International Journal of Infectious Diseases 16 (2012) e429–e435

Contents lists available at SciVerse ScienceDirect

International Journal of Infectious Diseases

journal homepage: www.elsevier.com/locate/ijid

Extrapolating respiratory tract infection incidences to a rural area of Ghana using a probability model for hospital attendance

R. Krumkamp [a,*], N.G. Schwarz [a], N. Sarpong [b], W. Loag [a], H. Zeeb [c], Y. Adu-Sarkodie [d], J. May [a]

[a] Bernhard Nocht Institute for Tropical Medicine, Hamburg, Germany
[b] Kumasi Centre for Collaborative Research in Tropical Medicine, Kumasi, Ghana
[c] BIPS - Institute for Epidemiology and Prevention Research, Bremen, Germany
[d] School of Medical Sciences, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

ARTICLE INFO

Article history:
Received 24 August 2011
Accepted 2 February 2012

Corresponding Editor: William Cameron, Ottawa, Canada

Keywords:
Respiratory tract infections
Children
Ghana
Incidence
Logistic models

SUMMARY

Objective: The aim of the current study was to extrapolate incidences for respiratory tract infections (RTI) using referral data from a local hospital in Ghana weighted by the individual likelihood of a hospital visit.
Methods: Diagnoses from children visiting a rural hospital in Ghana during August 2007 to September 2008 were recorded. A logistic regression model, based on a population study conducted within the hospital catchment area, was used to calculate the individual probability of clinic attendance and to extrapolate the number of recorded cases. Cumulative incidences for children living in the hospital catchment area were estimated.
Results: Upper RTI was the most common respiratory diagnosis, with an extrapolated incidence of 17 481 cases per 100 000 per year, followed by pneumonia with an incidence of 2496 per 100 000 per year. All diseases analyzed were most common in the first year of life.
Conclusions: In general the study results are in line with comparable studies. Several methodological issues biasing the results in different directions were identified. For example, opportunistic infections that are more often observed in hospital attendees are likely to be overestimated. However, the applied approach presents a tool for areas where disease monitoring systems are not established.

1. Introduction

Acute respiratory infections (ARIs) account for 22% of deaths in children worldwide, with pneumonia as the single leading cause of death.[1,2] Seventy percent of these deaths from ARI in childhood occur in African and Asian countries.[3] Even though many figures on the worldwide distribution of respiratory tract infections (RTIs) exist, the quality and the precision of these data differ tremendously between regions. In the heavily burdened developing countries in particular, data for a detailed picture on disease occurrence are lacking due to insufficient surveillance or disease monitoring systems.[4,5] Within the last few years, new strategies to estimate and extrapolate infectious disease rates have been developed. Much has been learned from the course of the HIV/AIDS pandemic, where new concepts were developed to obtain reliable data for countries with weak surveillance systems.[6] The hospital surveillance concept has frequently been applied to extrapolate HIV rates on the basis of HIV-positive women attending antenatal clinics. Here, the HIV prevalence among the general female population of reproductive age within a region is estimated by adjusting data observed among the antenatal clinic attendees. Via these groups, disease rates for the greater population can be derived.[7–9] However, this method has to account for several selection biases. Cases identified via the hospital screening approach are often not representative of the underlying population because they are more likely to have better access to health care services.[10–12] If these selection biases are left unaccounted for, the hospital data will lead to an unrepresentative picture of disease occurrence. A way to address these biases is to consider a cases' probability of attending the clinic when extrapolating disease rates for the underlying population.

The aim of this study was to estimate disease rates of different RTIs among children in Ghana's Ashanti Region, using case records from the Agogo Presbyterian Hospital (APH), and taking the individual probability of hospital attendance into account.

2. Methods

The study was performed in the area covered by the APH, Asante Akim North District, Ashanti Region, Ghana. This region has

* Corresponding author. Tel.: +49 40 42818 535.
E-mail address: krumkamp@bnitm.de (R. Krumkamp).

1201-9712/$36.00 – see front matter
doi:10.1016/j.ijid.2012.02.003

an estimated population of 142 400 inhabitants, spread over 1160 square kilometers. The area has a tropical climate and is mainly covered by secondary rain forest and cultivated land.[13]

2.1. Datasets

The current study was based on datasets from two studies: (1) a cross-sectional study conducted throughout the APH's catchment area (community survey) to assess the probability of a child visiting the clinic, and (2) a hospital survey to assess diagnoses made during the study period at the APH. Since the community survey is a random sample of individuals living in the hospital catchment area, it is assumed to be representative of the population of potential clinic attendees.

2.1.1. Community survey

A random sample of parents living in the catchment area of the APH were interviewed. The number of selected individuals for each village was proportional to the size of the village population according to the Ghana census for the year 2000.[14] Questions were asked about household characteristics, education, jobs and financial situation, and health behavior. Furthermore, parents were asked where they would seek medical advice in the case that their child was to show particular disease symptoms. The geographic position of the interviewee's residence was assessed via a GPS (global positioning system) receiver. The interviews were conducted between May and September 2008. The dataset comprises 7226 intreviews.

2.1.2. Hospital survey

Children up to 14 years of age, who were visiting the outpatient department or who were admitted to APH's pediatric ward between September 2007 and August 2008, were included in the hospital survey. Comparisons with the official hospital records showed that about 70% of the pediatric hospital attendees were included in the survey. For each child, all symptom-based diagnoses made were documented. Since the current analysis was focused on RTIs, diagnoses included were: upper respiratory tract infection (URTI), bronchitis, pneumonia, lobar pneumonia, and tuberculosis (TB). Questions applied in the community survey were also asked to obtain socioeconomic information from the hospital attendees. This was part of a sub-study, only applied to a fraction of the study participants (20.2%). In the case of a child repeatedly recorded with the same diagnosis, only the first observation was included in the calculations.

2.2. Study design

Different analytical steps were conducted for the study, which are outlined below. The analysis was conducted with SAS 9.2 (SAS Institute Inc., Cary, NC, USA).

2.2.1. Literature review on factors influencing hospital attendance

The aim of the literature review was to identify factors with a known influence on the utilization and accessibility of health care services in developing countries. Knowledge of these factors was necessary to set up a plausible logistic regression model for calculating probabilities of hospital attendance. The underlying literature search was conducted in MEDLINE with PubMed using the following search syntax: ((("Health Services/utilization"[Mesh:NoExp] OR "Health Services Accessibility"[Mesh:NoExp]) AND "Developing Countries"[Mesh]) AND "Review"[Publication Type:NoExp]).

The literature search was conducted on July 11, 2010 and was limited to review articles and to journal articles published in English in the last 15 years. Studies were excluded from the analysis if they focused on mental health, on treatment options for chronically ill patients (e.g., HIV cases), or on the access to disease prevention programs.

2.2.2. Constructing additional study variables

At the time of writing, no detailed maps or satellite images of the Ashanti Region were available for measuring the road distances between the interviewees' residences and the APH. Therefore, the straight-line distance (in kilometers) between these places was calculated from the recorded GPS coordinates using Vincenty's formula, which describes the distance between geographical coordinates, taking the earth's curvature into consideration.[15,16]

In the datasets, no reliable measures or single proxies for the individual socioeconomic status (SES) were available, therefore an SES index was constructed via a principal component analysis (PCA) using data on ownership of assets, individual living conditions, and the educational level of the interviewees.[17–19] The score was constructed with data from the community survey and finally applied to the individuals in the hospital survey to determine hospital attendees' SES. However, this was only possible for those hospital attendees who had complete socioeconomic data for inclusion in the formulae.

2.2.3. Bivariate analysis and regression modeling

The community survey was analyzed via bivariate analysis and multiple regression modeling to construct a model that describes the individual likelihood of clinic attendance. The model was finally applied to calculate the probability of clinic attendance for the cases identified in the hospital survey.

The variables identified in the literature review were considered for the model. The model regressand (i.e., dependent variable) was based on the study question regarding whether children with 'coughing for 1 week' or with 'severe breathing' would attend the APH. In a bivariate analysis, the association between variables considered for the regression model and the model regressand were calculated. Dichotomous variables were analyzed via cross tables (odds ratios (ORs) and the 95% confidence limits (CLs)); normally and non-normally distributed continuous variables were compared with *t*-tests and Kruskal–Wallis tests, respectively.

Because of the high quantity of missing data in the hospital survey, several logistic regression models with different sets of variables were constructed, to which the model results were applied. Thus, the constructed probability models had to represent the combinations of variables observed in the hospital survey. The individual probability (p_i) of clinic admission was calculated via the logistic function, as given in Equation 1.[20,21] By doing so, the probability of an individual with a particular variable combination attending the APH was estimated. The CLs were calculated based on the standard error of the regression models. The model construction was done content-wise, based on epidemiological evidence assessed in the literature review. To finally decide on the variables to be included in the probability models, the log-likelihood ratio test was used to compare the model fit of simpler models against more complex ones.[21]

$$E(Y_i|X_{i1}, \cdots, X_{ik}) = p_i = 1/(1 + \exp(-(\beta_0 + \beta_{i1}X_{i1} + \cdots + \beta_{ik}X_{ik}))) \quad (1)$$

2.2.4. Extrapolating infectious disease rates

The estimated probabilities were used to weight each hospital case in order to extrapolate the additional, yet unobserved cases in the underlying population. The weighted cases were accumulated in order to obtain the expected number of cases (C) (Equation 2). As mentioned above, the hospital survey has an estimated completeness (α) of 70%. In order to account for the missing diagnoses the