transmitting it; usually used in the plural.

gastro-enteritis :

Inflammation of the mucous membrane of both stomach and intestine.

genu, pl. genua, gen. genus :

1. The place of articulation between the thigh and the leg.
2. Any structure of angular shape resembling a flexed knee.

genus, pl. genera (subgenera) :

In natural history classification, the taxonomic level of division between the family, or tribe, and the species; a group of species alike in the broad features of their organisation but different in detail, and incapable of fertile mating.

Gram-negative :

Refers to the inability of a bacterium to resist decolourisation with alcohol after being treated with Gram's crystal violet. However, following decolourisation, these bacteria can be readily counter-stained with safranin, imparting a pink or red colour to the bacterium when viewed by light microscopy. This reaction is usually an indication that the outer structure of the bacterium consists of a cytoplasmic (inner) membrane surrounded by a relatively thin peptidoglycan layer, which in turn, is surrounded by an outer membrane.

Gram-positive :

Refers to the ability of a bacterium to resist decolourisation with alcohol after being treated with Gram's crystal violet stain, imparting a violet colour to the bacterium when viewed by light microscopy. This reaction is usually an indication that the outer structure of the bacterium consists of a cytoplasmic membrane surrounded by a thick, rigid bacterial cell wall comprised of peptidoglycan.

haematuria :

Any condition in which the urine contains blood or red blood cells.

haemolysin :

1. Any substance elaborated by a living agent and capable of causing lysis of red blood cells and liberation of their hemoglobin.
2. A sensitising (complement-fixing) antibody that combines with red blood cells of the antigenic type that stimulated formation of the hemolysin, affecting the cells in such a manner that complement fixes with the antibody-cell union and causes dissolution of the cells, with liberation of their hemoglobin.

haemolytic :

Destructive to blood cells, resulting in liberation of hemoglobin.

herpangina :

A disease caused by types of *Coxsackie virus* and marked by vesiculopapular lesions about 1 to 2 mm in diameter which are present around the fauces and soon break down to form grayish yellow ulcers; accompanied by sudden onset of fever, loss of appetite, dysphagia, pharyngitis, and sometimes abdominal pain, nausea, and vomiting.

histology :

The science concerned with the minute structure of cells, tissues, and organs in relation to their function.

hypochondriasis :

A morbid concern about one's own health and exaggerated attention to any unusual bodily or mental sensations; a delusion that

one is suffering from some disease for which no physical basis is evident.

ileum :

The third portion of the small intestine, about 12 feet in length, extending from the junction with the jejunum to the ileocecal opening. It is distinct from jejunum in being typically smaller in diameter with thinner walls, having smaller and less complex plicae circulares, its mesentery having more fat and its arteries (ileal arteries) forming more tiers of arterial arcades with shorter vasa recta.

immunogenic :

The ability of an infectious agent to induce specific immunity.

inclusion :

1. Any foreign or heterogeneous substance contained in a cell or in any tissue or organ, not introduced as a result of trauma.
2. The process by which a foreign or heterogeneous structure is misplaced in another tissue.

incubation period :

1. The time interval between invasion by an infectious agent and appearance of first sign of the disease in question.
2. In a "vector", the period between entry of the infectious agent into the vector and the time at which the vector becomes infective; i.e. transmission of the infectious agent from the vector to a fresh final host is possible.

infection :

The entry and development or multiplication of an infectious

agent in the body of man or animals. Infection is not synonymous with infectious disease; the result may be inapparent or manifest. The presence of living infectious agents on exterior surfaces of the body is called "infestation". The presence of living infectious agents upon articles of apparel or solid articles is not infection, but represents "contamination" of such articles.

inoculate :

1. To introduce the agent of a disease or other antigenic material into the subcutaneous tissue or a blood vessel, or through an abraded or absorbing surface for preventive, curative, or experimental purposes.
2. To implant microorganisms or infectious material into or upon culture media.
3. To communicate a disease by transferring its virus.

intranasal :

Within the nasal cavity.

intussusception :

1. The taking up or receiving of one part within another, especially the enfolding of one segment of the intestine within another.
2. Often, specifically, the process of incorporation of new material in the growth of the cell wall.

jejunum :

The portion of small intestine, about 8 feet in length, between the duodenum and the ileum. The jejunum is distinct from the ileum in being more proximal, of larger diameter with a thicker wall, having larger, more highly developed plicae circulares, being more

vasculer (redder in appearance) with the jejunal arteries forming fewer tiers of arterial arcades and longer vasa recti.

latent period :

Delay between exposure to a disease-causing agent and the appearance of manifestation of the disease. After exposure to ionising radiation, for instance, there is a latent period of 5 years, on average, before development of leukemias, and more than 20 years before development of certain malignant conditions. The term "latent period" is always used synonymously with induction period, that is the period between exposure to a disease causing agent and the appearance of manifestations of the disease. It has also been defined as the period from disease initiation to disease detection. In infectious disease epidemiology this corresponds to the period between exposure and the onset of infectiousness.

lesion :

1. A wound or injury.
2. A pathologic change in the tissues.
3. One of the individual points or patches of a multifocal disease.

leukocyte (leucocyte) :

A type of cell formed in the myelopoietic, lymphoid, and reticular portions of the reticuloendothelial system in various parts of the body, and normally present in those sites and in the circulating blood (rarely in other tissues). Under various abnormal conditions, the total numbers or proportions, or both, may be characteristically increased, decreased, or not altered, and they may be present in other tissues and organs. Leukocytes represent three lines of development from primitive elements: myeloid,

lymphoid, and monocytic series. On the basis of features observed with various methods of staining with polychromatic dyes (e.g., Wright's stain, and others), cells of the myeloid series are frequently termed granular leukocyte's, or granulocytes; cells of the lymphoid and monocytic series also have granules in the cytoplasm but, owing to their tiny, inconspicuous size and different properties (frequently not clearly visualised with routine methods), lymphocytes and monocytes are sometimes termed nongranular or agranular leukocyte's. Granulocytes are commonly known as polymorphonuclear leukocytes (also polynuclear or multinuclear leukocyteS), inasmuch as the mature nucleus is divided into two to five rounded or ovoid lobes that are connected with thin strands or small bands of chromatin; they consist of three distinct types: neutrophils, eosinophils, and basophils, named on the basis of the staining reactions of the cytoplasmic granules. Cells of the lymphocytic series occur as two, somewhat arbitrary, normal varieties: small and large lymphocytes; the former represent the ordinary forms and are conspicuously more numerous in the circulating blood and normal lymphoid tissue; the latter may be found in normal circulating blood, but are more easily observed in lymphoid tissue. The small lymphocytes have nuclei that are deeply or densely stained (the chromatin is coarse and bulky) and almost fill the cells, with only a slight rim of cytoplasm around the nuclei; the large lymphocytes have nuclei that are approximately the same size as, or only slightly larger than, those of the small forms, but there is a broader, easily visualised band of cytoplasm around the nuclei. Cells of the monocytic series are usually larger than the other leukocyte's, and are

characterised by a relatively abundant, slightly opaque, pale blue or blue-gray cytoplasm that contains myriads of extremely fine reddish-blue granules. Monocytes are usually indented, reniform, or shaped similarly to a horseshoe, but are sometimes rounded or ovoid; their nuclei are usually large and centrally placed and, even when eccentrically located, are completely surrounded by at least a small band of cytoplasm.

malaise :

A feeling of general discomfort or uneasiness, an "out-of-sorts" feeling, often the first indication of an infection or other disease.

mesentery (mesentric) :

1. A double layer of peritoneum attached to the abdominal wall and enclosing in its fold a portion or all of one of the abdominal viscera, conveying to it its vessels and nerves.
2. The fan-shaped fold of peritoneum encircling the greater part of the small intestines (jejunum and ileum) and attaching it to the posterior abdominal wall at the root of the mesentery (radix mesenterii).

morbidity :

1. A diseased state.
2. The ratio of sick to well in a community.
3. The frequency of the appearance of complications following a surgical procedure or other treatment.

myocarditis :

Inflammation of the muscular walls of the heart.

neutrophil, neutrophile (neutrophils) :

1. A mature white blood cell in the granulocytic series, formed by myelopoietic tissue of the bone marrow (sometimes also in extramedullary sites), and released into the circulating blood, where they normally represent from 54of leukocytes. When stained with the usual Romanovsky type of dyes, neutrophil's are characterised by: 1) a nucleus that is dark purple-blue, lobated (three to five distinct lobes joined by thin strands of chromatin), and with a rather coarse network of fairly dense chromatin; 2) a cytoplasm that is faintly pink (sharply contrasted with the nucleus) and contains numerous fine pink or violet-pink granules, i.e., not acidophilic or basophilic (as in eosinophils or basophils). The precursors of neutrophil's, in order of increasing maturity, are: myeloblasts, promyelocytes, myelocytes, metamyelocytes, and band forms. Although the terms neutrophilic leukocytes and neutrophilic granulocytes include younger cells in which neutrophilic granules are recognised, the two expressions are frequently used as synonyms for neutrophil's, which are mature forms unless otherwise indicated by a modifying term, such as immature neutrophil. See Also: leukocyte, leukocytosis.

2. Any cell or tissue that manifests no special affinity for acid or basic dyes, i.e., the cytoplasm stains approximately equally with either type of dye.

oncogene (oncogenic) :

1. Any of a family of genes which under normal circumstances code for proteins involved in cell growth or regulation (e.g., protein kinases, GTPases, nuclear proteins, growth factors) but may foster malignant processes if mutated or activated by contact with

retroviruses. Identified oncogenes include the ras, originally associated with bladder tumours, and the p53, a mutated version of a gene on chromosome 17 which normally corrects for mutations caused by ultraviolet radiation. p53 now has been shown to be involved in cancers of the breast, cervix, ovary, and lung, among others. Oncogenes often work in concert to produce cancer, and their action may be exacerbated by retroviruses, jumping genes, or inherited genetic mutations.

2. Found in certain DNA tumour viruses. It is required for viral replication.

orphan virus :

It is a virus that is not known to cause any disease.

osinophilic :

Staining readily with eosin dyes; denoting such cell or tissue elements.

palpebral :

Relating to an eyelid or the eyelids.

papule :

A small, circumscribed, solid elevation on the skin.

pathogenesis :

The pathologic, physiologic, or biochemical mechanism resulting in the development of a disease or morbid process.

pleurisy :

Inflammation of the pleura.

prevalence :

The number of events e.g. instances of a given disease or other condition in a given population at a designated time; sometimes used to mean "prevalence rate". When used without qualification, the term usually refers to the situation at a specified point in time. Note that this is a number not a rate.

protozoan cyst (protozoa) :

infectious form of many protozoan parasites such as *Entamoeba histolytica*, *Giardia lamblia*, *Balantidium coli*, etc., usually passed in the feces and provided with a highly condensed cytoplasm and resistant cell wall.

pulmonary :

Relating to the lungs, to the pulmonary artery, or to the aperture leading from the right ventricle into the pulmonary artery.

pus :

A fluid product of inflammation, consisting of a liquid containing leukocytes and the debris of dead cells and tissue elements liquefied by the proteolytic and histolytic enzymes (e.g., leuko-protease) that are elaborated by polymorphonuclear leukocytes.

remission :

1. Abatement or lessening in severity of the symptoms of a disease.
2. The period during which such abatement occurs.

septicaemia :

Systemic disease caused by the spread of microorganisms and

their toxins via the circulating blood; formerly called "blood poisoning".

serotyping :

This classical technique distinguishes between strains of viruses or bacteria by difference in their antigenic structure, recognised by reaction with specific antisera. The established schemes use polyclonal antisera but newer schemes based on monoclonals are being developed. Serotyping requires the production and maintenance of appropriate banks of antisera which is both time-consuming and costly, and is thus usually restricted to reference laboratories.

sputum, pl. sputa :

1. Expecterated matter, especially mucus or mucopurulent matter expecterated in diseases of the air passages.
2. An individual mass of such matter.

streptococcus (streptococcal) :

A genus of nonmotile (with few exceptions), nonsporeforming, aerobic to facultatively anaerobic bacteria (family Lactobacillaceae) containing Gram-positive, spherical or ovoid cells which occur in pairs or short or long chains. Dextrorotatory lactic acid is the main product of carbohydrate fermentation. These organisms occur regularly in the mouth and intestines of humans and other animals, in dairy and other food products, and in fermenting plant juices. Some species are pathogenic. The type species is *Streptococcus pyogenes*.

stomatitis :

Inflammation of the mucous membrane of the mouth.

syndrome :

The aggregate of signs and symptoms associated with any morbid process, and constituting together the picture of the disease.

tenesmus :

A painful spasm of the anal sphincter with an urgent desire to evacuate the bowel or bladder, involuntary straining, and the passage of little fecal matter or urine.

trachea, pl. tracheae (tracheal) :

The air tube extending from the larynx into the thorax (level of the fifth or sixth thoracic vertebra) where it bifurcates into the right and left main bronchi. The trachea is composed of from 16 to 20 rings of hyaline cartilage connected by a membrane (annular ligament); posteriorly, the rings are deficient for one-fifth to one-third of their circumference, the interval forming the membranous wall being closed by a fibrous membrane containing smooth muscular fibres. Internally, the mucosa is composed of a pseudostratified ciliated columnar epithelium with mucous goblet cells; numerous small mixed mucous and serous glands occur, the ducts of which open to the surface of the epithelium.

trophozoite :

The amoeboid, vegetative, asexual form of certain Sporozoea, such as the schizont of the plasmodia of malaria and related parasites.

ureamia :

1. An excess of urea and other nitrogenous waste in the blood.
2. The complex of symptoms due to severe persisting renal failure that can be relieved by dialysis.

vacuolation (vacuoles) :

Any of various membrane-delimited compartments within a cell.

vesicle :

1. A small (less than 0.5 cm) circumscribed elevation of the skin containing fluid.
2. A small sac containing liquid or gas.

viable :

Capable of living; denoting a fetus sufficiently developed to live outside of the uterus.

viraemia :

The presence of a virus in the bloodstream.

virus, pl. viruses :

1. Formerly, the specific agent of an infectious disease.
2. Specifically, a term for a group of infectious agents which with few exceptions are capable of passing through fine filters that retain most bacteria, are usually not visible through the light microscope, lack independent metabolism, and are incapable of growth or reproduction apart from living cells. They have a prokaryotic genetic apparatus but differ sharply from bacteria in other respects. The complete particle usually contains only DNA or RNA, not both, and is usually covered by a protein shell or capsid that protects the nucleic acid. They range in size from 15 mm up to several hundred mm. Classification of virus's depends upon characteristics of virions as well as upon mode of transmission, host range, symptomatology, and other factors. For virus's not listed below, see the specific name.

3. Relating to or caused by a virus, as a virus disease.

zoonosis (zoonotic) :

An infection or infestation shared in nature by humans and other animals that are the normal or usual host; a disease of humans acquired from an animal source.

Appendix B

Publication

This article is done in collaboration with the workers on this data set.

It is to be submitted for publication in the very near future.

Study of infection in day nurseries

SUMMARY

A prospective study of enteric infection among 167 children in two Salford day-nurseries was carried out. Over the one-year study period, symptoms of illness were recorded, and fortnightly stool samples were tested for the presence of a large number of known or potential enteric pathogens.

The study demonstrated that the spectrum of pathogens associated with mild diarrhoeal illness in this setting is strikingly different from that found in studies of children admitted to hospital with gastroenteritis.

The main findings were that the incidence of diarrhoea was about half that in developing countries, with much shorter episodes, and that the majority of diarrhoeal incidents could be attributed to *Echovirus*, *Coxsackie B* virus, *Cryptosporidium* sp. and *Campylobacter* sp. There was also much asymptomatic carriage of these agents. Isolation of enteropathogenic *Escherichia Coli* as a sole isolate was shown to have no association with gastro-enteritis.

Rotavirus was largely absent, and *Salmonella* sp. were entirely absent. Reasons for this are discussed in the text.

INTRODUCTION

This prospective study of diarrhoeal illness and enteric infection in day-nursery children was carried out in 1986–1987. At that time, children shown to be excreting enteric pathogens such as enteropathogenic *E. Coli* were excluded from day-nurseries. This study was designed to estimate the incidence of diarrhoeal illness among children of age 1 to

4 years, and to establish which agents were responsible for the majority of incidents of community diarrhoeal illness, and the extent of asymptomatic transmission of enteric agents.

Inner city Social Services day nurseries were chosen for the study, because of the known high risk of enteric infections in such populations, and also because such populations are sufficiently stable to permit prospective studies to be carried out.

Since the study was performed the policy of excluding symptomless excretors has been abandoned even for *Shigella sonnei*. Instead, emphasis is placed on exclusion of children whilst acutely symptomatic together with promotion of good hygiene. Nonetheless, the results of this study are still of current interest. As a longitudinal study of one-year duration (repeated in two nurseries), with fortnightly sampling, it meets the design criteria for such studies used by Bern *et al* [3] in their meta-analysis of diarrhoea morbidity and mortality in developing countries. To our knowledge, there has been no comparable study of diarrhoea in the community in the developed world, although morbidity statistics from general practice in the UK are available [8, 9].

In this paper, we wish to present our major findings, i.e. on diarrhoeal incidence, incidence of infection with potential enteric pathogens, extent of asymptomatic excretion of potential pathogens, and the extent to which enteric illness episodes could be attributed to the various agents. We shall consider briefly the extent of the inevitable under-recording of infection rates caused by the imperfect sensitivity of tests for agents in stool, and by the less than 100% compliance with the intended stool samples programme.

A subsequent paper on this study is planned, in which an ambitious statistical analysis will be described, which led to corrections of

infection rates, and to calculations of the basic reproductive rate R_0 for individual infectious agents.

METHODS

SUBJECTS AND SETTING

The study was approved by the Research Ethics Committee of the South Manchester Health Authority.

The study population was recruited from two Social Services Department day care nurseries in a socially deprived area of Greater Manchester. This setting was chosen to maximise the isolation rate of faecal pathogens, in view of the recognised association between social deprivation and frequent childhood diarrhoea [8]. Nursery places were provided free of charge to children who had been referred to the Social Services Department by a Health Visitor or Social Worker as giving cause for concern. Such causes for concern included: the child's name being on the Child Protection Register; parents having difficulties coping; single parent families; and children with disabilities.

The support of nursery staff for the study was sought by organising a series of presentations on the study's aims and methods. The parents of all current attenders at nursery 1 during April 1986, and at nursery 2 during April 1987, were sent a covering letter and consent form. Parents whose children were enrolled subsequently were also asked to consent. Parents who did not return a consent form were spoken to by staff when they brought their child to the nursery, to give a more detailed description of the study and to encourage them to consent to their child's participation. A single nursery nurse was responsible for collecting faecal specimens and collating demographic and microbiological data.

FAECAL SPECIMEN AND DATA COLLECTION

The aim was to obtain a faecal specimen for each child who was enrolled at the nursery every two weeks for a one year follow up period or until the child left the nursery. Hence the study data consist of a cohort of observations abridged into 26 fortnights, with both left and right censored cases. If symptoms of gastroenteritis were reported during a child-fortnight (C-F), this was recorded as a single episode. Multiple episodes within a C-F were not permitted. This may represent a slight under-recording of episodes, as previously published work suggests that symptoms recurring after a four day interval can be regarded as a separate episode [2].

A two week period was chosen because it had been shown by population surveys to be the optimal trade-off between under-reporting due to memory failure associated with longer periods of recall, and over-reporting, which has been attributed to the reporting of short duration symptoms episodes that turn out not to be judged as 'illness' once a few days have elapsed [4]. Sampling at least fortnightly was required by Bern *et al* [3] in their meta-analysis of diarrhoea in developing countries.

If symptoms of gastroenteritis occurred in a fortnight, every effort was made to collect one or more diarrhoeal specimens and all isolates resulting from these specimens were recorded. When a child was absent from the nursery because of illness, enquiries were made and the Environmental Health Department was notified if gastroenteritis was suspected.

Specimens were obtained directly from children in diapers and normally from a potty or occasionally from the lavatory bowl in children who had gained bowel control. There was a separate lavatory for each nursery class, and there were three classes per nursery. Stool collec-

tion charts were placed in each lavatory to enable staff to participate in stool collection. The practice in both schools was to encourage all the children to visit the toilet together at set times of the day. Even though the research nurse was present on at least two days per week, it was necessary for some help to be given by staff to achieve complete specimen collection. Specimens were delivered personally by the research nurse to the laboratory on the day of collection. Cover was provided for the research nurse's vacations and sickness absences by local Environmental Health Officers (EHOs), who obtained specimens and transported them to the laboratory in the same way. Where a child was known to be absent from the nursery because of gastroenteritis, EHOs visited the household and attempted to obtain a faeces sample.

Family and demographic data were obtained from the nursery records and by interviewing parents. Histories of symptoms of gastroenteritis were obtained from the report of parents and/or staff. Gastroenteritis was taken to be: 'diarrhoea', defined as at least one liquid stool; 'loose stools', defined as at least one stool which was not properly formed and was atypical in that regard for the child; 'vomiting', defined as any reported emesis; or any combination of these symptoms. Although a record was kept of children who were examined by their General Practitioner, no access to clinical notes from these examinations was sought.

LABORATORY METHODS

Stools were routinely cultured for enteropathogenic *E. coli*, *Salmonellas*, *Campylobacter*, *Shigellas* and *Yersinia* using standard methods. Unstained saline preparations were examined by light microscope for

ova, cysts and parasites. In addition, stools were examined for *Cryptosporidium* sp. using smears stained by modified Ziehl-Neelsen stain.

For electron microscopy (EM), all specimens were stored at +4°C. 20% faecal suspensions in faecal transport medium were centrifuged at 15000 g for 15 minutes at +8°C. 2 ml of supernatant was transferred to a polycarbonate ultracentrifuge tube and spun at 65000 g in an ultracentrifuge for one hour at +8°C. The supernatant from the ultracentrifuge spin was tipped off and the pellet resuspended in the small amount of fluid remaining. Viruses were adsorbed onto Formvar-carbon coated electron microscope specimen grids, stained with 3% PTA and examined at 63000 magnification in an AE1 EM801 electron microscope.

For virus isolation, a thawed 20% faecal emulsion (0.1 ml) was inoculated into 3 cell lines: primary baboon kidney (BK), diploid fibroblasts (MRC 5), and continuous human epithelial cells (HEp 2). These were examined twice a week for evidence of cytopathic effect (CPE), and negative cultures discarded after 2 weeks. Specimens showing viral CPE were identified by standard methods.

Many diagnostic developments have occurred since this study was undertaken, such as the use of ELISA for *Rotavirus* and *Adenovirus*, RT-PCR for SRSV's etc. Results may have differed if such additional techniques had been available, particularly with regard to viral types, such as Adeno types 40 and 41 [17]. Electron microscopy does not distinguish between viruses of the same morphological type. We have taken (untyped) Adenovirus found only with the EM, and not in tissue culture, to be an enteric agent, and all the (typed) Adenovirus that were also found in tissue culture to be respiratory adenovirus.

STATISTICAL METHODS

All data manipulations and statistical tests were performed with the statistics package SPSS-PC [13]. Confidence intervals were obtained by assuming a Poisson distribution for numbers of episodes of diarrhoea or infection. Fisher's exact test was used for testing for association between infections and episodes of gastroenteritis.

Where no significant association was found based on small numbers of cases, such as in the case of *Calicivirus* this should not be taken as evidence that no association exists. Data including small numbers of cases may, however, be of use in future meta-analyses which seek to combine cases from several studies in order to perform a significance test with sufficient statistical power to reach a conclusion of no association.

RESULTS

STUDY POPULATION AND COMPLIANCE

Before quoting the findings of the study, we describe the study population, and the degree of compliance obtained. A total of 93 children attended day-nursery 1 and 80 attended day-nursery 2 during the study periods. Six parents did not consent to their child's inclusion; the participation rate was therefore 167/173 (96.5%). There were no statistically significant differences between included children in the two nursery populations in sex ratio, attendance, duration of enrolment, proportion of missing specimens or rates of diarrhoea, loose stools or vomiting.

Table B.1 shows the study population classified by age and sex. The modal age at entry to the study was 2 years for males and 1 year for females, with 68.3% of the study population aged 2 years or less. The mean age for males was 2.37 years and for females it was 2.32 years.

Six (3.6%) children were 6 months of age or less at the beginning of the study.

The lower proportion of females in the study population (36%) was unlikely to have resulted from chance, (binomial test on an expectation of 50%, $p=0.0003$). The preponderance of males may have reflected their greater tendency to suffer behaviour problems deemed by the Social Services Department to represent a need for a nursery place.

Turning to other indicators of the type of population under study, table B.2 shows numbers of siblings. The proportion of children with at least one sibling in the study population was 68.9%. As maternal age was not recorded, this can not be compared with national statistics to give an indicator of relative family size.

Table B.3 shows single-parent and unemployment status. The percentage of unemployed fathers is 64.5%, which compared with the national average for 1987 of 13.5%. The percentage of single-parent families was 47%, which compared with the national average of 9% [15, 10].

This gives a picture of a deprived population, with many more single-parent families and unemployed family members than the national average.

GASTROENTERITIS REPORTING AND STOOL SAMPLING

Table B.4 shows the incidence density of reported gastroenteritis. Altogether, there were 136 child-fortnights in which gastroenteritis was reported. Of these, 26 C-Fs were associated with an isolate of the same organism and/or serotype from the same child in the preceding fortnight. These were taken to be continuing excretion of organism within the same episode. Hence there were a total of 110 first symptomatic episodes, including 16 (14.5%) where no faeces specimen was obtained.

Length of attendance did not appear to bias the incidence density of reported symptoms of gastroenteritis. Those who attended for 12 weeks or less reported 29/646 (4.49%) symptomatic C-Fs compared to 82/1819 (4.51%) symptomatic C-Fs reported by those who attended between 13 to 26 weeks.

Figure 1 shows the seasonal variation of gastroenteritis incidence. There was an autumn peak. The continuous line at the top of Figure 1 shows the percentage absence from nurseries in successive fortnights commencing January 1st. Attendance shows no overall trend and although there are dips in attendance at the 13th and 24th fortnight, they are not apparently related to the fortnightly period prevalence of gastroenteritis. The seasonal variation shown therefore appears genuine.

Males aged two years or under showed both the lowest attendance rates (Table B.5) and the highest incidence density of gastroenteritis (Table B.4). This is consistent with previous work ([6]).

Table B.4 shows a decline in incidence density of gastroenteritis with age. Overall, there was a slightly higher incidence density among females, 1.20 episodes per child-year versus 1.11 in males (Chi square for difference in proportions = 0.15, $p=0.70$). This overall figure, however, conceals a reversal in sex ratio of incidence density with age. This finding is consistent with a lower hazard of infection for females, so that they acquire infection at a higher average age.

The dependence of infection rates on age and sex found in this study was consistent with a picture of infection rate declining with age, as full or partial immunity is progressively acquired to particular serotypes. Males, perhaps because of greater body-contact, have a greater hazard of acquiring infection, so that their infection rate peaks earlier than

that for females.

Birth order was investigated but found to be unimportant, as was time from enrolment. It had been surmised that children might acquire more infection immediately on entry to the nursery.

We now turn to the stool sampling. Table B.5 shows numbers of fortnights enrolled and stool samples taken, for each of the four age and sex groups. Each group contained at least 400 enrolled fortnights, providing considerable statistical power to make comparisons between these groups.

Despite the best efforts of the nursery nurse, faecal specimens were collected on only half of the possible occasions. The reasons for non-collection included: periods of high nursery staff absence (it was impossible for the nurse to collect specimens from all the children if she received no help from staff); the fact that some children's bowel habit was such that they did not defaecate during the day; and some children's refusal to participate in specimen collection. As noted earlier, in fortnights when symptoms occurred, great efforts were made to obtain samples, and only 14.5% of such samples were missing.

In the age group two years or under, a faeces specimen was significantly less likely to be obtained from males (Chi square 12.9, $p < 0.001$).

AGENTS OF INFECTION

Table B.6 shows isolates and their association with reported symptoms of gastroenteritis occurring during the same C-F. Of the 1175 faecal specimens which were obtained, 100 (8.5%) yielded repeat isolates of an organism and/or serotype from the same child during consecutive C-Fs. Hence there were 1075 C-Fs at risk of first isolation of a potential pathogen. There were 443 (40.6%) C-Fs during which at least one

potential pathogen was isolated for the first time. Among these, 253 (74.9%) were isolates of just one organism; these are reported in the column headed 'sole isolates'. Multiple organisms were isolated during the remaining 190 C-Fs (25.1%); the association between specific potential pathogens and symptoms of gastroenteritis is reported for these organisms in the column headed 'with other organisms' of table B.6.

Overall, there were 49 C-Fs with both reported symptoms and an isolate of at least one potential pathogen; 35 C-Fs with symptoms but no isolate; 370 C-Fs with an isolate but no symptoms; and 595 C-Fs with neither symptoms nor an isolate. There was a highly significant difference between the percentage of symptomatic C-Fs occurring in association with isolates of at least one potential pathogen ($59/443 = 13.3\%$) and the percentage of symptomatic C-Fs ($35/630 = 5.6\%$) occurring in the absence of an isolate (Chi square = 21.2, $p < 0.0001$). A significant association between presence of agents and diarrhoea was of course expected.

The significance of this association was similarly tested for each specific pathogen, using Fisher's exact test because of small expected frequencies, and results are reported in Table B.6.

Significant associations with symptoms were found for *Echovirus*, *Coxsackie B* virus, *Campylobacter*, *Cryptosporidium*, *Shigella sonnei* and *Rotavirus*.

The purpose of presenting symptomatic rates separately for sole isolates and multiple isolates was to explore the possibility of synergy between organisms. Significant associations with symptoms were found for *Astrovirus* and enteropathogenic strains of *E. coli* when they were isolated with other organisms, but not when they were isolated alone.

Table B.7 shows that approximately two fifths of reported cases of

gastroenteritis in nursery school children could be attributed to organisms which were significantly associated in this study. A further two fifths of cases occurred in the absence of an isolate, and the remaining fifth occurred in the presence of organisms where clear evidence of association was lacking.

Considering the 137 symptomatic fortnights in our study, isolation of *Giardia*, *Cryptosporidium* and *S. sonnei* were significantly associated with GP consultation in Fisher's exact tests. Although approximately 50% of those with symptoms and a sole isolate of campylobacter or *Rotavirus* consulted the doctor, this association did not reach significance, presumably due to the small numbers of isolates. The other organisms found to be significantly associated with diarrhoea (*Echovirus* and Coxsackie B. virus) were not associated with GP consultation, probably because they produced mild symptoms.

Evidence of person to person transmission within the nurseries was sought by inspecting the data for occurrence of the same organism and/or serotype on subsequent fortnights. Although particular serotypes do sometimes cluster for short periods of time, this might equally be due to the presence of a serotype sweeping through the local community, with transmission occurring outside the day-nursery. Even if all transmissions were assumed to occur inside the nursery, no major outbreaks were seen; there is thus no compelling evidence for major gastrointestinal outbreaks occurring within the day-nurseries.

THE STATISTICAL PROBLEM OF MEASUREMENT BIAS

Studies of this type have common problems of compliance and pathogen test sensitivity, which lead them to underestimate infection rates. In this study, the probability r of a stool sample for a given fortnight being

present is only 48%, but rises to 85.5% when symptoms are present.

When samples are provided, there may also be false negative results, due to the imperfect sensitivity of the tests. It was possible to estimate this sensitivity s from the data, and thus to calculate the probability of successfully detecting a positive stool as $p = rs = 0.293$, but as 0.522 when symptoms were present. If an infectious period is so short that there is an opportunity to collect only a single diarrhoeal sample, the probability of detecting it is p , so that the true rate $t = o/p$, where o is the observed rate. Thus asymptomatic infection rates for many pathogens should be scaled up by a factor of roughly 3.4, and symptomatic infection rates by 1.92. There is more than one opportunity to detect a long-lasting infection such as *Giardia*, and so the scaling factor required is not so great.

Turning to percentages of symptomatic infections, consider the definition

$$\text{estimated proportion symptomatic} = \frac{\text{no. of symptomatic infections}}{\text{total no. of infectious episodes}}.$$

For agents with short infectious periods, failure to detect the agent will not bias the estimated proportion of symptomatic infections, because both numerator and denominator are scaled down equally. Non-compliance scales the denominator down more than the numerator, and so increases the estimate of the proportion symptomatic. However, for agents such as *Giardia* with long infectious periods, we have mentioned that detection of an infectious episode is still likely, because of the several opportunities of detection. Low detection sensitivity thus scales down the numerator more than the denominator, giving spuriously low estimates of the proportion of symptomatic infections.

This may explain why *Giardia intestinalis*, with its long excretion period, is not seen to associate more strongly with symptoms.

Table B.8 shows infection rates and percentages of infection that were symptomatic, after corrections based on the logic given here were applied. The crude correction given here does not greatly increase the association of *Giardia* with symptoms.

The method of estimating sensitivity from data was to look for child-fortnights where infection was known to be present, because a particular serotype had been detected for some previous and subsequent fortnights (i.e. leading and trailing positives). The intermediate fortnights then constitute independent Bernoulli trials, and the sensitivity is estimated as the proportion of such trials where a positive sample was obtained. Pooling data for similar pathogens, a figure of $105/172 = 0.61$ was obtained, which did not differ significantly between agents. This is our estimate of s . The figure for *Giardia* alone was also 0.61.

Since the probability of detecting an agent is 0.52 when symptoms are present, we would expect a little under 48% of observed symptom episodes to have no agent associated (under 48%, because episodes can last for more than one fortnight, with more opportunities to sample). The calculated non-detection probability is then 39.8%, very close to the observed figure of 37.2%.

This statistical consideration of bias shows us that our results are consistent with all or nearly all episodes of illness having an associated agent.

DISCUSSION

The number of GP consultations in the UK in 1986 with ‘intestinal diseases presumed to be infective, of either unspecified viral or unknown origin’ was 189.6 per 1000 person-years at risk [9]. The 35 consultations for gastroenteritis in our study yielded a rate of 372 (95% CI 260 to 516)

which is significantly higher. This gives a picture of our population as having slightly higher than average risk of gastrointestinal illness; this was, after all, one reason why such a population was studied.

In a meta-analysis of third-world data, Bern [3] found a diarrhoea incidence of 2.6 episodes per person-year among children 4 years old and younger. This is significantly higher than our estimate of 1.43 (95% CI 1.19 to 1.43). Thus as one might expect, even a high-risk UK population falls short of third-world incidence of gastrointestinal illness. Further, episodes of diarrhoea seen in our study were much shorter than those in the third world. Morris *et al* ([11]) found that the percentage of time during which diarrhoea occurs in children aged 0–5 in Ghana to be of the order of 20%, with diarrhoea persisting for more than 40% of the time for a large majority of children. Our study showed the median percentage of person-time during which diarrhoea occurs (the longitudinal prevalence) to be 6.7%, and no children had diarrhoea more than 40% of the time.

The main pathogens causing gastrointestinal illness in these day-nursery children are seen from table B.8 to be *Echovirus*, *Coxsackie B* virus, *Cryptosporidium*, *Campylobacter*, *Shigella sonnei* and *Rotavirus*.

Not surprisingly, the attributable fractions of infections found in this community due to the various agents differ from that found by the same public health laboratory for hospitalised patients ([5]) done a few years before this study. Ellis *et al* [5] found that 34% of patients were excreting *Rotavirus*. There are many other studies of hospitalised children yielding similar results.

The explanation could well be that infection with *Rotavirus* often leads to hospital admission, because of dehydration caused by the associated symptoms of vomiting and diarrhoea. Thus a large proportion of

hospital cases of diarrhoea are infected with *Rotavirus*. However, in the community we also see diarrhoeal cases with milder symptoms that do not require hospitalisation, caused by feebler agents such as *Echovirus*, and the number of such cases far exceeds those due to *Rotavirus*.

The near-absence of *Rotavirus* and the complete absence of *Salmonella* is however still a little puzzling. The study spanned two calendar years, giving opportunity for the agents to appear. The lack of *Rotavirus* was noticed early in the study, so that procedures could be checked. The result is thus clearly not due to experimental error.

Velazquez *et al* [16] discuss the effect of early *Rotavirus* infection on later infection and its symptoms. Infection and gastroenteritis are both reduced after the first infection. However, such a protective effect following very early *Rotavirus* infection would still by no means account for the low rate of gastroenteritis due to *Rotavirus*, or for the small number of isolations of the agent from stool. Statistical fluctuation in the prevalence of agents from year to year is probably the explanation.

There are many studies with which our results on asymptomatic excretion can be compared. They may be classified into: cohort studies such as this, in which a cohort of children are followed through time; prospective clinical studies in which children with diarrhoea, usually hospitalised, are studied over a period; and outbreak studies, in which samples are taken during an outbreak of one pathogen.

Information on the percentage of symptomatic infections is however quite sparse. In cohort studies, samples are typically not taken regularly as in this study, but only when episodes of gastroenteritis occur. In [1], an additional sample was taken only once, at the start of the study. Such studies therefore yield no information about the extent of asymptomatic carriage.

In prospective studies of hospital populations, often percentages of diarrhoeal and control groups infected are quoted. A non-zero percentage of infected non-diarrhoeal children shows that asymptomatic infection is possible, but such data do not yield an estimate of the percentage of asymptomatic carriage without making further assumptions. There are very many papers that support significant asymptomatic carriage for *Campylobacter* and *Cryptosporidium* in this way [14, 12, 7].

In outbreak studies, the whole population is sampled, and an estimate of symptomatic carriage can be obtained for the pathogen responsible. These studies also relate to the community, rather than to hospitalised cases, and so are comparable with ours. One might however expect that an abnormally high proportion of cases would be symptomatic during an outbreak, because of the publication bias associated with large numbers of symptomatic cases.

In conclusion, the limited data available are in broad agreement with our results on asymptomatic transmission.

A striking result from Table B.8 is that even the organisms most strongly associated with gastroenteritis (*Campylobacter* and *Cryptosporidium*) were asymptomatic during approximately 50% of the C-Fs of their first isolation.

There were other organisms such as SRSV and *Calicivirus* that were not significantly associated with symptoms, but which were nonetheless detected. Numbers of such cases detected are undoubtedly an underestimate, so that these asymptomatic infections may constitute an important reservoir of infection within the community.

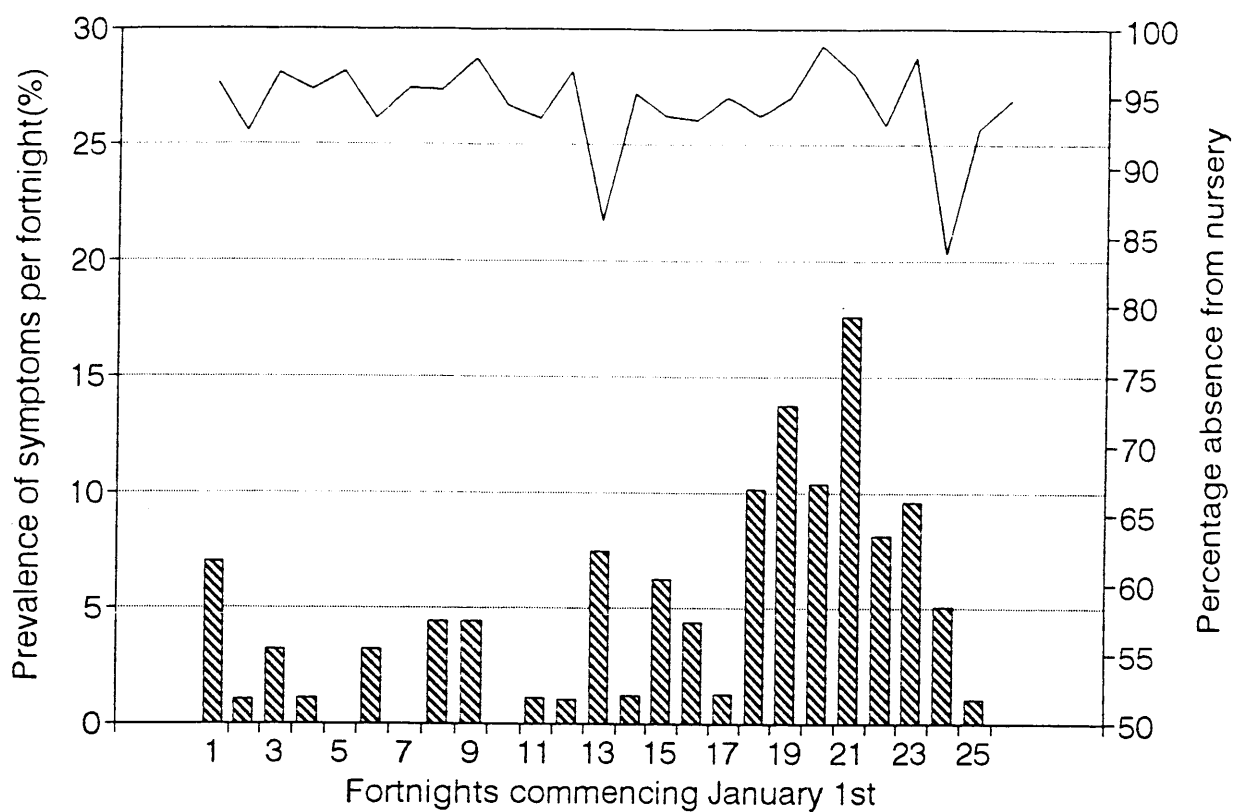


Figure 1 Seasonal incidence of gastrointestinal illness and absence from the nursery.

Acknowledgments

Mariam Al-Mannai would like to thank The University of Bahrain, Bahrain, for sponsoring her to do a Ph.D. at the University of Salford.

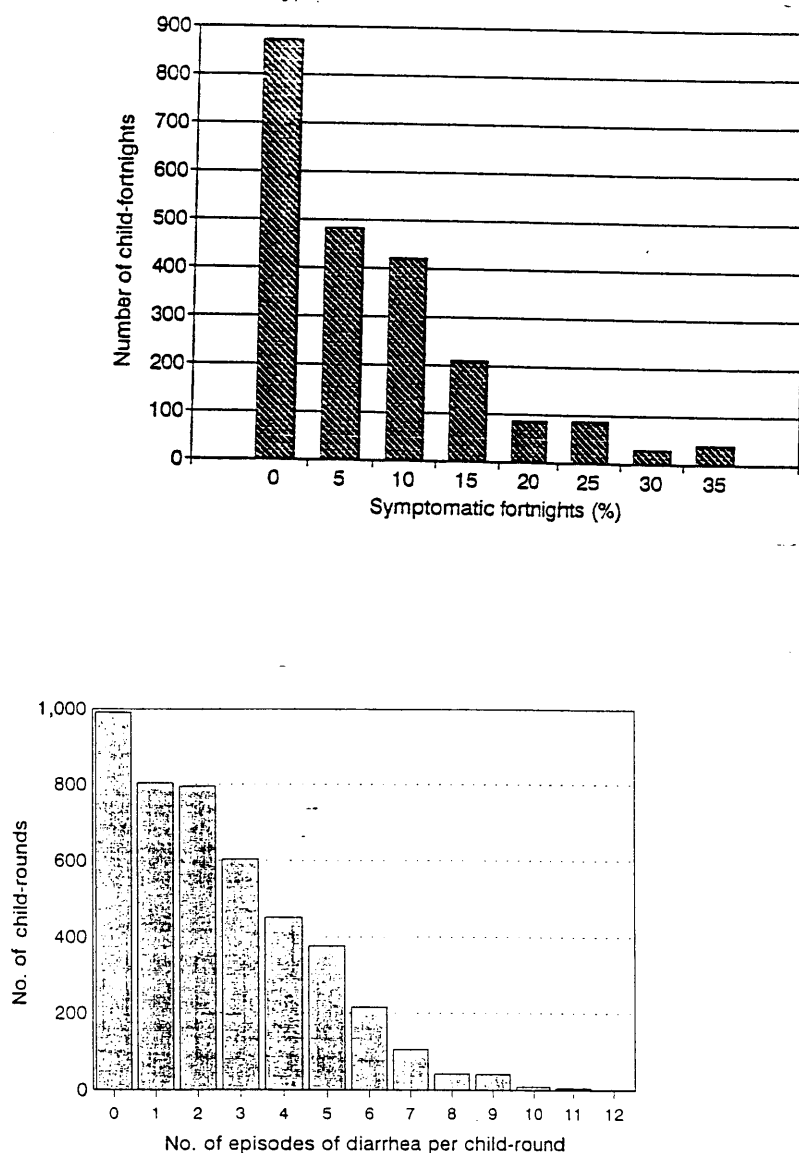


Figure 2 duration of gastrointestinal illness in our study, with superimposed curve from [11].

Bibliography

- [1] Araya, M., Figueroa, G., Espinoza, J., Xarur, X. and Brunser, O. Acute diarrhoea and asymptomatic infection in Chilean preschoolers of low and high socio-economic strata *Acta Paediatr Scand* 1986, **75**, 645–651.
- [2] Baqui A. H., Black R. E., Yunus M. D., Hoque A. R. A., Chowdhury H. R., Sack R. B., Methodological issues in diarrheal disease epidemiology: definition of diarrhoeal episodes. *Int J Epidemiol* 1991;**20**: 1057-63.
- [3] Bern, C., Martines, J. de Zoysa, I., Glass, R. I. The magnitude of the global problem of diarrhoeal disease: a ten-year update, Bulletin of the World health Organization, 1992, **70**, 705–714.
- [4] Boerma J. T., Black R. E., Sommerfler A. E., Rutstein S. O., Bicego G. T., Accuracy and completeness of mother's recall of diarrhoea occurrence in pre-school children in demographic and health surveys. *Int J Epidemiol* 1991;**20**: 1073-80.
- [5] Ellis M. E., Waston B., Mandal B. K., Dunbar E. M., Craske J., Curry A., Microorganisms in gastroenteritis. *Arch Dis Child* 1984, **59**, 848-55.

- [6] Glass R. I., Lew J. F., Gangarosa R. E., LeBaron C. W., Ho-M. S. Estimates of morbidity and mortality rates for diarrheal diseases in American children. *J Pediatr* 1991 Apr; 118 (4 (Pt 2)): S27-33.
- [7] Lindblom G. B., Ahren C., Chagalucha J., Gabone R., Kiajser B., Nilsson L. A., Sjogren E., Campylobacter jejuni/coli and enterotoxigenic Escherichia coli (ETEC) in faeces from children and adults in Tanzania. *Scand J Infect Dis* 1995, **27** (6), 589-93.
- [8] McCormick, A. and Rosenbaum, M. Morbidity statistics for general practice 1981-1982. Third national study: socio-economic analyses. OPCS series MB5, no. 1, HMSO, London.
- [9] McCormick, A. and Rosenbaum, M. Morbidity statistics for general practice 1981-1982. Third national study: socio-economic analyses. OPCS series MB5, no. 2, HMSO, London.
- [10] Morris, J. (Editors) Social Trends, **17**, 1987, HMSO, London.
- [11] Morris, S. S., Cousens, S. N., Kirkwood, B. R., Arthur, P., Ross, D. A. Is prevalence of diarrhea a better predictor of subsequent mortality and weight gain than diarrhea incidence?, American Journal of Epidemiology, 1996, **144**, 582-588.
- [12] Nath G., Shukla B. N., Reddy D. C., Sanyal S. C., A community study on the aetiology of childhood diarrhoea with special reference to Campylobacter jejuni in a semiurban slum of Varanasi, India. *J Diarrhoeal Dis Res* 1993 Sep, **11**(3), 165-8.
- [13] Norusis, M. J. *The SPSS Guide to Data Analysis for SPSS/PC+* 1988, SPSS inc, USA.

- [14] Pazzaglia G., Bourgeois A. L., El-Diwany K., Nour N., Badran N., Hablas R., *Campylobacter* diarrhoea and an association of recent disease with asymptomatic shedding in Egyptian Children. *Epidemiol-Infect* 1991 Feb, **106**(1), 77-82.
- [15] Ramprakash, D. and Morris, J. (Editors) *Social Trends*, **14**, 1984, HMSO, London.
- [16] Velazquez, F. R., Matson, D., Calva, J. J., Guerrero, M. L., Morrow, A. L., Carter-Campbell, S., Glass, R. I., Estes, M. K., Pickering, L. K., Ruiz-Palacios, M. Rotavirus infection in infants as protection against subsequent infections, *New England Journal of Medicine*, 1996, **335**, 1022–1028.
- [17] Wood, D. J. and Bailey, A. S. Detection of adenovirus types 40 and 41 in stool specimens by Immune Electron Microscopy. *Journal of Medical Virology* 1987, **21**, 191-199.

Age	< 1	1	2	3	4	Tot
M	5 (4.7)	28 (26.7)	36 (34.3)	28 (26.6)	8 (7.6)	105 (64.0)
F	2 (3.4)	23 (39.0)	20 (33.9)	10 (16.9)	4 (6.8)	59 (36.0)
Tot	7 (4.2)	51 (31.1)	56 (34.1)	38 (23.2)	12 (7.3)	164 (100.0)

Table B.1: The study population. Age at entry to the study, years (%). Records of ages of 2 males and 1 female are missing.

Siblings	0	1	2	3	4	5	6	Tot
Number of subjects	52	61	32	7	9	5	1	167
Percentage	(31.1)	(36.5)	(19.2)	(4.2)	(5.4)	(3.0)	(0.6)	(100)

Table B.2: The study population. Percentage of siblings.

	Number employed			
Children live with:	0	1	2	Total
Mother alone	27	38	-	65
Mother and grandparents	3	8	-	11
Father alone	3	0	-	3
Mother and father	37	29	16	82
Others	2	4	-	6
Total	72	79	16	167

Table B.3: The study population. Single parent and employment status.

	Aged 2 years or less		More than 2 years		Total
	Male	Female	Male	Female	
Fortnights with reported G-E	39	33	35	29	136
Fortnights enrolled	402	442	1077	544	2465
Eepisodes of G-E per child-year	2.52	1.94	0.84	1.39	1.43
95% confidence interval	1.73 to 3.31	1.28 to 2.60	0.78 to 1.08	0.88 to 1.89	1.19 to 1.68

Table B.4: Incidence density of reported gastroenteritis.

	Aged 2 years or less		More than 2 years		Total
	Male	Female	Male	Female	
Fortnights enrolled	402 (16.3%)	442 (17.9%)	1077 (43.7%)	544 (22.1%)	2465 (100%)
Fortnights attended	359 (15.4%)	417 (17.9%)	1039 (44.5%)	518 (22.2%)	2333 (100%)
Faecal specimens collected	207 (17.6%)	250 (21.3%)	454 (38.6%)	264 (22.5%)	1175 (100%)
Attendance	89.3%	94.3%	96.5%	95.2%	94.6%
Specimen collection	57.6%	60.0%	43.7%	51.0%	50.36%

Table B.5: Enrolment, attendance and specimen collection.

Organism	SOLE ISOLATE			WITH OTHER ORGANISMS			TOTALS		
	Symp	No Symp	%	Symp	No Symp	%	Symp	No Symp	%
Adenovirus	0	4	0.0	4	8	33.3	4	12	25.0
Astro virus	0	4	0.0	3	1	75.0*	3	5	37.5
Calici virus	1	4	20.0	0	4	0.0	1	8	11.1
Corona virus	0	5	0.0	2	5	28.6	2	10	16.7
Reovirus	1	5	16.7	1	2	33.3	2	7	22.2
Rota virus	1	1	50.0	1	0	100.0	2	1	66.7*
SRSV	0	1	0.0	5	6	45.5	5	7	41.7
Adenovirus (resp)	10	65	13.3	3	30	9.1	13	95	12.0
Coxsackie B	8	33	19.5*	10	23	30.3*	18	56	24.3*
Echovirus	9	35	20.5*	5	23	17.9	14	73	16.1*
Campylo	5	6	45.5*	0	4	0.0*	5	10	33.3*
Cl difficile	0	0	0.0	0	1	0.0	0	1	0.0
E coli	3	34	8.1	7	24	23.6*	10	58	14.7*
S. sonnei	1	0	100.0	1	0	100.0*	2	0	100.0*
Yersinia	0	0	0.0	1	0	100.0	1	0	100.0
Crypto	5	4	55.5*	5	5	50.0*	10	9	52.6*
Giardia	3	6	33.3	1	5	16.7	4	11	26.7
Total	46	207	18.2*	49	141	25.8	95	348	38.5

Table B.6: Association between symptoms and first isolation of specific organisms during the same fortnight. * denotes $p < 0.05$, Fisher's Exact Test. Note absence of Salmonella.

Agent	Number(%)
Echovirus	11 (11.7)
Coxsackie B virus	10 (10.6)
Cryptosporidium	9 (9.6)
Campylobacter	6 (6.4)
Shigella sonnei	3 (3.2)
Rota virus	2 (2.1)
Non-significant organism	18 (19.1)
No organism isolated	35 (37.2)
Total	94 (100.0)

Table B.7: Proportion of first symptom episodes attributable to significantly associated organisms.

Agent	Infection Rate	percent symptomatic	symptomatic rate
Echovirus	1.76	10.4	0.183
Coxsackie B virus	0.93	13.1	0.122
Cryptosporidium	0.19	54	0.103
Campylobacter	0.3	27	0.081
Shigella sonnei	0.02	100	0.02
Rota virus	0.09	21.7	0.0195
Giardia	0.25	8.2	0.021

Table B.8: Infection rates as number per person-year, percentages symptomatic, and symptomatic infection rates as number per person year after a crude bias correction.

Agent	Detection method
Adenovirus (untyped)	EM
Astrovirus	EM
Calicivirus	EM
Coronavirus	EM
Reovirus	EM
Rotavirus	EM
SRSV	EM
Adenovirus (respiratory)	EM + cell culture
Coxsackie A and B	EM + cell culture
Echovirus	EM + cell culture
Campylobacter	bacterial culture
Escherichia Coli	bacterial culture
Salmonella	bacterial culture
Shigella	bacterial culture
Yersinia pestis	bacterial culture
Cryptosporidium	light microscopy
Giardia intestinalis	light microscopy

Table B.9: Agents tested for and their detection methods.

	Aged 2 years or less		More than 2 years	
	Male	Female	Male	Female
Male, at least 2 years of age	-	ns	< 0.0001	< 0.01
Female, at least 2 years of age	-	-	< 0.0001	ns
Male, more than 2 years of age	-	-	-	< 0.01

Table B.10: Significance tests for difference between age/sex group incidence density of Gastroenteritis no. Chi square tests for difference in proportions.

Appendix C

Literature Review

Here, we look at some of the published literature about the use of logistic regression in modelling childhood diarrhoea. We also discuss some of the methods used in estimating the basic reproductive rate, R_0 , and the use of the Weibull model in survival analysis.

C.1 Logistic Regression

The use of logistic regression models in determining the risk factors of childhood diarrhoea has been established by many authors such as Van den Broeck *et al.* (1996), Bhutta *et al.* (1997), Bhutta (1996), Booth *et al.* (1996), Borgnolo *et al.* (1996), Bhutta *et al.* (1996), Tshikuka *et al.* (1996), Brussow *et al.* (1995) to mention a few. Most of these studies are case-control. Unfortunately most of these studies were conducted in the underdeveloped countries of the world where poverty and illiteracy are dominant factors. There are no grounds to compare rural parts of countries like the Philippines, Bangladesh, Pakistan and Peru with a city of the developed world like Manchester, even though the study population here was chosen from a very deprived area and the children

were reported to be needing special care. Hence only three articles will be discussed here.

Gorter *et al.* (1991), performed a case control study of the risk factors for child diarrhoeal diseases related to water sanitation in a rural area of Nicaragua. The study population consisted of 1229 children under 5 years of age who were reported to be suffering from diarrhoea and were matched with the same number of children of the same age present with other illnesses. Several factors were considered here none of which were considered earlier in this chapter. Logistic regression modelling was used to determine the relation between the frequent childhood diarrhoea and mother's level of schooling, availability of a latrine in the child's house, the number of children under five who lived with the child in the same house and the distance between the child's house and the nearest source of water. This study revealed that the children who lived further than 500 meters of the nearest water supply were at higher risk of inducing diarrhoea. Mother's level of education had negative correlation with the frequency of diarrhoea in her children. These factors, of course were not considered in our study because water was available in each child's accommodation and mother's level of education was not available for us.

Another case-control study was conducted by Knight *et al.* (1992) to determine the risk factors for the transmission of diarrhoea in a rural district of the poorest state of Malaysia. The study population consisted of 170 children between 4 and 59 months of age who were reported to have diarrhoea matched with the same number of children of the same age and sex who were reported to have other illnesses not related to diarrhoea. Several social factors were

considered. Some of these factors are used in this study, such as child's birth order and the number of people the child lived with in the same house. Other factors like the presence of animals in the house, mother's ethnic group and guardian's educational level were also considered. Logistic regression was used to determine the relation between these and other factors and frequent childhood diarrhoea. It was found that the only highly correlated factor with the frequent occurrence of diarrhoea in these children was the child's birth order, where in the Manchester data birth order was not significantly related with the high frequency of the occurrence of diarrhoea.

Gupta *et al.* (1996) conducted another study in 3 villages in India. The purpose of the study was to determine the clinico-epidemiologic aspects of mucoid diarrhoea. Three different types of diarrhoea were reported, mucoid diarrhoea, dysentery diarrhoea and watery diarrhoea. Five hundred and fifty three children under the age of 4 years were observed for a follow up period of twenty nine months. The average number of episodes of diarrhoea was 1.7 per child per year. Multivariate logistic regression was used to compare the three different types of diarrhoea. The results were that mucoid and dysentery diarrhoeas were similar in both clinical and laboratory findings. The clinical findings were the association between these types of diarrhoea and dehydration, abdominal pain, vomiting and fever, where the laboratory findings were the results of laboratory analysis of faecal specimens. The latter were *Salmonella*, *Shigella* and *E. histolytica*. In the Manchester study, *Salmonella* and *E. histolytica* were not found in the laboratory analysis, and all the clinical findings were summed

up together as gastro intestinal diseases.

C.2 The basic reproductive rate R_0

There are few different methods for estimating the basic reproductive rate R_0 for certain infections. This section will look at some of these methods. Anderson and May's book, Anderson and May (1991), is considered as one of the most important references in mathematical epidemiology.

Bradley *et al.* (1992), to determine the distribution pattern of infection in a rural community in Africa where hookworm infection was an endemic. The study was conducted in Zimbabwe. Stool specimens were obtained from 519 individuals in the community and examined to identify infected individuals to estimate prevalence. The resulting data were grouped into 4 broad age group and a cohort equally representing the 4 age groups and two sex groups was randomly selected to participate in the subsequent expulsion programme. A crude estimate of R_0 was obtained by estimating the parameters of the model

$$M(a) = R_0 M^* [1 + (1 - z) M^* / k]^{(k-1)} (1 - e^{-\mu})$$

where $M(a)$ is the mean worm burden at age a , M^* is the average worm burden, k negative binomial distribution parameter, z inverse measure of the strength of the density dependence and μ death rate of adult parasite.

Under the assumption that infection intensity rises rapidly with age such that M^* (average worm burden) is approximately equal to the equilibrium worm burden, and the number of worms was

distributed according to the negative binomial with constant k and mean as defined in the model's equation. By fitting observed age intensity to the model an estimate of R_0 of 2 was calculated.

Gupta *et al.* (1995) showed that R_0 could not be estimated for agents that do not produce lifelong immunity. One example of these infections was *Plasmodium falciparum* malaria. Using age structured mathematical models, the authors explored the effects of a delay in the development of immunity on the age exposure profile of an endemic infectious agent with a given host population. As the number of infections in a community can only grow if every infected individual can infect at least one other person, the basic reproductive rate R_0 , was estimated using the equation

$$R_0 = m\alpha^2 b L_m D c$$

where m is the number of mosquitoes per human host, α is mosquito biting rate, b is the probability that an infected mosquito will produce and infect a host, L_m is the average life span of mosquitoes, D is the average duration of infectiousness in the human host and c is the probability an infectious feed will lead to sporozoite production in the vector.

For the infections that produce lifelong immunity, the suggested formula for estimating R_0 is

$$R_0 = L/A$$

where L is the average life span of the host and A is the average age at which an individual typically acquire an infection.

The models described in this article can be applied to any infection that does not produce life-long protective immune response such as the common sexually transmitted diseases.

Brunham *et al.* (1994) estimated the basic reproductive rate for two sexually transmitted diseases, *Meisseria gonorrhoeae* and *Chlamydia trachomatis*, by two different methods. One of these methods used a mathematical model of transmission dynamics and depends on estimating three variables. The formula used was of the form

$$R_0 = \beta C D$$

where β is the probability of infection transmission per partnership, C the effective rate of partner change and D the duration of infectivity. This model does not assume any sexual mixing partners.

The second method was based on estimating R_0 by a mathematical model relating it to the prevalence of infection the formula used was

$$\pi = (1 - \frac{1}{R_0})\mu$$

where π is the probability that the sex's partner is infected and μ is the ratio of the mean rate of partner change among individuals with sexually transmitted diseases.

These two models were applied to sexual behaviour data from Nairobi, Kenya. The results of the first equation was $R_0 = 2$ for *N. gonorrhoeae* and $R_0 = 3.6$ for *C. trachomatis*. The second formula yielded R_0 of 1.8 for *N. gonorrhoeae* and 1.43 for *C. trachomatis*.

The first method of estimating R_0 does not depend on the assumption of absence of immunity. Therefore R_0 estimated using this equation should be more reliable.

Gupta *et al.* (1994a) suggested that

$$R_0 = L/A$$

for any infection if it produced lifelong immunity, where L is the average life-span of the host and A is the average age at infection. This was also suggested by Anderson and May (1991). A value between 50 and 100 for *falciparum malaria* was suggested. Such high value was given because human's first encounter with the parasite is before the age of one year. If a parasite has more than one serotype, then the average age of the first infection may be approximated by

$$A_n = \frac{L}{\sum R_{0i}}$$

where R_{0i} is the basic reproductive rate of the i th strain of the parasite and L is as defined above.

These models were tested on data from a highly malairal area of Papua New Guinea. It was shown that a low A with *P. falciparum* did not reflect high transmissibility of the parasite but may be an indication of the circulation of many distinct antigenic strains of the parasite each with moderate to low transmissibility. In such circumstances, immunity to malaria may appear short lived when in reality immunity to a specific strain is lifelong and reinfection occurs as exposure to different strains.

Gupta *et al.* (1994b) demonstrated that simple models of strain heterogeneity could generate periodic behaviour as a consequence of the interaction between parasite strains and host immunological defences. As the degree of cross-immunity between strains increased, the system moves from a regime of independent strain

transmission and coexistence, through increasingly coupled behaviour, to the displacement of the strain of lower transmissibility by the strain with a higher basic reproductive rate.

In this study, a simple homogeneous model for malaria was extended to encompass two distinct parasite strains. The Ross-Macdonald model was extended to discriminate the states of infectiousness and infection in the human host. This model follows the following set of differential equations:

$$dx/dt = amby(1 - x) - (h + \mu)x$$

$$dx_1/dt = amby(1 - x) - sx_1$$

$$dy/dt = ax_1(1 - y) - \mu_M y$$

where

x is the proportion of immune human population,

x_1 proportion of infectious human population,

y population of infectious vectors,

μ natural death rate of human host,

m number of female mosquitoes per human host,

$\frac{1}{n}$ average duration of immunity,

$\frac{1}{s}$ average duration of infection,

a the rate at which the mosquito bites the host,

b proportion of infected bites on the host that produce an infection

and

μ_M vector average mortality rate.

This model gave a definition for the basic reproductive rate of

$$R_0 = a^2 mb / s \mu_M$$

Another model was developed in the case where there were two strains of the parasite. In this model the host population was divided into three different groups, immune to the first strain x_1 , immune to the second strain x_2 and immune to both x_3 . y_1 and y_2 are the number of vectors carrying serotypes 1 and 2 respectively. The model had the form

$$dx_1/dt = \lambda_1(1 - \sum x_i) - (c\lambda_2 + \mu)x_1$$

$$dx_2/dt = \lambda_2(1 - \sum x_i) - (c\lambda_1 + \mu)x_2$$

$$dx_3/dt = c(\lambda_1x_2 + \lambda_2x_1) - \mu x_3$$

$$dx_{11}/dt = \lambda_1(1 - \sum x_i + cx_2) - sx_{11}$$

$$dx_{12}/dt = \lambda_2(1 - \sum x_i + cx_1) - sx_{12}$$

$$dy_1/dt = ax_{11}(1 - \sum y_i) - \mu_M y_1$$

$$dy_2/dt = ax_{12}(1 - \sum y_i) - \mu_M y_2$$

where $\lambda_i = mab_i y_i$, and

c is the degree of cross protection where $c = 1$ means no cross protection and $c = 0$ means total cross protection.

Then the basic reproductive rate of the i th strain R_0^i can be estimated by

$$R_0^i = a^2 m b_i / s \mu_M.$$

These two models were tested on data from Papua New Guinea.

Thein-Hlaing *et al.* (1991), in a study conducted in Myanmar (Burma) estimated the basic reproductive rate for *Ascaris lumbricoides*. This study was conducted on children under 15 years of age in two villages and between 2 and 21 years of age in a third village. The purpose of this study was to determine the impact of

3 month chemotherapeutic treatments against *Ascaris lumbricoides* as opposed to longer periods of treatment. R_0 was estimated to assess the choice of the mathematical model for the treatment against the parasite. The estimated R_0 for the age groups in this study was not found to be different from that estimated in other studies under the same social and economical conditions (Thein-Hlaing *et al.* (1984)).

Woolhouse *et al.* (1991) considered the effects of heterogeneous rates of contact with infective water bodies on the basic reproductive rate of human *Schistosomes*. It was found that an individual's contribution to R_0 depended on both the net contact rate and on how these contacts were distributed among water sites.

A theoretical analysis of the epidemiology of human *Schistosome* infection was developed. This analysis was based on a mathematical model that emphasised the relation between heterogeneities in water contact rates and the dynamics of the *Schistosome* transmission. The dynamics of the parasite transmission was represented by the model

$$\frac{dM_i}{dt} = \alpha \left(\sum_j \eta_{ij} N y_j \right) - \gamma M_i$$

$$\frac{dy_i}{dt} = \beta \left(\sum_j \eta_{ij} M_j \right) (1 - y_i) - \mu y_i$$

$i = 1, 2, 3, \dots, p$ and $j = 1, 2, 3, \dots, l$

where

α was the number of parasites acquired per infected snail,

β average infection of snails per contact per *Schistosome*,

γ per capita mortality rate of *Schistosome*,

μ the average rate of infected snails,

$\mathbf{P}$ a vector of p rows of human population size,

$\mathbf{L}$ a vector of l rows of number of water sites,

M_i the number of parasites found in the i th individual,

y_j the fraction of infected snails in the j th water site and

η_{ij} the rate of water contact of individual i and water site j .

This model assumed no exact correlation between individual contributions to man-snail and snail-man transmission rates. It produced an approximation for R_0 namely

$$R_0 = \left(\frac{\alpha N \beta}{\gamma \mu} \right) \epsilon$$

where ϵ was the dominant eigenvalue of the matrix LXL .

This analysis was applied to field data recording the water contact behaviour of a human community in Zimbabwe. It was shown that heterogeneities in contact rate increased the expected R_0 by a factor of 4 to 5 over the value expected given homogeneous contact rates.

C.3 Survival analysis and Weibull distribution

The use of Weibull distribution in survival analysis has been implemented not only in the fields of medicine and epidemiology, but in many other fields such as engineering and economics. This section will look at some of these applications with more emphasis on the medical area.

Mackisack and Stillman (1996) used the Weibull distribution to predict the characteristic-life of wooden power poles treated with copper chrome arsenate. The failure time in this study was time

until the pole was changed due to degradation and not to accident. The Weibull distribution parameters were estimated using maximum likelihood.

Narendranathan and Stewart (1993) used survival analysis in modelling unemployment durations. In this study, the Weibull distribution was used and the failure time was until the subject left the unemployment stage. It was found the Weibull model did not give satisfactory representation of the base-line hazard.

Smales *et al.* (1993) applied survival analysis in dentistry. Their study was to compare the predictive accuracy of the standard Weibull distribution model with that of a mixture model for resin-bonded etched-metal bridge survival. The failure time was time until the bridge failed to function. Data were collected from examining patients on the survival of 189 such bridges placed between 1983 and 1988. Data records were updated in 1992. There were two types of bridges: anterior and posterior made from seven different metal alloys. The predictive estimates for the standard Weibull and mixture models were found to be different. The predictive accuracy of the mixture model was much better than that of the standard Weibull.

Chiarotti *et al.* (1992) conducted a study to estimate seroconversion time based on different parametric methods and to assess their influence on the estimation of the incubation time between HIV infection and the onset of AIDS on 712 HIV-positive Italian haemophiliacs. Seroconversion time was estimated using the mid-point of each seroconversion interval. Parametric Weibull distribution and Kaplan-Meier non-parametric survival analysis were

conducted. The failure time here was onset of AIDS. The study suggested that different estimates of seroconversion times do not modify the observed shape of the distribution of progression to AIDS and that they simply result in a shift in time since seroconversion to reach a given cumulative incidence.

Chiarotti *et al.* (1994) conducted a study to determine the median time between the infection of HIV and the onset of AIDS. The study subjects were Italian haemophiliacs. Three different parametric models were used. these model were Weibull distribution, log-logistic and generalised exponential. Survival analysis was performed and the failure time was the onset of symptoms of AIDS. These models could represent three different situations with respect to the hazards of developing AIDS following seroconversion. The hazard under the Weibull distribution increased monotonically without bounds where in the other model it increased to an asymptotes in the generalised exponential and to a maximum then started decreasing in the log-logistic.

Hauk *et al.* (1997) applied two part survival model to administrative data for determining rate and predictors for maternal-child transmission of HIV. A model that combined a logistic regression for the transmission of HIV and a Weibull regression model for the survival analysis was used. The failure time was time to diagnosis of AIDS. The results found here were helpful in indicating predictors of transmission and in estimating transmission rates with reduced concern for adequacy of follow-up.

Willems, Cales and Verbeke (1997) fitted a Weibull model and used survival analysis to evaluate the *in vitro* peel/shear bond strength

of a selection of orthodontic bracket adhesives to human premolar teeth. The failure time here was scored according to the adhesive remnant index. Twenty-two different types of bracket adhesives on intact human premolars were used to evaluate the *in vitro* peel/shear bond strength. The site of bond failure was also investigated. This study determined which was the best material to use in theory but also suggested this material to be tried clinically.

Feuer *et al.* developed three statistical models to study the impact that two clinical trial had on survival in two types of cancers. Hodgkin's disease and testicular cancer. These models were segmented regression in conjunction with Cox semi-parametric proportional hazard model, the two-parameter (shape and scale) Weibull distribution and the exponential cure model. These models allowed to determine when survivals began to improve. The Weibull model had a steadily decreasing hazard when the shape parameter was less than one was found to be more appropriate for Hodgkin's disease and the cure model was found more appropriate for the other type of cancer. The advantages of the Weibull model and cure model is that they were parameterised to be applied in testing about how survival curves change as new treatments became available.

Appendix D

References

- Adegbola RA, Demba E, DeVeer G and Todd J (1994), Cryptosporidium infection in Gambian children less than 5 years of age. *Journal of Tropical Medicine and Hygiene*, Vol 97, No 2, pp. 103-107.
- Allen AVH and Ridley DS (1970), Further observations on the formol-ether concentration method for faecal parasites, *Journal of Clinical Pathology*, Vol 23, pp. 545-546.
- Anderson RM and May RM (1991), *Infectious Diseases of Humans, Dynamics and Control*, Oxford University Press, Oxford.
- Araya M, Baiocchi N, Espinoza J and Brunser O (1991), Persistent diarrhoea in the community: Characteristics and risk factors. *Acta Paediatrica Scandinavia*, Vol 80, No 2, pp. 181-189.
- Bagdonavicius VB and Nikulin MS (1997), Transfer functionals and semiparametric regression models, *Biometrika*, Vol 84, No 2, pp. 365-378.
- Baker RD and Stevens RH (1995), A Random Effect Model for Ana-

- lysis of Infectious Disease Final-State Data, *Biometrics*, Vol 51, No 3, pp. 965-968.
- Ball F (1986), A Unified Approach to the Distribution of Total Size and Total Area under the trajectory of Infectives in Epidemic Models, *Advances in Probability*, Vol 18, pp. 289-310.
- Baqui AH, Black RE, Yunus MD, Hoque ARA, Chowdhury HR, and Sack RB (1991), Methodological Issues in Diarrhoeal Disease Epidemiology: Definition of Diarrhoeal Episodes, *International Journal of Epidemiology*, Vol 20, pp. 1057-1063.
- Becker NG (1982), *Analysis of Infectious Diseases Data*, Chapman and Hall, New York.
- Belding DC (1965), *Text Book of Parasitology*, 3rd edition, Appleton Century-Crofts, New York.
- Bhutta ZA (1996), Impact of age and drug resistance on mortality in typhoid-fever, *Archives of Disease in Childhood*, Vol 75, No 3, pp. 214-217.
- Bhutta ZA, Punjwani N and Lindblad BS (1996), Concomitant Bacteremia as a risk factor for diarrheal disease mortality in Karachi: A case- control study of hospitalized children, *Acta Paediatrica*, Vol 85, No 7, pp. 809-813.
- Bhutta ZA, Nizami SQ, Thobani S and Issani Z (1997), Factors determining recovery during nutritional therapy of persistent diarrhoea: the impact of diarrhoea severity and intercurrent infections, *Acta Paediatrica*, Vol 86, No 8, pp. 796-802.

- Boerma JT, Black RE, Sommerfler AE, Rutstein SO and Bicego GT (1991) Accuracy and completeness of mother's recall of diarrhoea occurrence in pre-school children in demographic and health surveys, *International Journal of Epidemiology*, Vol 20, pp. 1073-1080.
- Booth M, Guyatt HL, Li YS and Tanner M (1996), Morbidity attributable to Schistosoma-faponium infection in 3 villages in Dongting Lake Region, Hunan province PR China, *Tropical Medicine and Intervational Health*, Vol 1, No 5, pp. 646-654.
- Borgnolo G, Bradbone F, Scornavacca G, Franco D, Vinci A and Iuculano F (1996), A case-control study of *Salmonella* gastro-intestinal infection in Italian children, *Acta Paediatrica* Vol 85, No 7, pp. 804-808.
- Borriello SP and Barclay FE (1986), An in-vitro model of colonisation resistance to Clostridium difficile infection, *Journal of Medical Microbiology*, Vol 21, No 4, pp. 299-309.
- Brunham RC, Nagelkerke NJD, Plummer FA and Moses S (1994), Estimating the basic reproductive rates of Neisseria gonorrhoeae and Chlamydia trachomatis: The implications of acquired immunity, *Sexually Transmitted Diseases*, Vol 21, No 6, pp. 353-356.
- Brussow H, Rytz A, Freire W, Rahim H and Dirren H (1995), Risk factors associated with the recall of respiratory diseases in Equadorian children, *Journal of Diarrhoeal Disease Research* Vol 13, No 5, pp. 212-218.
- Burattini MN, Massad E, Coutinho FAB and Baruzzi RG (1993), Malaria prevalence amongst Brazilian Indians assessed by a new

- mathematical model, *Epidemiology and Infection*, Vol 111, No 3, pp. 525-537.
- Cheng TC (1973), *General Parasitology*, Academic Press, New York.
- Chiarotti F, Palombi M, Schinaia N, Ghirardini A and Prospero L (1992), Effects of different parametric estimates of seroconversion time on analysis of progression to AIDS among Italian HIV-positive haemophiliacs, *Statistics in Medicine*, Vol 21, pp. 591-601.
- Chiarotti F, Palombi M, Schinaia N, Ghirardini A and Bellocchio R (1994), Median time from seroconversion to AIDS in Italian HIV-positive haemophiliacs: Different parametric estimates, *Statistics in Medicine*, Vol 13, pp. 163-175.
- Clayton D and Hill M (1993), *Statistical Models in Epidemiology*, Oxford University Press, Oxford.
- Coghlan JD (1992), Yersinia, Pasteurella and Francisella, in Greenwood, D, Slack RCB and Peutherer, JF eds., *Medical Microbiology, A guide to Microbial Infections: Pathogenesis, Immunity, Laboratory Diagnosis and Control* 14th edition, Churchill Livingstone, Hong Kong.
- Collett D (1991), *Modelling Binary Data*, Chapman and Hall, London.
- Collett D (1994), *Modelling Survival Data in Medical Research*, Chapman and Hall, UK.
- Cox RD and Oakes D (1984), *Analysis of Survival Data*, Chapman and Hall, London.

- Crowder M (1996), Some tests based on extreme values for a parametric survival model, *Journal of the Royal Statistical Society Series B - Methodological*, Vol 58, No 2, pp. 417-424.
- Cruickshank R, Duguid JP, Marmion BP, and Swain RHA (1975), *Medical Microbiology, The Practice of Medical Microbiology*, Vol 11, 12th edition, Churchill Livingstone, U.K.
- Dagan R, Bar-David Y, Kassis I, Savor B, Greenberg D and Afflalo Y (1991), Cryptosporidium in Beduin and Jewish infants and children in southern Israel. *Israel Journal of Medical Sciences*, Vol 27, No 7, pp. 380-385.
- Derijk WG, Wood IM and Thompson VP (1995), Clinical survival analysis of bonded bridges using a censored Weibull method, *Journal of Dental Research*, Vol 74, p.176.
- Derijk WG, Wood M and Thompson VP (1996), Maximum-likelihood-estimates for the lifetime of bonded dental prostheses, *Journal of Dental Research*, Vol 75, No 9, pp. 1700-1705.
- Dietz K (1988), The dynamics of Spread of HIV Infections in the Heterosexual Population, in Jager JC and Ruitenberg, EJ. eds., *Statistical Analysis and Mathematical Modelling of AIDS*, Oxford University Press, Oxford.
- Dietz K (1995), Some Problems in the Theory of Infectious Disease Transmission and Control, in Mollison D ed., *Epidemiologic Models: Their Structure and Relation to Data*, Cambridge University Press, Cambridge.
- Downs AM, Ancelle RA, Jager JC, Heisterkamp SH, vanDruten JAM,

- Ruiterberg EJ and Brunet J-B (1988), in Jager JC and Ruiten-
berg EJ eds, *Statistical Analysis and Mathematical Modelling of
AIDS*, Oxford University Press, Oxford.
- Duerden B, Reid TMS, Jewsbury JM, and Turk DC (1987), *A New
Short Textbook Of Medical Microbal and Parasitic Infections*, Hod-
der and Stoughton, London.
- Effron B and Tibshrani RJ (1993), *An introduction to the Bootstrap*,
Capman and Hall, New York.
- Ellis ME, Waston B, Mandal BK, Dunbar EM, Craske J, Curry A et al.
(1984) Microorganisms in gastroenteritis. *Archives of Diseases in
Childhood*, Vol 59, pp. 848-855.
- English Life Tables (1996), No 15, Office for National Statistics, Lon-
don, The Stationary Office.
- Faruque AS, Mahalanabis D, Islam A, Hoque SS and Hasnat A (1992),
Breast feeding and oral rehydration at home during diarrhoea to
prevent dehydration. *Archives of Diseases in Childhood* Vol 67,
No 8, pp. 1027-1029
- Ferguson JK, Jorm LR, Allen CD, Whitehead PK and Gilbert GL
(1995), Prospective study of diarrhoeal outbreaks in child long-
daycare centres in western Sydney. *Medical journal of Australia*,
Vol 163, No 3, pp. 137-140.
- Feuer EJ, Kessler LG, Baker SG, Triolo HE and Green DT (1991), The
Impact of breakthrough clinical-trials on survival in population
based tumor registries, *Journal of Clinical Epidemiology*, Vol 44,
No 2, pp. 141-153,

FM1 Birth Statistics (1986), Vol OPCS.

Fontanet AL, Johnston DB, Walker AM, Rooney W, Thimasarn K, Sturchler D, Macdonald M, Hours M and Wirth DF (1993), High prevalence of mefloquine-resistant falciparum malaria in eastern Thailand. *Bulliten of World Health Organisation*, Vol 71, No 3-4, pp. 377-383.

Freund JE and Walpole RE (1987), *Mathematical Statistics*, 4th edition, Prentice-Hall International, USA.

Fuerst, R (1978), *Forbisher and Feurst's Microbiology in Health and Disease*, 14th edition, W.B. Saunders Company, U.S.A.

Gamble, DR (1984), Enteroviruses: Polio, Echo and Coxsackie Viruses, in Brown, F. and Sir Wilson, G. eds., *Topley and Wilson's Principles of Bacteriology, Virology and Immunity*, 7th edition, Vol 4, Edward Arnold, London.

Ghosh S, Sengupta PG, Mondal SK, Banu MK, Gupta DN and Sircar BK (1997), Risk behavioural practices of rural mothers as determinants of childhood diarrhoea. *Journal of Community Diseases*, Vol 29, No 1, pp. 7-14

Glass RI, Lew JF, Gangarosa RE, LeBaron CW and Ho MS (1991) Estimates of morbidity and mortality rates for diarrheal diseases in American children. *Journal of Pediatrics*, Vol 118, No 4, Part 2, pp. S27-S33.

Gonzalez JJ, Koch MG, D'orner D, Lage-Stehr J, Myrveit M and Vavik L (1988), in Jager JC and Ruitenberg EJ eds, *Statistical Analysis and Mathematical Modelling of AIDS*, Oxford University Press, Oxford.

- Gorter AC, Sandiford P, Smith GD and Paw JP (1991), Water Supply sanitation and diarrhoeal disease in Nicaragua: Results from a case-control study, *International Journal of Epidemiology*, Vol 20, pp. 527-533.
- Griffin T and Morris J (Editors) (1987), Social Trends, Vol 17, HMSO, London.
- Gross RJ (1992), Escherichia, in Greenwood D, Slack RCB and Peutherer JF eds., *Medical Microbiology, A guide to Microbial Infections: Pathogenesis, Immunity, Laboratory Diagnosis and Control*, 14th edition, Churchill Livingstone, Hong Kong.
- Gupta DN, Sircan BK, Sengupta PG, Ghosh S, Bann MK, Mondal SK, Saha DR, De SP, Sikdar SN, Manna B, Dutta S and Saha NC (1996), Epidemiological and Clinical Profiles of acute invasive diarrhoea with special reference to mucoid episodes: a rural community-based longitudinal study, *Transactions of the Royal Society of Tropical Medicine and Hygiene*, Vol 19, pp. 544-547.
- Gupta S, Trenholm K, Anderson RM and Day KP (1994A), Antigenic diversity and the transmission dynamics of Plasmodium falciparum, *Science*, Vol 263, pp.961-963.
- Gupta S, Trenholm K, Cox M, Anderson R and Day K (1995), The effects of Antigenic Diversity on Endemic Prevalence, in Mollison, D ed., *Epidemiologic Models: Their Structure and Relation to Data* Cambridge University Press, Cambridge.
- Hauck WW, McKee LJ and Turner BJ (1997), Two-part survival models applied to administrative data for determining rate of and

- predictors for maternal-child transmission of HIV, *Statistics in Medicine*, Vol 16, No 15, pp. 1683-1694.
- Holmes S (1979), *Henderson's Dictionary of Biological Terms*, 9th edition, Longman, London, UK.
- Hosmer DW (1989), *Applied Logistic Regression*, Wiley, New York.
- Jacquez JA, Koopman JS, Simon CP and Longini IM(jr) (1996), Modelling Progression of HIV Infection: Staging and the Chicago MACS Cohort, in Isham V and Medley, G eds., *Models for Infectious Human Diseases: Their Structure and Relation to Data*, Cambridge University Press, Cambridge.
- Johnson NL and Kotz S (1992), *Univariate discrete distributions*, J. Wiley and Sons, New York.
- Joklik WK, Willett HP and Amoss DB (1984), *Zinsser Microbiology*, 18th edition, Appleton Century- Croft, Norwalk, U.S.A.
- Kahle W (1996), Estimation of the parameters of the Weibull distribution for censored samples, *Metrika*, Vol 44, No 1, pp. 27-40.
- Kault DA (1995), Modelling AIDS reduction Strategies, *International Journal Of Epidemiology*, Vol 24, No 1, pp. 188-197.
- Klienbaum DG (1994), *Logistic regression: A self-learning text*, Springer-Verlag, New York.
- Klienbaum DG, (1995), *Survival analysis: A self-learning text*, Springer-Verlag, New York.
- Knight SM, Toodayan W, Caique WC, Kyi W, Barnes A and Desmarchelier P (1992), Risk factors for the transmission of diarrhoea

- in children: A case control study in rural Malaysia, *International Journal of Epidemiology*, Vol 21, pp. 812-818.
- Knox EG (1988) Modelling the AIDS epidemic, in Jager, JC and Ruitenberg, EJ eds, *Statistical Analysis and Mathematical Modelling of AIDS*, Oxford University Press, Oxford.
- Kurstak E, Kurstak C, vanDen Hurk J and Morisset R (1981), Animal Rotavirus, in Kurstak, E ed., *Comparative Diagnosis of Virus Disease*, Academic Press, U.K.
- Last JM (1988), *A Dictionary of Epidemiology*, Second Edition, Oxford University Press, New York.
- Last JM (1995), *A Dictionary of Epidemiology*, 3rd Edition, Oxford University Press, New York.
- Lerman Y, Slepon R and Cohen D (1994), Epidemiology of acute diarrheal diseases in children in high standard of living rural settlement in Israel. *Pediatric Infectious Diseases Journal*, Vol 13, No 2, pp. 116-22.
- Lewis MJ (1992), Shigella, in Greenwood D, Slack RCB and Peutherer JF eds., *Medical Microbiology, A guide to Microbial Infections: Pathogenesis, Immunity, Laboratory Diagnosis and Control*, 14th edition, Churchill Livingstone, Hong Kong.
- Li YH, Klein JP and Moeschberger ML (1996), Effects of model misspecification in estimating covariate effects in survival analysis for small sample sizes, *Computational Statistics & Data Analysis*, Vol 22, No 2, pp. 177-192.

- Lindblom GB, Ahren C, Changalucha J, Gabone R, Kiajser B, Nilsson LA, Sjogren E *et al.* (1995) Campylobacter jejuni/coli and enterotoxigenic Escherichia coli (ETEC) in faeces from children and adults in Tanzania. *Scandinavian Journal of Infectious Diseases*, Vol 27, No 6, pp. 589-593.
- Mackie TJ (1960), *Handbook of Bacteriology: A guide to the Laboratory Diagnosis and Control of Infection*, Edited by Cruickshank, R, 9th Edition, Livingstone Limited, Edinburgh.
- Mackisack MS and Stillman RH (1996), A cautionary tale about Weibull Analysis, *IEEE Transactions on Reliability*, Vol 45, No 2, pp. 244-248.
- Magder L and Brookmeyer R (1993), Analysis of Infectious Diseases Data from Partner Studies with Unknown Source of Infection, *Biometrics*, Vol 49, No 4, pp. 1110-1116.
- Marin CM, Segura JL, Bern C, Freedman DS, Guillermo Lescano A, Benavente LE, Cordero LG, Clavijo L and Gilman JB (1996), Seasonal change in nutritional status among young children in an urban shanty town in Peru, *Transactions of the Royal Society of Tropical Medicine and Hygiene*, Vol 90, No 4, pp. 442-445
- Marubini E and Valsecchi MG (1994) *Analysing Survival data from clinical trials and observational studies*, John Wiley and sons, UK.
- Memis CA, Playfair JHL, Roitt IM, Waklin D and Williams R (1993), *Medical Microbiology*, Mosby-Year Book Europe Ltd., Hong Kong.
- Moreau T, Maccario J, Lellouch J and Huber C (1992), Weighted log

- rank statistics for comparing 2 distributions, *Biometrika*, Vol 79, No 1, pp. 195-198.
- Morris SS, Cousens SN, Lanata CF, and Kirkwood BR (1994), Diarrhoea - Defining the Episode, *International Journal of Epidemiology* Vol 23, No 3, pp. 617-623.
- Mradley M, Chandiwana SK, Bundy DAP and Medley GF (1992), The Epidemiology and population biology of *Necator americanus* infection in a rural community in Zimbabwe, *Transactions of the Royal Society of Tropical Medicine and Hygiene*, Vol 86, pp. 73-76.
- Mudholkar GS and Hutson AD (1996), The exponentiated Weibull Family - Some properties and a flood data Application, *Communications in Statistics - Theory and Methods*, Vol 25, No 12, pp. 3059-3083.
- Murray PR, Kobayasi GS, Pfaller MA and Rosenthal KS (1994), *Medical Microbiology*, 2nd edition, Mosby, St. Louise, U.S.A.
- Musa JD, Iannino A and Okumoto K (1987), *Software Reliability: Measurement, Prediction, Application*, McGraw-Hill Book Company, USA.
- Narendranathan W and Stewart MB (1993), Modelling the probability of leaving unemployment: Competing risks models with flexible base-line hazards, *Applied Statistics*, Vol 42, No 1, pp. 63-83.
- Nath G, Shukla BN, Reddy DC, and Sanyal SC (1993), A community study on the aetiology of childhood diarrhoea with special reference to *Campylobacter* Jejuni in a semiurban slum of Varanasi,

- India. *Journal of Diarrhoeal Diseases Research*, Vol 11, No 3, pp. 165-8.
- Nemo GJ and Diwan AR (1979), Immunology of Persistent and Recurrent Viral Infections, in Dick, C ed., *Immunological Aspects of Infectious Diseases*, M&P Press, Lancaster, pp. 470-471.
- Nester EW, Roberts CE, Lidstorm ME, Pearsall NN and Nester MT (1983), *Microbiology*, 3rd. edition, Saunders College Publishing, Japan.
- Nichols C and Thom BT (1984), Screening for *Cryptosporidium* in stools, *Lancet*, Vol I, p. 735.
- Norusis MJ (1993), *SPSS Base System: Syntax Reference Guide to Release 6.0*, SPSS inc, USA.
- O'Brien AD , LaVeek GE, Thompson MR and Formal SB (1982), Production of *Sh. dysenteriae* type 1-like cytotoxin by *e. coli*, *Journal of infectious diseases*, Vol 146, pp. 763-769.
- Oskrochi GR (1995), Analysis of censored correlated observations, *Journal of Statistical Planning and Inference*, Vol 47, No 1-2, pp.165-180.
- Otani K, Othaki M and Fujita M (1996), Use of the Weibull hazard model to estimate age - specific probability of permanent tooth loss, *Journal of Dental Research*, Vol 75, No 7, pp. 1458-1463.
- Parmer NKB and Machin D (1995), *Survival analysis : A practical approach*, John Wiley and sons, UK.
- Pazzaglia G, Bourgeois AL, el-Diwany K, Nour N, Badran N and Hablas R (1991) *Campylobacter* diarrhoea and an association of

- recent disease with asymptomatic shedding in Egyptian Children. *Epidemiology and Infection*, Vol 106, No 1, pp. 77-82.
- Pedler SJ (1989), Bacteriology of Intestinal Disease, in Hawkey, PM and Lewis DA eds., *Medical Bacteriology: A practical Approach*, Oxford University Press, Oxford.
- Perez-Cuevas R, Guiscafre H, Romero G, Rodriguez L and Gutierrez G (1996), Mothers' health-seeking behaviour in acute diarrhoea in Tlaxcala, Mexico. *Journal of Diarrhoeal Diseases Research*, Vol 14, No 4, pp. 260-268.
- Pickering J, Wiely JA, Leib LE, Walker J and Rutherford GW (1988), in Jager, JC and Ruitenberg, EJ eds., *Statistical Analysis and Mathematical Modelling of AIDS*, Oxford University Press, Oxford.
- Ramiro-Cruz J, CaNo F, Bartlett AV and Mendez H (1994), Infection, diarrhea, and dysentery caused by *Shigella* species and campylobacter jejuni among Guatemalan rural children, *Pediatric Infectious Diseases Journal*, Vol 13, No 3, pp. 216-223.
- Ramprakash D and Morris J (Editors) (1994), Social Trends, Vol 14, HMSO, London.
- Roberts M and Heesterbeek H (1993), Bluff your way in Epidemic Models, *Trends in Microbiology*, Vol 1, No 9, pp. 343-348.
- Samadi AR, Chowdhury AI, Huq MI and Shahid NS (1985), Risk factors for death in complicated diarrhoea of children. *British Medical Journal*, Vol 290, No 6482, pp. 1615-1617.

- Sen-Gupta PG, Nair GB, Mondal S, Gupta DN, Sen D, Sikar SN et al. (1991), Epidemiology of campylobacteriosis in a cohort of rural population near Calcutta. *Epidemiology and Infection*, Vol 106, No 3, pp. 507-512.
- Shao XM (1997), Parametric survival analysis for gating kinetics of single potassium channels, *Brain Research*, Vol 770, No 1-2, pp. 96-104.
- Singleton P (1987), *Dictionary of Microbiology and Molecular Biology*, 2nd edition, J. Weiley and Sons, UK.
- Skirrow MB (1992), Campylobacter and Helicobacter, in Greenwood, D, Slack, RCB and Peutherer, JF eds., *Medical Microbiology, A guide to Microbial Infections: Pathogenesis, Immunity, Laboratory Diagnosis and Control*, 14th edition, Churchill Livingstone, Hong Kong.
- Smales RJ, Berekally TL and Webster DA (1993), Predictions of resin-bonded bridge survival, comparing tow statistical models, *Journal of Dentistry*, Vol 21, pp. 147-149.
- Stanley NF (1981), Diagnosis of Reovirus Infections, Comparative Aspects, in Kurstak, E ed., *Comparative Diagnosis of Virus Disease*, Academic Press, U.K.
- Stedman (1996), *Stedmans Consise Medical Dictionary*, Software, 3rd edition, Novel Network Incorporation.
- Tangermann RH, Gordon S, Wiesner P and Kreckman L (1991) An outbreak of cryptosporidiosis in a day-care center in Georgia, *American Journal of Epidemiology* Vol 133, No 5, pp. 471-476.

- Thein-Hlaing, Than-Shaw and Myat-Lay-Kyin (1991), The impact of three-monthly age-targeted chemotherapy on *Ascaris Lumbricoiles* infection, *Transactions of the Royal Society of Tropical Medicine and Hygiene*, Vol 85, pp. 519-522.
- Thein-Hlaing, Than-Shaw, Htay-Htay-Aye, Myint-Lwin and Thein-Maung-Myint (1984), Epidemiology and Transmission dynamics of *Ascaris lumbricoides* in Okpo village, rural Burma, *Transactions of the Royal Society of Tropical Medicine and Hygiene*, Vol 78, pp. 497-504.
- Thomas JC and Neumann CG (1992), Choosing an appropriate measure of diarrhoea occurrence: examples from a community-based study in rural Kenya, *International Journal of Epidemiology*, Vol 21, No 3, pp. 137-140.
- Tshikuka JG, Scott ME, Graydonald K and Kalumba ON (1996), Multiple infection with *Plasmodium* and helminths in communities of low and relatively-high socio-economic status, *Annals of Tropical Medicine and Parasitology*, Vol 90, No 3, pp. 277-293.
- Turk DC, Porter IA, Duerden BI and Ried TMS (1982), *A Short Text-book of Medical Microbiology*, 5th edition, Hodder and Stoughton, London.
- Velmirovic B (1984), *Infectious Diseases in Europe, A fresh Look*, WHO Regional Office for Europe, Copenhagen.
- Willems G, Carels CEL and Verbeke G (1997), In vitro peel/shear bond strength of orthodontic adhesives, *Journal of Dentistry*, Vol 25, No 3-4, pp. 263-270.

- Woolhouse MEJ (1996), The Interpolation of Immunoepidemiological Data for Helminth Infections, in Isham, V and Medley, G eds., *Models for Infectious Human Diseases: Their Structure and Relation to Data*, Cambridge University Press, Cambridge.
- Woolhouse MEJ, Watts CH and Chandiwana SK (1991), Heterogeneities in transmission rates and the epidemiology of *Schistosme* infection, *Proceedings of the Royal Society of London, Seires B*, Vol 245, pp. 109-114.
- Yeager BA, Lanata CF, Lazo F, Verastegui H and Black RE (1991), Transmission factors and socioeconomic status as determinants of diarrhoeal incidence in Lima, Peru. *Journal Diarrhoeal Diseases Research*, Vol 9, No 3, pp. 186-193.