INTERNATIONAL JOURNAL OF HEALTH PLANNING AND MANAGEMENT
Int J Health Plann Mgmt 2002; **17**: 1–21.
Published online in Wiley InterScience (www.interscience.wiley.com). **DOI**: 10.1002/hpm.686

The relative prevalence of disease symptoms for ill persons: Evidence from Benin

Christophe Muller*
School of Economics, University of Nottingham, University Park, Nottingham NG7 2RD, UK

SUMMARY

In developing countries, the usual modelling of the correlates of health problems is not a good fit for the health phenomena encountered and the available data. Indeed, three common situations occur: (a) it is often the observed symptoms that are used to determine medical interventions instead of specific disease diagnostics or general health indicators; (b) the ill persons described by the data are often affected by multiple health problems; and (c) the correlates of the full spectrum of all symptoms need to be considered together. In this paper, these issues are dealt with by proposing a statistical approach based on competing scores of symptoms that explain their relative prevalence among the observed ill persons.

Using multinomial logit models, the relative prevalence of four symptoms was estimated for four age classes of ill persons in Benin. Socio-demographic characteristics, household equipment and consumption behaviour are shown to influence the relative prevalence of symptoms and therefore could be used to decide what treatment to use. Moreover, living standards and economic activities are important and the pattern of symptoms among poor or agricultural ill persons differs from that of the rich or the non-peasants. The proposed method can be used to assist the definition of target groups and to guide the allocation of scarce resources in poor countries. Copyright © 2002 John Wiley & Sons, Ltd.

KEY WORDS: [Q1] Q1

INTRODUCTION

Illnesses and diseases severely affect population welfare in developing countries. For many LDCs (less developed countries), the usual modelling approaches to investigate the correlates of health problems are not a good fit for the health phenomena encountered in practical situations or the available data. In this paper a practicable statistical approach is proposed to investigate the correlates of symptoms, adapted to the complexity of the health situation and to the data limitations in LDCs. This approach can help to guide empirical therapy and the allocation of scarce resources in very poor countries.

* Correspondence to: Dr C. Muller, School of Economics, University of Nottingham, University Park, Nottingham NG7 2RD, UK. E-mail: christophe.muller@nottingham.ac.uk

Contract/grant sponsor: PARADI.

In LDCs it is generally the symptoms that are observed rather than specific diagnostics. These symptoms therefore constitute the only basis for many feasible medical interventions. Identifying high-risk demographic or economic groups for general symptoms and separating the relative risks of prevalence of symptoms for these groups can assist the allocation of public health funds.

Despite this concern, micro-econometric studies for developing countries are rarely devoted to the study of a detailed set of symptoms. Behrman and Deolalikar (1988), Strauss and Thomas (1995, 1998), who survey the economic literature on health problems in LDCs, and Sickles and Taubman (1997) for all countries, show that most economic studies deal with general indicators of individual health. In micro-econometric health studies, the health status is measured by using: (i) anthropometric measures; (ii) records of incapacity for carrying out some activities; (iii) clinical measures; (iv) respondent-reported disease symptoms and general health self-evaluation. Clinical measures incorporating accurate medical information are very rare. Such measures are generally necessary for precise illness diagnosis. For example, blood tests are often the only way of knowing with certainty the infection agent (e.g. for malaria). Typical medical information such as the pulse rhythm, the aspect of the throat and the tongue, the precise degree of fever, etc, would also enable analysts to narrow the range of possible diseases. The self-reports of symptoms or of illness duration (or illness severity) are more common. Symptoms such as the occurrence of fever or wounds are rarely used in econometric studies, in contrast with illness duration that has been the object of much estimation. However, illness duration responses have been found to be correlated with socio-economic status, reflecting subjective aspects of any general self-assessment of health. Anthropometric measures (weight-for-age, weight-for-height, body mass index) are also often used in econometric analyses, but not necessarily related to precise diseases, in particular because they may be caused by malnutrition. Information about incapacity to function at particular tasks has only been used recently in LDCs. Unfortunately, it does not determine the origin of the incapacity. In contrast to such synthetic health indicators, medical practitioners are confronted with a set of endemic illnesses with distinct characteristics. Since health policies in LDCs are often directed to specific illnesses through specific medical programmes, distinguishing the prevalence of different diseases is a crucial concern in practice. Sometimes it is not possible to cure the disease, but it is possible to alleviate the symptoms. In this case, information about symptoms is of interest in itself and not only as proxy for subjacent diseases.

In addition to the limited information on disease description many medical data in LDCs are based on samples of ill persons rather than individual samples representative of the total population.[†] Since these populations are natural targets of health policies, it is of interest to elucidate which factors are associated with each symptom, using these samples of ill persons despite the selectivity problem that they involve. Indeed, hospitals, health centres and dispensaries are mostly visited by ill persons.[‡]

[†]Besides, most sources of morbidity statistics only provide information on special population groups (Lilienfeld and Stolley, 1994).

[‡]This means that, in this approach, data sets are analysed conditionally on the assumption that selectivity effects can be neglected.

Once data about such samples are available, one can study the relation of disease with the individual characteristics of the ill person. The present knowledge of socio-economic factors influencing specific diseases is growing but still limited. The epidemiology literature (Salvato, 1982; Mausner and Kramer, 1985; Lilienfeld and Stolley, 1994; Souhami and Moxham, 1994) and the economic literature (Behrman and Deolalikar, 1988; Strauss and Thomas, 1995, 1998; Feachem *et al.*, 1992; Zweifel and Breyer, 1997; Folland, Goodman and Stano, 1997; Philipson, 2001) consider biological, genetic and environmental factors as well as economic and demographic characteristics. Moreover, the authors generally study one given disease without accounting for other health problems, whereas the populations of ill persons in LDCs are often affected by multiple health problems (Drasar *et al.*, 1981). In developing countries, ill persons often suffer from several diseases and injuries simultaneously. However, clinical symptoms are difficult to observe for all the diseases and only the main symptom is generally recorded during health surveys. In Africa, medical programmes are often directed against a few major diseases (malaria (Asenso-Okyere and Dzator, 1997), bilharzias (Audibert and Etard, 1998), AIDS, tuberculosis, etc.) that are first identified by general symptoms (e.g. fever for malaria). Indeed, considering the high prevalence of malaria in some areas, it may be more efficient not to check it specifically by using blood tests requiring medical equipment, before undertaking treatment. This is especially the case for remote rural areas far away from medical facilities. In this situation, the knowledge of socio-economic factors correlated to the relative prevalence of a particular symptom can be useful.

However, the correlates to be included in prevalence equations are subject to debate. Epidemiological models link host factors, environment factors and the disease agent (Salvato, 1982; Mausner and Kramer, 1985; Lilienfeld and Stolley, 1994). The disease agent is unobserved in our data. Host factors affect susceptibility to disease while environment factors influence exposure, and sometimes also indirectly influence susceptibility. Estimations of probability models of the incidence or prevalence of a specific disease are common in the health economics literature (Heller and Drake, 1979; Cebu Study Team, 1992; Rampey *et al.*, 1992; Geoffard and Philipson, 1995; Philipson, 2001). They show that member's age, family size and other socio-demographic variables, and economic behaviour are all influential in explaining the disease occurrence.

However, the full spectrum of symptoms and their relative prevalence are not considered in these studies. This is not only a problem for LDCs where ill persons are often affected by several health problems, but also in general contexts because the prevalence of the main symptom does not inform on pre-clinical conditions while at that time the person may be infectious to others. We shall specify a model incorporating the spectrum of all symptoms so as to integrate pre-clinical conditions in the analysis. Moreover, studying prevalence rather than studying incidence may be more appropriate for new medical interventions in LDCs. Indeed, where little health infrastructure is available, medical intervention treatment to initiate is first directed towards the stock of existing ill persons rather than towards new cases occurring during a specific period.

How can the effects of correlates of health problems be modelled to account for the complexity of health situations and the data limitations in LDCs? What are the correlates of the relative prevalence of symptoms of ill persons in the case of Benin? The aim of this paper is to answer these questions by proposing and estimating a multinomial logit model of the relative prevalence of symptoms. The estimation results suggest a way of prioritizing health care interventions, which may not be obvious from existing health indicators for a particular area.

THE MODEL

The theoretical framework

Our approach is based on the assumption that every observed symptom is characterized by its 'latent potentiality'[¶] (or score) that is explained by environmental factors and host characteristics. Indeed, as suggested by Mausner and Kramer (1985) or Lilienfeld and Stolley (1994), the development of a specific disease must account for successive stages: susceptibility, presymptomatic phase, clinical disease, and perhaps disability or death. During the subclinical stages, no symptom can be observed. Thus, symptom scores characterize the development level of diseases. The score of the ith disease (S_i) is an indicator of its latent severity and describes the spectrum of the disease. The first part of this spectrum (susceptibility or pre-symptomatic level) corresponds to unobserved events occurring in the human organism from the time of exposure. Although unapparent, these events associated with early infection are important because they may play a role in the transmission of infectious agents. When the disease is actually observed, the score can be interpreted as the level of clinical symptoms. We assume that the declared symptom corresponds to the highest score obtained among the set of symptoms. The reason to assume that one symptom dominates is that only one symptom has been recorded in the questionnaire for the data described below.[§]

To formalize these assumptions, Eqn (1) shows the determination of the scores of symptoms ($S_1, \ldots, S_q$) by the correlates X_{ih} for each ill person h and each symptom i.

$$S_{ih} = X'_{ih}b + \varepsilon_{ih} = X'_{h}b_i + \varepsilon_{ih} \tag{1}$$

where $i\ (= 1, \ldots, q)$ is the symptom index and $h(= 1, \ldots, N)$ is the ill person index. X_{ih} (or X_h) is the vector of regressors explaining the score level of symptom i for ill person h. b (or b_i) is the vector of parameters, ε_{ih} is an error term accounting for unobserved heterogeneity of ill persons and households, unobserved attributes of environment and symptoms, and measurement errors. The specification $X'_h b_i$ is another way to express $X'_{ih}b$ when the same set of characteristics is considered for

[¶]MIMIC-models (van de Ven and van der Gaag, 1982; van Vliet and van Praag, 1987; van de Ven and Hooijmans, 1991) treat the general health status itself as a latent variable.

[§]Similar 'hidden diseases' occur for mortality statistics established in terms of 'cause of death'. Traditionally a single cause of death has been recorded for routine statistical tabulations. It has been criticized as not providing a complete representation of events (Krueger, 1966), and since 1978 multiple cause-of-death data are available for all recorded deaths in the USA.

all symptoms ($X'_{\mathrm{ih}} = X'_{\mathrm{h}}$ for all i). Defining different coefficients b_i for the different symptoms is consistent with the existence of health processes specific to each symptom. For each observed ill person, only one symptom is observed. However, Eqn (1) is assumed valid also for the unobserved symptoms for which it is a latent equation.

Because the scores cannot be observed, Eqn (1) cannot be directly estimated. However, it can be used, as shown below, as the basis for the definition of equations for the relative odds of symptoms. The independent variables X_{ih} are described below and correspond to the available information from the questionnaire. As only differences in scores are useful in determining the observed symptom, we do not consider in these equations the explanatory variables whose effects are not specific to the considered symptoms since they disappear with the comparison of scores (in the difference $S_{\mathrm{ih}} - S_{\mathrm{jh}}$ for two symptoms i and j with $b_i = b_j$ for the considered variable). Then, the effects of a correlate x_{ik} can be considered as an interaction effect of the kth regressor x_k with a dummy variable of the ith symptom. This approach is consistent with the assumption that any host or environment variable specifically affects the symptom because this affect results from specific biological processes. Alternatively, as we indicate in Eqn (1), it can be incorporated by using different coefficients b_{ki} (vector b_i is composed of the coefficients $b_{\mathrm{ki}}, k = 1, \ldots, K$ and K is the number of regressors) for describing the effect of variable x_{k} on symptom $i(i = 1, \ldots, \mathrm{I})$. To ensure identification, we drop the intercept term associated with the alternative 'Other Symptoms' (see below). We now show how to derive the relative probabilities of the symptoms from this model by specifying the error terms.

The econometric specification

The observed symptom j for ill person h is such that its score exceeds the score of all other symptoms for the same person: $S_{\mathrm{jh}} > S_{\mathrm{ih}}$ for all symptoms $i \neq j$. This expresses that the only declared symptom in the survey is the one with the more advanced development. The probability of observing the prevalence of symptom j for person h is therefore probability $P_{\mathrm{jh}} = \Pr(S_{\mathrm{jh}} > S_{\mathrm{ih}}$, for all $i \neq j)$. Only the differenced model, $S_{\mathrm{ih}}(.) - S_{\mathrm{1h}}(.)$, can be identified where S_{1h} is the score function of a reference alternative (number 1 here). Indeed, the identification of all the score functions would require the observation of all pre-clinical and clinical symptom levels for a given ill person.

To convert these score differences in probabilities of symptom prevalence, we use a multivariate generalization of the simple logistic model, the multinomial logit model (MNL). Consistently with the usual practice for MNL, the errors $\varepsilon_{\mathrm{ih}}$ are assumed to independently follow identical Gumbel distribution. Then, the probability of prevalence of symptom j for person h can be shown to be that of a MNL (see Maddala, 1983; Jones, 1999).

$$P_{\mathrm{jh}} = \frac{\exp(X'_h b_j)}{\sum_{i=1}^{q} \exp(X'_h b_i)} \quad (2)$$

This specification is chosen for its simplicity and the small information requirement for its estimation.[†] When longitudinal data with multiple spells are available, models of competing risks inspired from duration models would be appropriate (Lin, 1997; Omori, 1998). This is not the case with cross-section data such as those we dispose of and that are typically available in LDCs. With cross-section data, the MNL model, or any other multinomial probability model, has never been used to estimate the relative prevalence probabilities of symptoms (or diseases). What has been done is the estimation of the probability of incidence of a *given* disease or symptom, without considering the possibility of multiple health problems. This is usually carried out by using a simple logit model or a simple probit model (Becker, 1979; Prentice and Pyke, 1979; Whittemore, 1995; Breslaw and Holubkov, 1997; Zhao *et al.*, 1998). This is therefore the first study using MNL for modelling the prevalence of several symptoms. Naturally, if the situation of competing health risks is important, considering only one disease may produce biased estimations.

However, the advantage of considering several risks is limited by the fact that only relative probabilities can be identified in this model. Alternatively, one can identify all the probabilities described in Eqn (2) by fixing to zero the vector of coefficients of the reference alternative. The model will be estimated by using the method of maximum likelihood. Maximum likelihood estimates of the parameters b_i can be calculated by optimizing the log-likelihood that is globally concave. We now discuss the regressors incorporated in the model.

The independent variables

We include in X the available variables describing the host factors. These variables are the socio-demographic characteristics of the ill person and of the household, variables describing the economic situation or the individual behaviour likely to influence the health status, and the household health equipment. Unfortunately, the environment factors are unobserved. We control for their influence by including dummy variables for seven districts.

When looking at factors effects in many econometric studies, one must remember that they deal with the incidence of general illness or with the incidence of a specific disease or symptom in a particular population. In this paper, we study the *relative* prevalence of symptoms in a population of ill persons,[‡] which is slightly distinct,

[†]However, it has also unattractive features, notably the property of independence from irrelevant alternatives that implies that the ratios P_{jh}/P_{kh} for two symptoms j and k depend neither on parameters nor on attributes associated with other symptoms. Moreover, no correlation between the error terms ε_{ih} has been allowed. We attempted to estimate multinomial probit models to eliminate these shortcomings, but the discriminating information present in the data was insufficient to reach significant estimates. Under this strong constraint of information, we choose to restrict the analysis to the results of the multinomial logit model.

[‡]This makes the comparison with past published results delicate. Indeed, the comparison would be straightforward in the case of: a fixed structure of the population of ill persons, a fixed probability of 'Other Symptoms' in this population, a steady-state situation enabling us to assimilate incidence and prevalence with constant duration of disease, and no influence of diseases on one another. Thus, part of the differences between our results and, say, binary logit estimates of the incidence of a particular symptom, would be explained by the above distinctions in the estimated model and in the considered population. Other differences in results would stem from the fact that different sets of correlates may be used. However, it is interesting to verify that factors present in studies of absolute incidence of morbidity are also influential in our relative prevalence approach.

although the same type of factors can reasonably be considered. The age of the ill person is the main demographic variable, consistently with the medicine, epidemiological and health economics literatures, in which health processes are strongly dependent on age groups (Souhami and Moxham, 1994). Age should also affect the composition of diseases in a population of ill persons. For example, the risk of infantile and other illnesses often strongly decreases with child growth (see the cases of Rhinovirus infectious diseases in Rampey *et al.*, 1992; diarrhoea and illness of children in Heller and Drake, 1979; diarrhoea and respiratory diseases of infants in Cebu Study Team, 1992). To deal with these age effects, we not only separate the population of ill persons into four age groups, but we also introduce the age variable for each estimation based on a specific age group. In absolute incidence studies (Mausner and Kramer, 1985), females are associated with a lower relative prevalence of wounds than males. Accordingly, the gender of the ill person is incorporated in X.

Other socio-demographic characteristics matter. Marital status has been found to be associated with health status in studies of absolute incidence. Sickles and Taubman (1997) show that marriage has a positive effect on life-tables. Using Kenyan data, Gage (1997) shows that children of married mothers have a lower probability of polio dropout and of acute malnutrition. Mausner and Kramer (1985) insist that marital status is associated with lower mortality for both sexes. Several explanations have been proposed for the influence of marital status on health status. Single persons often have a more dissolute lifestyle. Marriage brings positive interactions between spouses (caring and companionship), although also stress and anxiety. Surviving widows or widowers experience acute grief provoking health problems and premature death. Finally, marriage is in itself a selection mechanism often excluding persons with chronic disease. We include dummy variables for the head's spouse and for single heads. Other marital statuses are unobserved.

Household composition influences health. Indeed, it is well known that crowding people indoors reinforces the occurrence of respiratory diseases (Sutton, 1981). This implies that, at constant dwelling area, large size households should have a higher incidence of these diseases and the same is true for most infectious illnesses. This is clearly not the only influence channel of household composition since in LDCs large household size is often correlated with high income or socio-economic status, but also with considerable economic burden. In that case, high income might induce both large household and large dwelling, and the effect of household size may not be interpretable only in terms of density. Moreover, health management decisions by the parents may incorporate the birth order, the number of children or the masculinity ratio of household, directing care and resources allotted to infants and young children (Mausner and Kramer, 1985; Strauss and Thomas, 1995). The number of active members is an indicator of the household capacity to meet its members needs both financially and in terms of caring time. Our available indicators describing household composition are the household size and the number of active members.

Economic variables also affect health status. Household income or assets, as in Heller and Drake (1979), Appleton (1995), Gage (1997), Smith (1998), are strongly correlated with absolute incidence of general or specific illnesses, the richer households being uniformly in better health for all symptoms. This may partly be the consequence of investments in housing improving house insulation from mosquitoes and parasites. Thompson *et al.* (1997) have shown in particular that the sole expedient of closing the eaves of houses in Mozambique significantly reduces the malaria risk. Richer households and landlords are more likely to undertake this type of housing investment. We dispose of a dummy variable for landlord and of the household per capita expenses, which account for the above considerations.

Not only income but also the occupations of the ill person and other household members are important determinants of health status (Mausner and Kramer, 1985; Sickles and Taubman, 1997; Cavelaars *et al.*, 1998). For example, Andersson and Bergstrom (1997) found that agricultural women in Central African Republic are associated with lower birth weight of children. We observe and incorporate dummy variables for corn producers and for pure agricultural households.

Health equipment is also influential. The use of latrines and septic tanks is generally associated with better health of household members (Cebu Study Team, 1992; Appleton, 1995). Smoke exposure has been found to be positively associated with febrile respiratory infections (Cebu Study Team, 1992). We include dummy variables for households disposing of a septic tank and for households using kerosene lighting (as opposed to firewood that generates smoke).

Finally, habits have been frequently invoked as explanations for morbidity (Sickles and Taubman, 1997). Alcoholism is known to increase the risks of vehicle accidents and liver diseases. Smoking is associated with higher incidence of lung cancer and stroke. Household alcohol expenses and tobacco expenses are included in the model. We are now ready to examine the data used for the estimation.

DATA DESCRIPTION

The country

Benin is a small rural country in western Africa with a population of 4.74 million in 1991, with 48% under 15 years old.[§] Per capita GNP is US$340, which makes Benin one of the poorest countries in the world. Agriculture is the cornerstone of Benin's economy and contributes 37% of the country's GNP. Moreover, three quarters of the active population is employed in the agricultural sector. The education level of the population is very low with around four-fifths of adult people illiterate.

Health status is also dramatically low. Average life expectancy at birth is estimated at 50 years. Children experience a heavy mortality toll with perinatal mortality rate equal to 6.9% and juvenile mortality rate equal to 17%. This high mortality is partly due to morbidity, especially from endemic diseases such as malaria, parasites and tuberculosis. Malnutrition is also widespread with 35% prevalence of malnutrition amongst children under 5 years old, and 10% of children showing weight insufficiency at birth.

However, despite the extent of poverty in the country, efforts have been made to improve the health of the population. Sixty-seven percent of children are vaccinated

[§]The following statistics are also given for 1991.

with their third dose of DCT and 34% of deliveries benefit from health assistance. Nonetheless, households spend on average only 5% of their final consumption expenditure on health. Moreover, only 4.3% of the GNP are devoted to health expenses with 41.8% of these health expenses being funded by international assistance. A better knowledge of factors associated with the main type of diseases would contribute to enhancing the efficiency of the too scarce resources allocated to health. We now turn to the sample of ill persons used for the estimation.

The ill person sample

The data are taken from a random health survey conducted by the government of Bénin covering the district of Ouidah, which has about 70 000 inhabitants, in the South East of Cotonou, from May to September 1992. This district is composed of nine communes where 2591 households were visited, corresponding to 11 502 individuals. 880 individuals reported having suffered an illness at the time of the collection.[¶] Owing to missing values, 786 observations of ill persons are used in the estimation (Bolduc *et al.*, 1996) describe this data. The household sample is representative of the district of Ouidah. The household members who reported an illness 2 weeks prior to the interview have been asked about their health and their socio-demographic and economic characteristics at the individual/household level. The interviews were conducted by doctors and medical staff from the Community Health Centre in Pahou. The medical knowledge of these enumerators suggests that the measurement of health status is more precise than usual.[†] The basis of enquiries is a mixture of self-report and medical examination. Symptom rather than disease has been recorded to avoid misreporting from individuals who self-treated for whom no specific disease diagnostic is available.

From this survey, we observe a sample of ill persons in households with illness. This implies that most information about other household members or other types of households is not available. Because we want to study competing symptoms for ill persons and not the incidence of symptoms in general, we do not incorporate any selectivity mechanism in the model. Another reason for it is that the selectivity mechanism cannot be identified from the available information.

The sample of ill persons is divided into four age classes in order to account for the specificity of health processes at different stages of the life cycle. 171 ill persons are babies (under 4 years old); 132 are young children (between 4 and 10); 71 are adolescents (between 11 and 18); and 412 are adults (over 18). Table 1 shows the mean and standard deviation of the main variables by age class of the ill persons. The average household has 5.08 members, of whom 38% are ill. Most of these members are children (4.25 by household) and the average age of ill persons is 27.3 years. The demographic and economic characteristics of households do not vary much with the age class of the ill person.[‡] By contrast, half the babies and children who are ill

[¶]The CREDESA health centre supervised the collection.

[†]See Cox and Cohen (1985) for the importance of observation problems in health surveys.

[‡]The age classes of ill persons correspond to different but not disjoint sub-populations of households since members of several age classes may belong to the same household.

Table 1. Descriptive statistics by age class of the ill person: 1: babies (1–3); 2: young children (4–10); 3: adolescents (11–18); adults (19+). Mean and standard deviation (in parentheses) are presented

Age class	1	2	3	4	Total
Household size	5.24	5.51	4.79	4.93	5.08
	(3.56)	(3.68)	(3.07)	(3.56)	(3.54)
Number of children	4.37	4.38	4.75	4.07	4.25
	(3.47)	(3.41)	(3.96)	(3.45)	(3.50)
Number of ill persons	1.18	1.39	1.25	1.10	1.18
	(0.49)	(0.71)	(0.65)	(0.37)	(0.51)
Per capita expenses (CFA)	3148	2816	3914	3157	3166
	(4712)	(3797)	(6312)	(4214)	(4487)
Total expenses (CFA)	10 356	10 255	11 009	10 502	10 474
	(3045)	(2669)	(2770)	(2739)	(2801)
Number of active persons	5.95	6.74	6.77	4.49	5.39
	(3.03)	(3.40)	(3.73)	(3.58)	(3.58)
Land area (arc[Q7])	8.94	7.18	21.46	6.96	8.74
	(24.65)	(20.18)	(97.41)	(23.90)	(36.90)
Proportion of ill persons	0.36	0.39	0.39	0.38	0.38
	(0.29)	(0.29)	(0.30)	(0.30)	(0.30)
Age (years)	1.68	6.60	14.79	46.74	27.30
	(0.76)	(1.98)	(2.50)	(18.59)	(24.69)
Gender female	0.50	0.50	0.58	0.60	0.56
	(0.50)	(0.50)	(0.50)	(0.49)	(0.50)
Education level (years)	0.00	0.43	0.87	0.55	0.44
	(0.00)	(0.50)	(0.58)	(0.75)	(0.66)
Duration of illness (days)	9.95	9.87	9.65	11.74	10.85
	(4.36)	(4.13)	(4.45)	(4.09)	(4.28)
Fever	0.51	0.67	0.62	0.37	0.47
	(0.50)	(0.47)	(0.49)	(0.48)	(0.50)
Cough	0.23	0.14	0.07	0.06	0.11
	(0.42)	(0.34)	(0.26)	(0.23)	(0.31)
Diarrhoea	0.08	0.01	0.00	0.00	0.02
	(0.27)	(0.09)	(0.00)	(0.07)	(0.14)
Wounds	0.02	0.05	0.10	0.07	0.06
	(0.15)	(0.22)	(0.30)	(0.25)	(0.23)
Abdominal pains	0.03	0.02	0.06	0.06	0.05
	(0.17)	(0.15)	(0.23)	(0.23)	(0.21)
Fatigue	0.01	0.01	0.00	0.04	0.02
	(0.11)	(0.09)	(0.00)	(0.19)	(0.15)
Articular pains	0.00	0.02	0.01	0.08	0.05
	(0.00)	(0.15)	(0.12)	(0.28)	(0.22)
Skin illness	0.04	0.00	0.00	0.02	0.02
	(0.18)	(0.00)	(0.00)	(0.14)	(0.13)
Teeth pains	0.02	0.00	0.03	0.02	0.02
	(0.13)	(0.00)	(0.17)	(0.13)	(0.12)
Cardiovascular illness	0.00	0.00	0.03	0.06	0.03
	(0.00)	(0.00)	(0.17)	(0.24)	(0.18)
Observations	171	132	71	412	786

Q7

are female and respectively 58% and 60% of adolescents and adults. Some symptoms are more frequent for specific age classes: cough, diarrhoea, skin diseases for babies; fever for children; fever, wounds, abdominal pains for adolescents; abdominal pains, fatigue, articular pains, cardiovascular illnesses for adults. The observed symptoms are described in the next subsection.

The symptoms

The average duration of illness is 10.85 days, although since this variable has been truncated to 15 days for the longest duration,† the mean underestimates the actual mean duration. Diseases are recorded in terms of symptoms rather than based on aetiology criteria that would have necessitated the knowledge of the specific agent of each disease. Classification of diseases on the basis of symptomatology was the traditional approach in Europe for a long time since enough aetiologic information was available (Mausner and Kramer, 1985). This approach is still useful for Africa where little detailed data are available. Table 2 shows the frequencies of the 33

Table 2. Frequency of the symptoms

Symptoms	Number of observations	%
Fever-convulsions-malaria	373	47.35
Cough-fever	87	11.11
Other symptoms	50	7.4
Wounds-abscess	46	5.9
Osteo articular pains	39	5.0
Abdominal pains	36	4.6
Cardiovascular illnesses	27	3.4
Anaemia-bleeding-fatigue	19	2.4
Diarrhoea	16	2.0
Skin illnesses	14	1.8
Dental pains	12	1.5
Eye pains	9	1.1
Oedema	8	1.0
Vomiting	5	0.6
Icterus	6	0.8
Measles	6	0.8
Throat ache-angina	5	0.6
Paralysis	5	0.6
Ear pains	4	0.5
Genito-urinal troubles	4	0.5
Rhinitis	2	0.3
Epilepsies-behaviour troubles	2	0.3
Dysentery	2	0.3
Hernia	1	0.1
Total	778	99.96

†It is usual to truncate this variable in health surveys because recalls of larger durations are considered too unreliable.

recorded symptoms. Most of the usual health problems are observed. Only three symptoms show a sufficient number of observations for econometric analysis: 'Fever', 'Cough' and 'Wounds'. They represent 61% of observations and we focus on their relative prevalence. We group all the other symptoms in a residual category 'Other Symptoms'.

'Fever' accounts for almost half of the symptom declarations. Fever (without cough) may often be attributed to malaria that is frequent in this area.[‡] The symptom 'Cough' is more difficult to attribute to specific diseases. It may be caused by dengue, influenza, tuberculosis, infantile illnesses, etc. 'Wounds' may also have various origins related to working activities, violent situations or hazardous activities.

The characteristics of the ill person, and of the household she/he belongs to, vary with the recorded symptom. Table 3 shows these characteristics for the four categories of symptoms. Age, as we have shown above, and gender of ill persons are all linked to the recorded symptoms. The average duration of illness is shorter for fever and cough than for wounds and other diseases.

The associations of symptoms with the observed individual or household characteristics suggest the possibility of discrimination between the relative prevalence of different symptoms. In the next section, we present estimates of a multinomial logit model describing this discrimination. This enables us to account for the multivariate interaction of correlates.

ECONOMETRIC RESULTS

The estimates

Table 4 shows the multinomial logit estimates. The interpretation of the coefficients is in terms of effect of the associated variable on the score of the considered symptom (Fever, Cough, Wounds) relative to the score of the reference symptom ('Other Symptoms'). Therefore, a positive parameter indicates a positive effect on the relative prevalence of the considered symptom with respect to the prevalence of 'Other Symptoms'. Dummies for districts have been included to account for health environment of variable quality, although the corresponding coefficients are not shown in the table since they are not significant.

Let us first consider the demographic variables. The differences in estimates for each age class and for the global sample supports the use of age classes for studying the relative prevalence of symptoms. Globally, younger ill persons are relatively more affected by fever and cough. Significant effects of age inside age classes show that when they grow older, ill babies are relatively more affected by fever and cough, which may be related to the higher occurrence of infantile diseases. Older children suffer relatively more from wounds, perhaps because of a greater autonomy of

[‡]Malaria is responsible for the greatest number of death in the world, especially for children. It is transmitted by bites of certain species of infected anopheles, and the germs are becoming increasingly resistant to usual treatments. Gomes (1993) stresses that few reliable data on malaria morbidity exist, particularly in sub-Saharan Africa. The ill health burden and the economic burden caused by malaria is a major source of poverty (Bonilla and Rodriguez, 1993).

Table 3. Mean and standard deviation (in parentheses) of the main variables by symptom

	Symptom				
	Fever	Cough	Wounds	Others	Total
Age (years)	20.27 (20.27)	15.86 (21.12)	32.17 (25.80)	39.43 (25.77)	27.30 (24.69)
Gender	0.56 (0.50)	0.46 (0.50)	0.37 (0.49)	0.61 (0.49)	0.56 (0.50)
Educated ill person	0.28 (0.45)	0.11 (0.32)	0.37 (0.49)	0.29 (0.45)	0.27 (0.44)
Education level (years)	0.48 (0.69)	0.30 (0.55)	0.61 (0.68)	0.41 (0.64)	0.44 (0.66)
Household head	0.29 (0.46)	0.20 (0.40)	0.48 (0.51)	0.52 (0.50)	0.37 (0.48)
Duration of illness (days)	9.81 (4.30)	9.63 (4.00)	11.91 (4.40)	12.43 (3.78)	10.85 (4.28)
Household size	5.16 (3.57)	5.43 (3.99)	5.59 (3.49)	4.80 (3.34)	5.08 (3.54)
Number of children	4.34 (3.43)	4.39 (3.78)	4.50 (3.03)	4.04 (3.58)	4.25 (3.50)
Head's spouse	0.08 (0.28)	0.06 (0.23)	0.02 (0.15)	0.14 (0.35)	0.10 (0.30)
Number of active persons	5.57 (3.57)	5.31 (2.90)	6.13 (4.44)	5.06 (3.62)	5.39 (3.58)
Landlord	0.48 (0.50)	0.57 (0.50)	0.39 (0.49)	0.44 (0.50)	0.47 (0.50)
Single head	0.05 (0.22)	0.08 (0.27)	0.04 (0.21)	0.06 (0.23)	0.06 (0.23)
Head's sex	0.35 (0.48)	0.36 (0.48)	0.39 (0.49)	0.36 (0.48)	0.36 (0.48)
Head's age (years)	44.35 (15.47)	45.41 (15.57)	42.07 (13.51)	44.61 (17.03)	44.43 (15.94)
Number of ill persons	1.16 (0.44)	1.37 (0.84)	1.22 (0.42)	1.14 (0.46)	1.18 (0.51)
Proportion of ill persons	0.36 (0.29)	0.39 (0.30)	0.35 (0.29)	0.39 (0.31)	0.38 (0.30)
Septic tank	0.38 (0.49)	0.40 (0.49)	0.28 (0.46)	0.43 (0.50)	0.39 (0.49)
Kerosene lighting	0.82 (0.38)	0.77 (0.42)	0.78 (0.42)	0.86 (0.34)	0.83 (0.38)
Alcohol expenses (CFA)	430 (1512)	443 (990)	749 (2416)	411 (997)	443 (1373)
Tobacco expenses (CFA)	110 (962)	23 (121)	100 (454)	33 (146)	72 (679)
Health expenses (CFA)	1966 (9969)	945 (1905)	1338 (3371)	1397 (3627)	1613 (7277)
Per capita expenses (CFA)	3036 (3998)	2767 (3038)	3290 (4743)	3442 (5358)	3166 (4487)
Total expenses (CFA)	10 528 (2670)	10 249 (2811)	10 225 (3070)	10 514 (2930)	10 474 (2801)
Land area	10.98 (48.22)	7.07 (20.54)	3.37 (7.46)	7.15 (24.02)	8.74 (36.90)
Corn producer	0.27 (0.45)	0.21 (0.41)	0.35 (0.48)	0.28 (0.45)	0.27 (0.45)

Table 4. Multinomial logit estimates

Explanatory variables	Age under 4			4 to 10			11 to 18			Over 18			Global sample		
	F	C	W	F	C	W	F	C	W	F	C	W	F	C	W
Constant	2.23	2.12	−154.0	−0.290	−1.33	1.86	18.0[a]	1.58	−1.24	4.33[a]	0.245	2.11	3.45[a]	3.75[a]	0.744
	(1.75)	(2.03)	(17 600.0)	(2.28)	(3.14)	(7.25)	(7.10)	(1590)	(1370)	(1.34)	(2.39)	(2.24)	(0.90)	(1.35)	(1.58)
Age	0.673[a]	0.739[b]	0.795	0.277	0.277	0.785[b]	−0.294	−0.771[a]	−0.156	−0.0342[a]	0.00 572	−0.0210	−0.0264[a]	−0.0326[a]	−0.00 730
	(0.32)	(0.39)	(2910)	(0.194)	(0.24)	(0.47)	(0.27)	(0.37)	(0.40)	(0.0083)	(0.015)	(0.013)	(0.0050)	(−0.0085)	(0.0090)
Female	0.581	−0.0355	−43.1	−0.307	−0.633	−0.607	−0.383	1.05	−7.88[b]	0.0512	−0.493	−0.346	9.22	−0.137	−0.657[b]
	(0.49)	(0.60)	(2920)	(0.68)	(0.91)	(2.18)	(1.20)	(2.65)	(4.50)	(0.28)	(0.55)	(0.47)	(0.19)	(0.29)	(0.35)
Head's spouse										−0.624[b]	−0.462	−2.32[a]	−0.447	−0.778	−1.77[b]
										(0.36)	(0.77)	(1.11)	(0.29)	(0.55)	(1.06)
Single head										0.447	2.59[a]	0.260	0.297	1.09[b]	−0.232
										(0.57)	(0.87)	(1.18)	(0.42)	(0.57)	(0.84)
Number of active	0.113	−0.621	−18.6	−1.09[a]	−3.49[a]	1.97	0.282	−1.19	1.51	−0.508[a]	−0.0812	0.231	−0.254	−0.604[b]	0.401
persons	(0.43)	(0.55)	(3480)	(561.0)	(1.22)	(1.89)	(1.17)	(3.37)	(2.59)	(0.22)	(0.39)	(0.32)	(0.18)	(0.28)	(0.27)
Household size	0.433	0.128	−51.2	0.155	0.200	−0.501	−2.09[a]	−0.759	−1.46	−0.0276	0.331	0.318	0.0364	0.137	0.243
	(0.36)	(0.41)	(2390)	(0.46)	(0.64)	(1.48)	(0.90)	(1.31)	(2.29)	(0.20)	(0.36)	(0.31)	(0.14)	(0.20)	(0.24)
Septic tank	−0.900[b]	−1.26[b]	39.6	−0.652	−0.278	−0.753	−1.62	14.7	14.9	−0.705[b]	−0.684	−1.55[a]	−0.481[b]	−0.691[b]	−1.06[a]
	(0.56)	(0.65)	(7700)	(0.81)	(1.03)	(1.98)	(1.53)	(1590)	(1370)	(0.38)	(0.71)	(0.63)	(0.29)	(0.41)	(0.49)
Kerosene lighting	−1.0[b]	−1.49[a]	32.0	−1.45	−3.11	−0.178	−1.21	−0.306	2.78	−0.148	−0.0297	−0.200	−0.230	−0.694[a]	0.0277
	(6.22)	(0.74)	(2760)	(0.80)	(1.25)	(1.97)	(1.24)	(2.12)	(2.77)	(0.31)	(0.62)	(0.54)	(0.22)	(0.35)	(0.39)
Beverage expenses	−0.374	−0.00486	5.480	0.403	−0.076	−4.51	−1.44[a]	−4.98	1.45	0.0158	−0.0456	0.00 0431	−0.0929	−0.163	−0.0316
(CFA)	(0.27)	(0.24)	(2840)	(0.71)	(0.97)	(4.15)	(0.50)	(3.44)	(1.15)	(0.095)	(0.024)	(0.16)	(0.068)	(0.13)	(0.11)
Tobacco expenses	2.04	−0.204	0.117	4.30	4.25	−0.0937				0.0560	−0.210	0.0852[b]	0.0599[b]	0.128	0.060[b]
(CFA)	(0.03)	(0.18)	(658.0)	(120)	(120)	(172.0)				(0.039)	(0.26)	(0.045)	(0.31)	(0.063)	(0.035)
Per capita expenses	0.455	0.521	−8.62	1.78	3.32[b]	−14.3	−0.525	−3.40	0.721	−1.03[b]	−0.676	−0.0142	−0.236	0.213	−0.174
($\times 10^{-4}$) (CFA)	(0.79)	(0.87)	(9430)	(1.31)	(1.92)	(532)	(1.35)	(3.09)	(2.26)	(0.56)	(893.0)	(0.91)	(0.34)	(0.46)	(0.71)

Continues

Q8

Table 4. Continued

Landlord	0.203	1.94	4.24	0.361	−12.4	−13.9	−3.30*	−1.60	1.30	−0.265	−0.629	-1.17^{a}	−0.294	−0.476	-0.741^{b}
	(1.16)	(1.38)	(16 500)	(2.64)	(919)	(1130)	(1.64)	(3.33)	(2.59)	(0.29)	(0.58)	(0.59)	(0.21)	(0.33)	(0.43)
Specialized peasant	-2.75^{a}	-2.52^{a}	113.0	−1.51	-2.50^{b}	10.1^{b}	7.86	−1.50	3.69^{b}	0.894^{a}	-1.42^{a}	−0.620	-1.26^{a}	-1.28^{a}	−0.0486
	(0.79)	(0.92)	(11 200)	(1.12)	(1.57)	(6.30)	(7.41)	(11.0)	(20.0)	(0.35)	(0.71)	(0.65)	(0.25)	(0.37)	(0.48)
Corn producer	−1.42	−0.576	−45.3	0.485^{b}	0.844^{a}	1.25	1.46	−18.8	−25.1	-0.723^{b}	−12.2	−0.406	-0.800^{a}	−0.321	0.21[Q8]
	(0.93)	(0.94)	(3460)	(1.12)	(1.34)	(2.23)	(1.89)	(1530)	(838.0)	(0.433)	(154.0)	(0.79)	(0.32)	(0.439)	(0.5)[Q8]

[a]Significant at 5% level.
[b]Significant at 10% level. Standard erros are in parentheses.
F, fever; C, cough; W, wounds.

movement at older ages. Older adolescents suffer relatively less from cough and older adults relatively less from fever. The effect of age is significant inside each age class at least for one symptom, suggesting that age classes of smaller width would be appropriate.

The effect of gender is less strong. However, as expected, females are associated with a lower relative prevalence of wounds than males. In particular, among adolescents, who lead a more active life, wounds are significantly less relatively frequent among females. The effects of the education of the ill person is not significant and has been omitted as well as other minor insignificant variables. No information about parent's education is available.

The spouse of the head is always a woman in our sample. These wives experience relatively less frequently Fever and Wounds than other adults. Ill persons living in households with single heads suffer relatively less often from Cough. The signs of the coefficients indicate that 'Others Symptoms' are relatively more frequent for the head's spouse or for ill persons in households led by a single head. Clearly, the marital status is not neutral for the relative prevalence of symptoms of ill persons, although it is difficult to understand precisely the mechanisms that this involves without further information.

Adolescent ill persons in larger households have a lower relative frequency of fever. Moreover, the number of active persons in the household negatively influences the relative frequency of fever for children and adults, and of cough for children. Thus, the structure of disease prevalence is clearly dependent on the household composition, while, as anticipated, the complex socio-economic and biological phenomena behind this effect remain to determine.

Let us now consider the economic variables. It is noticeable that several of them do not systematically affect the relative prevalence of symptoms. This is not as surprising as it may seem. Indeed, it is not obvious *a priori* whether the relative frequency of one of the three main symptoms with respect to the set of residual symptoms should be higher or lower for rich households. However, what the results show is that the structure of the symptoms of the poor is not the same as that of the rich. High-income level, measured by per capita expenses, is associated with a lower relative frequency of fever among adults and a higher relative frequency of cough among children. When the household head is the landlord, another wealth indicator, the relative frequency of wounds is lower for adults, and the relative frequency of fever is lower for adolescents. A possible channel of these effects of socio-economic status are investments in housing that improve the insulation of the house from mosquitoes and therefore reduce the prevalence of malaria.

Economic activities also affect the relative prevalence of symptoms. Babies, children and adults in purely agricultural households have relatively less frequent fever and cough, and children and adolescents relatively more frequent wounds. The higher prevalence of wounds is consistent with the more likely occurrence of accidents during agricultural work. Adults and children in corn producer households have relatively less frequent fever, and children have relatively less frequent cough. This may be related to the fact that peasant households specialized in cultivation of corn do not systematically use irrigation and generally live on dry land. Thus, they avoid contamination by parasites present or associated with water, such as mosquitoes transmitting the malaria parasite. By contrast, this health hazard touches relatively more other professions such as fishermen. The rarer occurrence of fever and cough for these peasants may also be related to their low household density that diminishes contagion risks.

Household equipment plays a significant role. The presence of a septic tank in the household is associated with relatively less frequent fever among babies and adults, cough among babies, and wounds among adults. Here, fever and cough may be contained by reduced exposure to contaminated faeces that would transmit infections. Using kerosene for lighting is related to relatively less frequent fever and cough for babies and children. This is consistent with avoiding the noxious effect of firewood use on air quality.

Finally, consumptions that are noxious for health are found related to the relative prevalence of the specific symptoms. Large alcohol expenses are associated with lower relative frequency of fever among adolescents, perhaps because liver diseases are included in category 'Other Symptoms'. Indeed, the probability of 'Other Symptoms' may be higher for those individuals because of higher prevalence of liver diseases, and therefore the relative probability of fever (i.e. its probability divided by the probability of 'Other Symptoms') is lower. The coefficient of alcohol expenses is also positive for the probability of wounds among adolescent and adults, while it is not significant at 5% level. An expected relationship between alcohol and wounds among these age groups might be more significant with a large sample. Large tobacco expenses are linked with greater relative frequency of wounds among adults, which might be associated with the sometimes more dissolute lifestyle of smokers. The variable tobacco expenses cannot be introduced for the adolescent illness because of quasi-collinearity problems in this relatively small sub-sample (71 individuals). Let us now examine which consequences for health interventions can be derived from the estimation results.

Consequences for health care

The estimation results may be used to assist health care directed towards specific symptoms or specific groups. In particular, better screening of ill persons can provide effective means of disease treatment. Let us mention just a few important examples. Age appears as the major screening variable for the management of ill person populations. In particular young populations of ill persons are more affected by cough and fever, and priority should be given to treating diseases associated with these symptoms in these populations. Naturally, some health care interventions are directed towards specific age groups (children at school, infants during mothers' visits at health centres, adults at work). For populations of ill persons mostly composed of female patients, equipment for the treatment of injuries is less important than when facing male ill persons. In practice, age and gender are not used as much in Benin as they could be as screening information for prioritizing medical interventions. For example, health centres in areas with different demographic composition are generally designed in the same way. Also, little emphasis on medical support at work exist for adults in Benin, wasting an opportunity of screening the predominantly male adult ill workers.

The estimates indicate a relatively higher prevalence of fever and wounds for adults in poor households. Then, for this category of ill persons, medical interventions could focus more on these health problems, perhaps by enhancing the capacity of health centres located in poor areas to fight malaria. This can be achieved by constitution of floroquine stocks, drainage to destroy mosquitoes' breeding sites, insulation of houses.

Subsidies and technical advice towards the installation of septic tanks and the use of kerosene for lighting could be useful since relatively lower occurrences of fever and cough are simultaneously associated with the presence of such equipment. Some support for this equipment acquisition could be coordinated with typical health care interventions instead of being separately planned as is presently the case in Benin.

Among ill persons living in agricultural households, cough is relatively less frequent, while wounds are relatively more common. Here the equipment of remote rural health centres could be adjusted, for example by constituting stocks of plasters and disinfectant, and purchasing small surgery equipment. Unfortunately, far from being adapted to the local needs, the equipment of rural health centres in Benin is too often characterized by shortages in basic medicines and health apparatus. More effort is needed in this domain and this equipment effort could be systematically implemented accordingly to the needs of the local populations.

Naturally, more thorough studies would be necessary before undertaking such health policies and a technique that could now be applied to other data sets is proposed. In particular, it would be necessary to determine the importance of missing variables that may influence the results, and could be measured in a more specialized survey. The precise causality mechanisms explaining the statistical association should also be carefully investigated. To this extent, the results of multinomial logit estimation provide hints useful to direct more specific studies of the statistical links between diseases and factors. Then, an accurate examination of causality factors should be undertaken.

CONCLUSION

Several issues specific to medical interventions in developing countries have been barely treated in the economic and epidemiological literature. Firstly, health policy relates to disease, but in developing countries disease is often not diagnosed, rather the observation of symptoms provides the basis of many medical interventions. Generally, economic studies do not deal with symptoms, but deal with general indicators of health. Secondly, studies usually deal with only one disease when the affected population may have multiple diseases. Thirdly, sampling is generally done of ill persons rather than of the general population. Indeed, much of the available health information concerns ill persons observed during consultations by medical agents. Fourthly, only the main symptom is recorded despite many symptoms usually being present simultaneously. Even when some symptoms are not observable, there exist pre-clinical conditions corresponding to these latent symptoms, and at that time the person may be infectious.

To deal with these issues a statistical approach is proposed based on the consideration of competing scores of symptoms to explain the relative prevalence of symptoms among ill persons. Using data from Benin, a multinomial logit model of relative prevalence is estimated that exhibits the main socio-economic characteristics associated with the prevalence of fever, cough and wounds, relatively to other symptoms. The age of the ill person determines the pattern of symptoms to a large extent. Household composition, living standards and economic activities affect the relative prevalence of symptoms. Finally, health equipment and noxious consumption also play important roles.

Such knowledge regarding the correlates of the relative odds of symptoms is useful since the cost-effectiveness of the treatment of a symptom in a population of ill persons depends on the relative frequency of the symptom. Characteristics associated with the relative prevalence of a given symptom may indicate that the corresponding diseases must be treated as a priority if this characteristic is pervasively present in the population of ill persons. The estimates can be used to screen ill persons and adapt medical interventions to the expected symptoms.

Our approach is likely to help researchers to progress in dealing with the mentioned issues by providing a framework where they can be integrated. The correlates of symptoms for ill persons corresponding to what is generally observed in this context, and therefore what is used for medical intervention were modelled. The simultaneous occurrence of symptoms and for pre-clinical conditions were accounted for by specifying competing scores of symptoms.

However, important difficulties remain. Indeed, the extent of the available information cannot be ignored even if our model is adapted to its characteristics. The selectivity of the sample of ill persons cannot be seriously treated with our data. This means that it is implicitly assumed that this selectivity does not significantly affect the relative prevalence of symptoms. What would be needed to correct this problem is to observe a representative sample of the population, which is rare for the type of information we deal with. A second difficulty is the sample size. With a larger sample and more observed characteristics, better estimates would be possible, for example by defining smaller age cells for ill persons or by using correlates that are direct instruments in medical interventions. Still more could be done if several periods were observed. In that case, the incidence of symptoms as well as the prevalence could be modelled and dynamic models of competing risks could be estimated. Moreover, causality relationships could be more explicitly investigated. However, all these data requirements are rarely satisfied in LDCs and cannot justify rejecting the more limited available information.

ACKNOWLEDGEMENTS

This research has benefited from a grant from the Programme PARADI. I thank participants at a seminar at Oxford University, G. Lacroix, W. Morrison and D. Whynes for their comments. Of course responsibility for errors remains mine. I am grateful to the CREDESA for providing me with the data.

REFERENCES

Andersson R, Bergstrom S. 1997. Maternal nutrition and socio-economic status as determinants of birthweight in chronically malnourished African women. *Trop Med Int Health* **2**(11).

Appleton S. 1995. *The Impact of Public Services on Health Care and Morbidity: Evidence from Kenya.* Working paper CSAE 95–2[Q19] Q9

Asenso-Okyere WK, Dzator JA. 1997. Household cost of seeking malaria care: a retrospective study of two districts in Ghana. *Soc Sci Med* **45**:[Q2] Q2

Audibert M, Etard J-F. 1998. The impact of schistosomiasis in Mali. *J African Econ*.[Q3] Q3

Becker N. 1979. An estimation procedure for household disease data. *Biometrika*[Q4] 271–277. Q4

Behrman JR, Deolalikar AB. 1988. Health and nutrition. In *Handbook of Development Economics*, Vol. I, Chenery H, Srinivasan TN (eds). Chap. 14, Elsevier:[Q3] Q3

Bolduc D, Lacroix G, Muller C. 1996. The choice of medical providers in rural Bénin: a comparison of discrete choice models. *J Health Econ*[Q3] Q3

Bonilla E, Rodriguez A. 1993. Determining malaria effects in rural Columbia. *Soc Sci Med* **37**: 1109–1114.

Breslaw NE, Holubkov R. 1997. Weighted likelihood, pseudo-likelihood and maximum likelihood methods for logistic regression analysis of two-stage data. *Stat Med* **16**: 103–116.

Cavelaars AEJM, Kinst AE, Geurts JJM, *et al.* 1998. Morbidity differences by occupational class among men in seven European countries: an application of the Erikson-Goldthorpe social class scheme. *Int J Epidemiol* **27**: 222–230.

Cebu Study Team. 1992. A child health production function estimated from longitudinal data. *J Develop Econ* **38**: 323–351.

Cox BG, Cohen SB. 1985. *Methodological Issues for Health Care Surveys*. Marcel Dekker: New York.

Drasar BS, Tomkins AM, Feachem RG. 1981. Diarrhoeal diseases. In *Seasonal Dimensions to Rural Poverty*, Chambers R, Longhurst R, Pacey A (eds). Frances Pinter Ltd[Q5]. Q5

Feachem RGA, Kjellstrom T, Murray CJL, Over M, Phillips MA. 1992. *The Health of Adults in the Developing World.* Oxford University Press: Oxford.

Folland S, Goodman AC, Stano M. 1997. *The Economics of Health and Health Care*. Prentice Hall:[Q5]. Q5

Gage AJ. 1997. Familial and socio-economic influences on children's well-being: an examination of pre-school children in Kenya. *Soc Sci Med* **15**(12):[Q2]. Q2

Geoffard P-Y, Philipson T. 1995. The empirical content of canonical models of infectious diseases: the proportional hazard specification. *Biometrika* **82**: 101–111.

Gomes M. 1993. Economic and demographic research on malaria: a review of the evidence. *Soc Sci Med* **41**: 1093–1108.

Heller PS, Drake WD. 1979. Malnutrition, child morbidity and the family decision process. *J Develop Econ* **6**: 203–235.

Jones AM. 1999. Health econometrics. In *Handbook of Health Economics*, Newhouse JP, Culyer AJ (eds). North Holland[Q2]. Q2

Krueger DE. 1966. New enumerators for old denominators—multiple causes of death. In *Epidemiological Approaches to the Study of Cancer and Other Chronic Diseases*, Haenszel W (ed.). Natl Cancer Inst., Monogr. No. 19: Washington, DC; 431–443.

Lilienfeld DE, Stolley PD. 1994. *Foundations of Epidemiology.* Oxford University Press: Oxford.

Lin DY. 1997. Non-parametric inference for cumulative incidence functions in competing risks studies. *Statist Med* **16**: 901–910.

Maddala GS. 1983. *Limited-Dependent and Qualitative Variables in Econometrics.* Cambridge University Press: Cambridge.

Mausner JS, Kramer S. 1985. *Epidemiology—An Introductory Text.* W.B. Saunders Company: Philadelphia.

Omori Y. 1998. The identifiability of independent competing risks models with multiple spells. *Oxford Bull Econ Statist* **60**: 1.

Philipson T. 2000. Economic epidemiology and infectious diseases. In *Handbook of Health Economics*, Vol. I, Culyer AJ, Newhouse JP (eds). Elsevier S.B.V.

Prentice RL, Pyke R. 1979. Logistic disease incidence models and case-control studies. *Biometrika* **68**: 403–411.

Rampey H, Longini IM, Jr., Haber M, Monto AS. 1992. Discrete-time for the statistical analysis of infectious disease incidence data. *Biometrics* **48**: 117–128.

Salvato JA. 1982. *Environmental Engineering and Sanitation.* Wiley: Chichester.

Sickles RC, Iaubman P. 1997. Mortality and morbidity among adults and the elderly. In *Handbook of Population and Family Economics*, Rosenzweig MR, Stark O (eds). Elsevier:[Q6] Q6

Smith JP. 1998. Socioeconomic status and health. *American Economic Association Papers and Proceedings* **88**:[Q2] Q2

Souhami RL, Moxham J. 1994. *Textbook of Medicine*. Churchill Livingstone: Edinburgh.

Strauss J, Thomas D. 1995. Human resources: empirical modeling of household and family decisions. In *Handbook of Development Economics*, Vol. III, Behrman J, Srinivasan TN (eds). Chapter 34. Elsevier Science:[Q5] Q5

Strauss J, Thomas D. 1998. Health, nutrition and development. *J Econ Lit* **XXXVI**: 766–817.

Sutton RNP. 1981. Respiratory diseases. In *Seasonal Dimensions to Rural Poverty*, Chambers R, Longhurst R, Pacey A (eds). Frances Pinter Ltd:[Q5] Q5

Thompson R, Begtrup K, Cuamba N, *et al.* 1997. The Matola malaria project: a temporal and spatial study of malaria transmission and disease in a suburban area of Maputa. Mozambique. *Am J Trop Med Hyg* **57**:[Q2] Q2

Van de Ven, Van der Gaag J. 1982. Health as an unobservable: a MIMIC-model of demand for health care. *J Health Econ* **1**: 157–183.

Van de Ven WPMM, Hooijmans EM. 1991. The MIMIC-health status index. In *Econometrics of Health Care*, Duru C, Paelinck HP (eds).[Q7] Q7

Van Vliet RCJA, Van Praag BMS. 1987. Health status estimation on the basis of MIMIC-health care models. *J Health Econ* **6**: 27–42.

Whittemore AS. 1995. Logistic regression of family data from case-control studies. *Biometrika* **82**: 57–67.

Zhao LP, Hsu L, Holte S, Chen Y, Quiaoit F, Prentice RL. 1998. Combined association and aggregation analysis of data from case-control family studies. *Biometrika* **85**: 299–315.

Zweifel P, Breyer F. 1997. *Health Economics*. Oxford University Press: Oxford.

Author Query Form (HPM/686)

Special Instructions: Author please write responses to queries directly on Galley proofs and then fax back. Alternatively please list responses in an e-mail.

Q1: Author: Please provide kwd?
Q2: Author: pp.?
Q3: Author: Vol./pp.?
Q4: Author: Vol.?
Q5: Author: Place?
Q6: Author: Place? pp?
Q7: Author: OK?
Q8: Author: Matter seems to be missing?
Q9: Author: Publisher & Place?