J. Hyg., Camb. (1981), 86, 49
Printed in Great Britain

Monitoring infectious diseases using routine microbiology data
I. Study of gastroenteritis in an urban area

BY HILARY E. TILLETT

Communicable Disease Surveillance Centre, Public Health Laboratory Service, 61 Colindale Avenue, London NW9 5EQ

AND MAIR E. M. THOMAS

Epidemiological Research Laboratory, Central Public Health Laboratory, Colindale Avenue, London NW9 5HT

(*Received* 2 *May* 1980)

SUMMARY

Sources of information for monitoring infectious disease are routine data, special surveys and *ad hoc* investigations. In practice much use is necessarily made of routine notifications and laboratory records although this reporting is often incomplete and may therefore be biased.

In a retrospective study of a 16-year series (up to 1968) of routine records concerning the diagnosis of gastroenteritis at one Public Health Laboratory we found it possible to identify biases. During school outbreaks of dysentery, laboratory investigation of diarrhoea increased appreciably and such response to publicity affects the use of routine data in surveillance. Although the patients examined were probably representative diagnostically, their selection may not have reflected the age incidence of disease. Valid geographical comparisons within the urban area were not feasible because medical practitioners differed in their use of laboratory facilities and in their habits of notification. Nevertheless, as far as can be established retrospectively, these data did reflect time trends in disease incidence and so had value for monitoring purposes.

Several of the biases defined are likely to apply to other sets of routine data. A further communication will describe a statistical method of correcting for quantifiable bias.

INTRODUCTION

In practice, when infectious diseases are being monitored, it is difficult to obtain full information about the incidence of transmissible diseases in the community or even to know with certainty which sections of the population are most affected. In Britain, data collected on a routine basis are often used for monitoring infectious disease although it is recognized that such data are often incomplete. More complete special surveys are restricted in range and period because of their cost. Routine statistics which have been collected continuously over many years include

0022-1724/81/0041-1980 $01.00

Table 1. *Notifiable infectious diseases in England and Wales*

Acute encephalitis	Ophthalmia neonatorum
Acute meningitis	Paratyphoid fever
Acute poliomyelitis	Plague
Anthrax	Rabies
Cholera	Relapsing fever
Diphtheria	Scarlet fever
Dysentery	Smallpox
Food poisoning	Tetanus
Infective jaundice	Tuberculosis
Lassa fever	Typhoid fever
Leprosy	Typhus fever
Leptospirosis	Viral haemorrhagic fever
Malaria	Whooping cough
Marburg disease	Yellow fever
Measles	

statutory notifications, laboratory reports and records of clinical diagnoses collated by the Royal College of General Practitioners (RCGP).

Statutory notifications now apply to the 29 conditions listed in Table 1 which medical practitioners are required to report officially. Clinically severe or exotic disease is likely to be notified. Milder and commoner illnesses may escape medical attention and reporting of cases actually diagnosed by a doctor is known to be incomplete (Sharp, 1972). Notifications are traditionally founded on clinical diagnosis and may lack microbiological verification.

Many British microbiology laboratories voluntarily send regular reports of their diagnostic findings to the Communicable Disease Surveillance Centre (CDSC) at Colindale, and this information is circulated confidentially in the weekly and quarterly *Communicable Disease Report* (*CDR*). This information, like that from notification, is incomplete since only selected cases have a laboratory investigation, and not all laboratories contribute reports to the CDSC. However, the CDSC system has the advantages that a very wide range of communicable diseases is covered and that their aetiology has usually been established.

Although particular epidemiological surveys are valuable, they are usually brief so that routine statistics have to be used for the study of longer trends. It is therefore important to try to recognize and measure the biases in such data. We have attempted to assess a 16-year series of routine laboratory data, recognizing that, because the study is retrospective, it cannot be comprehensive.

EXAMPLE OF ROUTINE DATA STUDIED

The records of the examination of faeces specimens at a single laboratory in London, Edmonton Public Health Laboratory, over the 16-year period 1953–68 have already been analysed to summarize the microbiological findings (Thomas & Tillett, 1975). This laboratory had a virtual monopoly of general practitioner (GP)-referred bacteriology specimens within the London Borough of Enfield which, prior to the 1964 local government reorganization in London, comprised the former boroughs of Edmonton, Enfield and Southgate. The study concentrated on the first case referred to the laboratory with symptoms compatible with a diagnosis of infective gastroenteritis from any household incident regardless of whether or not household contacts were tested. These first cases are defined as index cases. The information available in the records included laboratory results and name of GP. Information on age, sex and address of the patients was virtually complete and we knew something of symptoms and time from onset for most patients.

During the 16-year period the laboratory investigated 20273 such index cases and these have been analysed according to the microbiological findings, age and sex groups, and seasonal and annual patterns. Food poisoning and dysentery are among the 29 notifiable diseases in England and Wales. The laboratory had fairly complete information about the households of index cases infected with salmonellas or shigellas, because household contacts of index cases were tested and follow-up specimens collected from infected patients. More detailed analyses were therefore possible with these two groups of households (Thomas & Mogford, 1970; Thomas & Tillett, 1973*a*).

Summary of laboratory findings

Table 2 shows the diagnoses made by the laboratory for the 20273 index cases. The most common organism isolated was *Shigella sonnei*, found in 9·2 % of cases, followed by the salmonellas, found in 2·3 % of cases. Enteropathogenic *Escherichia coli* were routinely sought only in specimens from children under five years old. *Giardia lamblia* was diagnosed by light microscopy in 1·4 % of cases. Less common pathogens (listed in Thomas & Tillett, 1975) were found in 0·6 % of cases. A condition we have labelled 'fatty diarrhoea' was observed in 6·7 % of cases. The condition was recognized at microscopy when undigested fat globules or their derivatives were seen. This condition has been reported as infectious in behaviour but no pathogen was found (Thomas, 1952).

For most cases, about 80 %, the laboratory was unable to make a diagnosis. During the years concerned (1953–68) facilities for virology and electron-microscopy were almost unavailable. The role of campylobacters, yersinias and certain other bacteria in diarrhoea in the United Kingdom was then unknown, as was the extent of viral gastroenteritis.

Table 2. *Laboratory findings from diagnostic specimens received from 20273 index cases*

Laboratory diagnosis	Number of index cases	Percentage of total index cases
Dysentery due to *Sh. sonnei*	1868	9·2
Salmonellosis	475	2·3
E. coli infection	232	1·1
Giardiasis	288	1·4
Other pathogens	122	0·6
'Fatty diarrhoea'	1365	6·7
No diagnosis	16076	79·3
Total cases	20273*	100

* Including 143 double and 4 triple diagnoses.

POSSIBLE SOURCES OF BIAS IN LABORATORY DATA

Choice of period of study

Patients with relevant symptoms had first to consult their GP, and it is not possible to judge retrospectively whether patients were consistent over the 16 years in the readiness with which they did so. However, the period of study was chosen to start in 1953 when the 1948 National Health Service, with its open access to diagnostic services for GPs, was well established. At the end of 1968 the laboratory was closed. Throughout the 16 years of study the laboratory records had been kept in a standard fashion.

The responsibility for promotion of interests in the diagnosis of notifiable disease lay with the Medical Officers of Health (MOsH). The years 1953–64 saw only one change among the MOsH in the three old boroughs served by the laboratory. In the 1964 reorganization the MOH for Enfield became MOH for the new London Borough of Enfield, thus providing continuity.

Laboratory techniques

A large degree of consistency was maintained in the laboratory methods. During the 16 years some new media and techniques were tried out in parallel with those established, but as none improved the isolation rates only minor technical modifications were adopted. Over the years there was an increase in the number of *E. coli* serotypes recognized to be pathogenic, but the total isolations of *E. coli* were tending to decrease.

Selection of patients for laboratory investigation

The main area of selection of index cases for laboratory investigation was through the action of the GPs, and it is their collective consistency which will have had the greatest effect on the reliability of time trends.

Enfield and Haringey Family Practitioner Committee permitted inspection of records listing GPs practising as principals or partners in the Enfield boroughs. They had available books for 1954 and 1969, these being as close as possible to the beginning and end of the period of study. In 1954 there were 190 GPs listed; by 1968 only 62 of those names still appeared in the list. We then noted which of the laboratory index cases had been referred by these 62 GPs during 1954 and 1968 and during a sample year in the middle of the period, 1960. The results are summarized in Table 3.

Table 3. *Proportions of annual index cases referred by the 62 GPs practising in the area throughout the period of study*

	Year		
	1954	1960	1968
Total index cases	1295	1883	1211
No. referred by the 62 GPs	654	945	715
Percentage of total	51	50	59

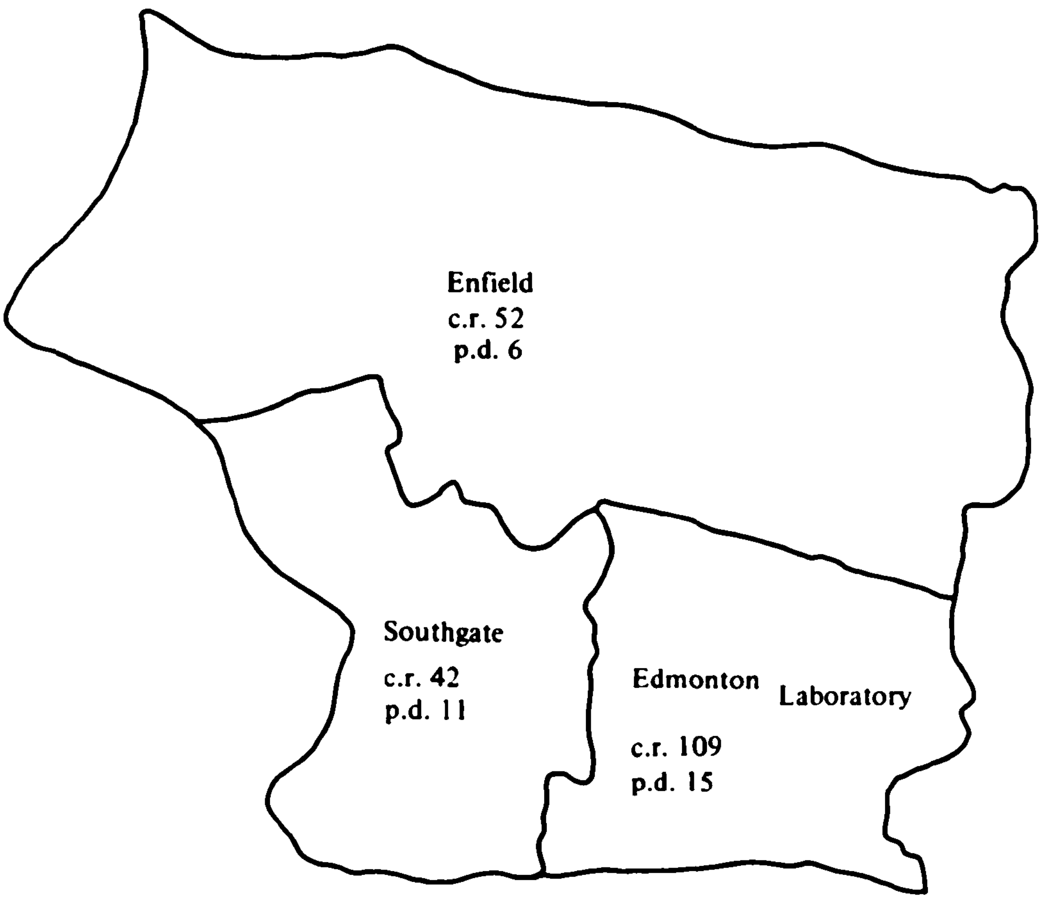

Fig. 1. Diagram to show the case rates and estimated population densities of the three old boroughs of Enfield, Southgate and Edmonton before their amalgamation in 1964.

There was some fluctuation in the proportions of annual index cases referred by these 62 GPs but no trend. Since the 62 GPs were referring about half the total annual number of index cases we conclude that new GPs collectively used the laboratory facilities as readily as their predecessors.

However, individual GPs differed greatly in their referral habits. In 1954 the numbers referred by each GP ranged from 0 by some, to 113 by one. We noted that many of those referring large numbers had practice addresses in the old borough of Edmonton near the laboratory. So the total index case rates have been estimated for the three old boroughs separately based on sample years before the 1964 amalgamation. These case rates together with estimated population densities are shown in Fig. 1. In Edmonton the 16-year case rate was 109 per 10000 population, whereas in Enfield it was 52 and in Southgate 42. Edmonton had the densest mid-period population of about 15000 per square mile and this might have been a factor influencing disease incidence. But the population of Southgate was considerably denser than that in Enfield and yet did not have a higher case rate. It appears that proximity to the laboratory influenced whether or not a case was investigated.

A similar geographical distribution was seen with index cases diagnosed as Sonne dysentery or food poisoning due to a salmonella. Since it was customary in these boroughs for laboratory confirmation to precede inclusion of notifications of these two diseases in the weekly returns made by the MOsH to the Registrar General, the notifications, which will have included positive household contacts, showed a similar geographical clustering around the laboratory.

Age group of patients referred and the results of laboratory diagnosis as compared with other studies

The population in the three boroughs derived from the 1961 census is shown in Table 4 together with laboratory index case rates. The age structure of the population in Edmonton, where most of the referred cases lived, is similar to that of the combined populations for the three boroughs, which has been used to calculate the rates. However, it cannot be checked that this was similar to the age structure of the practice lists of those GPs who regularly referred patients; therefore the following comparisons are tentative.

A national study of morbidity made by volunteer GPs in England and Wales covered a 12-month period during 1955 and 1956 (Logan & Cushion, 1958). Illnesses described as dysentery, food poisoning, gastroenteritis or diarrhoea and vomitting were listed for 3·2 per 100 patients. The diagnoses of dysentery and food poisoning accounted for 7 % and 2 % respectively of all such illnesses and are comparable with the Enfield findings of 9 % for Sonne dysentery and 2 % for salmonellosis. Age-specific rates were given for very broad age groups and it is only possible to compare the ratios of child to adult case rates which were 2·8 for the national study but higher, 4·5, for the Enfield study.

Three smaller surveys were reported in Britain during the years of the Enfield study and in each of these the patients had specimens examined bacteriologically.

A detailed study, reported from a group practice in a suburb near London, lasted for two years during which 738 patients with acute gastroenteritis were investigated (Tuckman *et al.* 1962). One fifth of cases yielded a bacterial pathogen, mostly *Sh. sonnei*, compared with 13 % of Enfield index cases. The ratio of child to adult case rates was about 3·7, whether or not contact cases were included. This was again lower than the 4·5 for Enfield index cases. At microscopy blood and pus cells were said to be 'seldom' seen in non-Sonne cases and in 'less than half' of Sonne cases, compared with 13 % in Enfield non-Sonne cases and 41 % in Sonne cases.

A study from Edinburgh investigated 589 cases (Knox *et al.* 1967). A bacteriological diagnosis was made in 106 cases of Sonne dysentery and one of salmonellosis. No microscopy was performed and the population at risk was not stated.

One further study was reported from an industrial area near London (Smither, 1953). Of 90 cases investigated, 72 (80 %) yielded no bacterial pathogen, 14 had Sonne dysentery, 1 had salmonellosis and 4 'other' pathogens were found. The age structure of the population at risk was not given.

The Enfield study covered a 16-year period during which considerable fluctuation in numbers of cases and prevalent pathogens was observed. In this section we are attempting comparison with practice-based studies lasting only one or two years and involving all patients visiting their GP, and not the routine selection of index cases examined in Enfield. However, there is no evidence to suggest that

Table 4. *Population of Enfield, Edmonton and Southgate at the* 1961 *census and average annual index case rates per* 10000 *population*

Age group (years)	Male	Female	Total
0–4	8936 (226)	8407 (180)	17343 (204)
5–9	7919 (149)	7757 (122)	15676 (136)
10–14	10602 (46)	10078 (34)	20680 (41)
15–39	42315 (31)	43303 (38)	85618 (35)
40+	60463 (20)	74077 (23)	134540 (22)
Total	130235 (49)	143622 (44)	273857 (46)

the Enfield-selected cases were grossly unrepresentative of bacterial pathogens then being sought in Britain. The national morbidity study and the largest of the three practice studies both indicated that Enfield GPs may have been referring a slightly larger proportion of their child than adult index cases. However, the GPs remained consistent in this since a moving average of the ratio of child to adult index cases remained constant over the 16 years (Tillett, 1977).

Influence of dysentery outbreaks

During the 16 years there were 16076 cases investigated for which the laboratory could make no diagnosis, and these cases varied in quarter years from 146 to 455. A plot of these quarterly undiagnosed cases is shown in Fig. 2, together with those quarterly index cases diagnosed as having Sonne dysentery. These two series follow remarkably similar patterns and are highly correlated. This could be explained if the undiagnosed cases included a large number of dysentery cases in which the laboratory had failed to isolate the shigella. But a statistical test was developed for use with data on multiple index cases (Tillett & Thomas, 1974) which showed that the laboratory was highly successful at isolating *Sh. sonnei* from true index cases: over 94% during the first week from onset of symptoms and over 74% subsequently. Since three-quarters of undiagnosed index cases were investigated within one week from onset, false negative diagnoses cannot explain this correlation between Sonne dysentery and undiagnosed cases.

In Enfield during this period Sonne dysentery was a nuisance and causing outbreaks in primary schools and nurseries (Thomas & Tillett, 1973*b*). It was not endemic in the area but occurred in epidemic waves every two to three years. A local policy was being implemented in the area by the laboratory and MOsH with the co-operation of school headteachers and nursery matrons, aimed at control and prevention of outbreaks. The most successful part of this policy was the exclusion from school of pupils with diarrhoea pending laboratory investigation. Another part of the policy was to encourage GPs to seek laboratory investigation

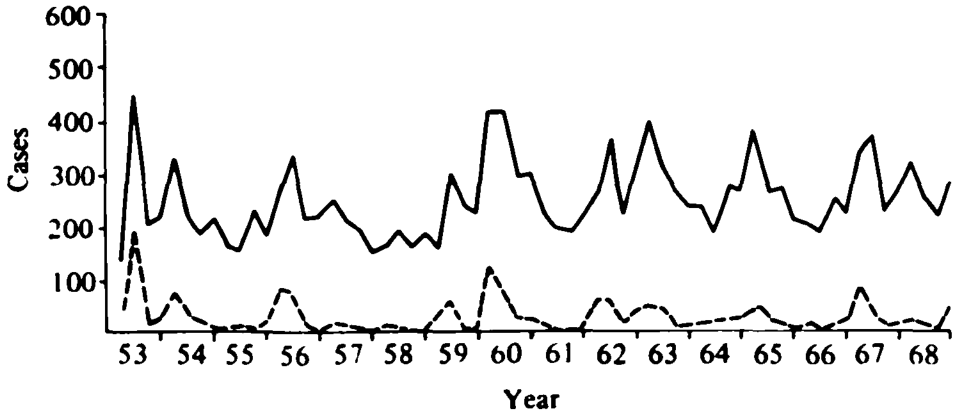

Fig. 2. Index cases for which the laboratory made no diagnosis (—) and cases of Sonne dysentery (– –).

of any patient they suspected of having dysentery. It therefore seems from Fig. 2 that GPs, when they became aware that Sonne dysentery was back in the area, referred a greater proportion of their diarrhoeal patients which included many non-dysentery cases.

This long series of 16076 examined but undiagnosed cases provided a unique opportunity to study patterns of gastroenteritis of unknown aetiology. But the apparent time patterns were being artificially affected by the influence of dysentery outbreaks on GP referral habits. Therefore we used multivariate statistical techniques to remove this artificial pattern and study the real underlying patterns. These will be described in the next paper (Tillett, 1981).

DISCUSSION

The Enfield laboratory data recorded in a consistent way over a long period provided considerable microbiological information together with the opportunity to try to study time trends in infective gastroenteritis in an urban area. Because routine laboratory data contain many biases, interpretation of these data has to be made with caution.

The main selection of index cases was by the decisions of GPs practising in the boroughs served by the laboratory. Retrospectivity we have established that collectively these GPs' referral habits remained broadly consistent and therefore the overall laboratory data do indicate trends in disease incidence, even though they do not measure actual incidence.

The largest bias in these laboratory data was due to a 'publicity' factor, here the reporting of dysentery outbreaks, which resulted in GPs referring a greater proportion of their patients with relevant symptoms for laboratory investigation. Other possible 'publicity' factors, such as the diagnosis of a case of typhoid or paratyphoid in the area, were not found to have influenced the data. This is probably because MOsH were responsible for investigation of contacts of such cases, and only a very few GPs would have had reason for concern.

Other collections of routine data on infectious diseases, such as notifications and national laboratory data as collected by CDSC, must be affected by similar factors either national or local, through professional interest or through the popular media. In the Enfield data, we identified one such factor and have been able to take it into account when analysing the data to study disease patterns (Tillett, 1981).

That individual GPs differ from one another in their referral habits has been reported elsewhere (Taylor *et al.* 1975). We were fortunate that new GPs entering the area were not collectively very different from the GPs giving up their practices in the same area. No information is available retrospectively to assess how consistent patients were in consulting their GP.

Geographical biases were evident even within this relatively small area. Over half the index cases lived within the borough in which the laboratory was situated, even though the diagnostic service was available for two other boroughs. The arrangements for submission of specimens by patients living in Southgate and Enfield were no more difficult. The specimen had to be taken to the nearest of four collection points, visited daily by the laboratory van. The impression of the laboratory director is that the GPs who regularly referred patients for laboratory investigation were often those who could conveniently visit the laboratory from their practices to call in and discuss interesting current diseases. A similar geographical imbalance has been observed in Scotland with GP referral rates of patients to hospital. Referral rates were higher for patients living closer to the out-patient clinic (Gruer, 1972).

This grographical disproportion in Enfield was also true for the notified cases of dysentery and salmonella food poisoning, since it was usual in Enfield to have laboratory confirmation before notification, so as to avoid corrections having to be made. Geographical variations in national routine statistics must be subject to even greater biases. Considering national notifications of dysentery over this same period, not all MOsH will have restricted weekly returns to bacteriologically confirmed cases, and few areas would have been actively seeking to investigate all symptomatic school children.

We suggest that notifications and laboratory data should not be used to measure relative disease incidence in different geographical regions but more appropriately to indicate trends within areas.

As regards age distribution of disease, it is not possible to check for such biases retrospectively, but we have made tentative comparisons with other studies (Logan & Cushion, 1958; Tuckman *et al.* 1962) and these indicate that Enfield GPs were probably referring a greater proportion of their child than adult cases, but they did this consistently and so time trends will not have been affected.

The monitoring of infectious diseases on a national scale requires use of routinely collected data, since special surveys are costly and therefore usually have to be of short duration. Routine data may be incomplete and contain biases, and in this paper we have investigated a particular series of laboratory data. But these data were easy to study because they were free from certain influences. They concerned a single laboratory which had continual responsibility for exactly the same area. The laboratory did not change its basic methods, nor were there many relevant changes in staff. The GPs responsible for selecting patients did not radically alter their referral practice and generally we conclude that the data were of value for studying disease patterns. We would expect national laboratory data to contain more biases. Geographical variations are likely to be large, but the data should reflect trends in incidence provided that possible influencing factors, such as publicity, are taken into account.

REFERENCES

GRUER, R. (1972). *Outpatient Services in the Scottish Border Counties.* Scottish Health Services Studies No. 23, Scottish Home and Health Department.

KNOX, J. D. E., MACNAUGHTON, G., LAWRENCE, A. R. & ROBERTSON, A. A. (1967). Diagnosis of diarrhoea in general practice. *Lancet* ii, 1392.

LOGAN, W. P. D. & CUSHION, A. A. (1958). *Morbidity Statistics from General Practice*, 1. Studies on medical and population subjects, no. 14. Her Majesty's Stationery Office.

SHARP, J. C. M. (1972). The notification of infectious disease. *Health Bulletin Scottish Home and Health Department* **30**, 147.

SMITHER, W. J. (1953). Gastroenteritis in general practice. A study of 90 unselected cases. *British Medical Journal* i, 376.

TAYLOR, R. J., HOWIE, J. G. R., BRODIE, J. & PORTER, I. A. (1975). Use of bacteriological investigations by general practitioners. *British Medical Journal* iii, 635.

THOMAS, M. E. M. (1952). Epidemic abdominal colic associated with steatorhoea. *British Medical Journal* i, 691.

THOMAS, M. E. M. & MOGFORD, H. E. (1970). Salmonellosis in general practice. Observation of cases and their households in Enfield. *Journal of Hygiene* **68**, 663.

THOMAS, M. E. M. & TILLETT, H. E. (1973*a*). Dysentery in general practice: a study of cases and their contacts in Enfield and an epidemiological comparison with salmonellosis. *Journal of Hygiene* **71**, 373.

THOMAS, M. E. M. & TILLETT, H. E. (1973*b*). Sonne dysentery in day schools and nurseries. *Journal of Hygiene* **71**, 593.

THOMAS, M. E. M. & TILLETT, H. E. (1975). Diarrhoea in general practice: a sixteen-year report of investigations in a microbiology laboratory with epidemiological assessment. *Journal of Hygiene* **74**, 183.

TILLETT, H. E. & THOMAS, M. E. M. (1974). Culture of the faeces in the diagnosis of Sonne dysentery: a statistical method for estimating the true isolation rate. *International Journal of Epidemiology* **3**, 177.

TILLETT, H. E. (1977). Ph.D. Thesis, University of London.

TILLETT, H. E. (1981). Monitoring infectious diseases using routine microbiology data. II. An example of regression analysis used to study infectious gastroenteritis. *Journal of Hygiene* **86**, 59.

TUCKMAN, E., CHAPPLE, P. A. L., FRANKLIN, L. M., MANSER, I. N., WOODALL, J. T., RANDALL, K. J. & MCDONALD, J. C. (1962). Acute gastrointestinal illness in general practice. *British Medical Journal* i, 135.