Brief Reports

Rapid Epidemiologic Assessment of Breastfeeding Practices: Probit Analysis of Current Status Data

by M. U. Ferreira,*** MD, M. A. Cardoso, ***** MSc, A. L. S. Santos,† MD, C. S. Ferreira,* MD, PhD, and S. C. Szarfarc,**†† PhD

* Departamento de Parasitologia, Instituto de Ciências Biomédicas da Universidade de São Paulo, São Paulo, Brazil
** Núcleo de Pesquisas Epidemiológicas em Nutrição e Saúde da Universidade de São Paulo, São Paulo, Brazil
*** Divisão de Nutrição, Instituto de Saúde, Secretaria de Estado da Saúde de São Paulo, São Paulo, Brazil
† Fundação de Assistência à Infância de Santo André, Santo André, Brazil
†† Departamento de Nutrição, Faculdade de Saúde Pública da Universidade de São Paulo, Brazil

Summary

We describe the use of probit analysis to estimate breastfeeding indicators from current status epidemiological data. A health centre-based sample of 2411 children aged 0–1 year was investigated in Santo André, a large town in the Metropolitan Area of São Paulo, southeastern Brazil. Mothers were interviewed during routine pediatric consultations and asked about their current infant feeding practices. Probit regressions were calculated by a public-domain microcomputer programme written by one of us. The median duration of total (i.e., exclusive plus partial) breastfeeding in this children’s sample, estimated as 108.8 days (95 per cent confidence interval: 95.5–123.2 days), is close to that recently reported in the city of São Paulo and nearby towns. However, the median duration of exclusive breastfeeding (28.9 days, 95 per cent CI: 17.9–38.3 days) is rather short when compared to recent estimates from this same region. Despite the nationwide efforts for promotion of exclusive breastfeeding, only 14 per cent (95 per cent CI: 12–17 per cent) of the infants were still being exclusively breastfed by 120 days of age. Therefore, a key feature of breastfeeding practices in this population sample, namely, the early introduction of supplementary foods, was identified by using simplified methods of data collection and analysis. This communication suggests that probit analysis of current status data may be further explored as a method for rapid epidemiologic assessment of breastfeeding practices.

Introduction

Diverse strategies of data collection and analysis have been used in epidemiological surveys of breastfeeding practices. Current status data are usually preferred, since simple yes/no questions (for instance, whether or not the infant is currently being breastfed) provide the information needed. Children’s age at the time of the interview can be calculated from the date of birth as recorded in birth certificates, health cards or any other official document. The advantage of this approach is that most memory errors and recall biases, which are commonly observed in retrospective information, can be avoided.

Acknowledgements

To Professor Carlos Augusto Monteiro (Departamento de Nutrição, Faculdade de Saúde Pública da Universidade de São Paulo) for scientific advice and critical reading of the manuscript.

Correspondence: Dr M. U. Ferreira, Departamento de Parasitologia, ICB II, Av. Prof. Lineu Prestes 1374, Cidade Universitária, 05508-900, São Paulo (SP), Brazil.
health centres in Santo André during a 3-month period were investigated. Despite the fact that the population under study does not constitute a random sample of children living in this town, substantial differences in relation to the whole population are not expected, due to the wide (>90 per cent) coverage of local public health services in this age group. Mothers were interviewed during routine pediatric consultations, and asked about their current infant feeding practices. Only yes/no questions were asked. Exclusively breastfed children were defined as those who were given no other liquid (including cow milk) or solid food in addition to human milk. Total breastfeeding included both exclusive and partial breastfeeding. Children's age at the time of the interview, measured in days, was calculated from the date of birth recorded in their health cards. Complete information was available for 2411 children.

A comprehensive description of the use of probit analysis in dose-response bioassays can be found elsewhere. The children were distributed over 12 age groups, regarded as independent 'batches' of subjects with the same age, corresponding to the average age calculated for each group. The observed proportions of breastfeeding children in each group were transformed into probits and a weighted linear regression of probits \( Y \) on ages \( x \) was calculated using the maximum likelihood iterative procedure described by Finney. The relation between probits and age is described by the equation \( Y = 5 + (x - m)/s \), \( m \) is the mean of the distribution of ages at weaning in a given population and \( s \) is its standard deviation. If this distribution follows the Gaussian model, the median will coincide with \( m \). Therefore, probit analysis is a parametric method which assumes the ages at weaning to be either normally distributed in the study population or susceptible to normalization by simple procedures such as log transformation. The probit transformations, the weighted linear regressions and the respective coefficients of determination \( (r^2) \) were calculated with the aid of a programme written in BASIC language by one of us (CSF) for IBM-PC compatible microcomputers. The same programme was used to calculate breastfeeding indicators, namely the median duration of both total and exclusive breastfeeding and the expected proportions of breastfed children at different ages, with the respective 95 per cent confidence intervals. Information about the programme may be obtained from the authors. Goodness of fit to the probit model was tested by the Kolmogorov–Smirnov statistic.

### Results

Table 1 presents the data set for probit analysis. The population sample was distributed over 12 age groups. The total number of children in each group, their average age, and the number of children being breastfed according to age were recorded. The age-specific prevalence rates of total and exclusive breastfeeding were fitted to the probit model, as shown in Fig. 1. Total breastfeeding data best fitted to the model after log transformation of ages, which suggests that, in this population, ages at weaning follow a log-normal rather than a Gaussian distribution. The respective results of Kolmogorov–Smirnov tests for goodness of fit were \( d = 0.017 \) for total breastfeeding \( (P > 0.20) \) and \( d = 0.021 \) for exclusive breastfeeding \( (P > 0.20) \). The respective regression equations were \( Y = 8.327 - 1.636 \times (r^2 = 0.95) \) for total breastfeeding (in this case \( x \) represents log-transformed ages) and

### Table 1

**Distribution of the children's sample (n = 2411) over 12 age groups and total number of children and of those breastfed in each age group**

<table>
<thead>
<tr>
<th>Age range (days)</th>
<th>Mean age of group (days)</th>
<th>No. of children</th>
<th>No. of breastfed children</th>
<th>No. of exclusively breastfed children</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-30</td>
<td>19</td>
<td>219</td>
<td>189</td>
<td>114</td>
</tr>
<tr>
<td>31-60</td>
<td>42</td>
<td>270</td>
<td>206</td>
<td>112</td>
</tr>
<tr>
<td>61-90</td>
<td>71</td>
<td>280</td>
<td>180</td>
<td>95</td>
</tr>
<tr>
<td>91-120</td>
<td>103</td>
<td>191</td>
<td>105</td>
<td>41</td>
</tr>
<tr>
<td>121-150</td>
<td>132</td>
<td>246</td>
<td>105</td>
<td>30</td>
</tr>
<tr>
<td>151-180</td>
<td>164</td>
<td>197</td>
<td>66</td>
<td>5</td>
</tr>
<tr>
<td>181-210</td>
<td>193</td>
<td>231</td>
<td>80</td>
<td>8</td>
</tr>
<tr>
<td>211-240</td>
<td>223</td>
<td>182</td>
<td>60</td>
<td>1</td>
</tr>
<tr>
<td>241-270</td>
<td>256</td>
<td>146</td>
<td>55</td>
<td>2</td>
</tr>
<tr>
<td>271-300</td>
<td>283</td>
<td>195</td>
<td>36</td>
<td>0</td>
</tr>
<tr>
<td>301-330</td>
<td>314</td>
<td>121</td>
<td>21</td>
<td>0</td>
</tr>
<tr>
<td>331-365</td>
<td>347</td>
<td>133</td>
<td>27</td>
<td>0</td>
</tr>
</tbody>
</table>
BRIEF REPORTS

Percentage of breast-fed children

0 50 100 150 200 250 300 350
Age (days)

Percentage of exclusively breast-fed children

0 50 100 150 200 250 300 350
Age (days)

Probit of breast-fed children

1.2 1.5 1.8 2.1 2.4 2.7
Log age (days)

Probit of exclusively breast-fed children

0 50 100 150 200 250 300 350
Age (days)

FIG. 1. Proportions (%) of breastfed children (a) and of exclusively breastfed children (b), and the corresponding linear regressions of probits on age values calculated for total (c) and exclusive breastfeeding (d). Diamonds represent empirical data, while continuous lines represent the values expected from the probit model.

K = 5.310−0.011x (r² = 0.97) for exclusive breastfeeding. The median duration of total breastfeeding in this population was 108.8 days (95 per cent confidence interval: 95.5–123.2 days), while the median duration of exclusive breastfeeding was as short as 28.9 days (95 per cent CI: 17.9–38.3 days).

Estimates of breastfeeding prevalence rates in different age groups could be obtained directly from the data presented in Table 1. However, these estimates would be based on the relatively small number of children included in each age group, and would represent 'period averages' rather than prevalences at specific age cut-off points. The probit model permits to estimate prevalence rates at specific ages, based on weighted linear regressions of the whole set of data (n = 2411), and therefore less subject to month-to-month fluctuations. We estimated by probit regression that only 14 per cent of the infants were exclusively breastfed by 120 days of age (95 per cent CI: 12.3–16.8 per cent) and 4 per cent of them were still exclusively breastfeeding by 180 days of age (95 per cent CI: 2.8–5.7 per cent). The respective prevalences of total breastfeeding were 47 per cent (95 per cent CI: 43.5–50.5 per cent) by 120 days and 35.8 per cent (95 per cent CI: 32.3–39.5 per cent) by 180 days of age.

Discussion

Probit analysis is a robust statistical method extensively used in dose–response bioassays. Calculations are easy to perform on personal microcomputers or programmable calculators. Moreover, confidence intervals can be calculated for each estimate. In epidemiological research, probit and logit regressions have been previously employed in cross-sectional surveys to determine the median age at menarche and at primary malaria infection. The ages at menarche in a large sample of girls were assumed to follow a Gaussian distribution, and probits fitted well the raw data, without any transformation. However, the distribution of ages at primary malaria infection tends to be skewed to the right, since some exposed people may exhibit a high degree of natural resistance against infection. Therefore, in this case log transformation of ages produced a better fit to the probit model. Recently, both log-probit and log-logit transformations were fitted to a large set of current status...
breastfeeding data from six countries, but substantially better goodness of fit was obtained with more complex non-parametric regressions and cubic splines. Nevertheless, since there is no theoretical basis to assume ages at weaning in a given population to follow a particular distribution, probit and other parametric models should be fitted to both raw and transformed data. Further analyses of this kind should support the choice of transformations adequate for each data set.

Since the start of the nationwide Brazilian programme for breastfeeding promotion, in March 1981, the median duration of total breastfeeding in the Metropolitan Area of São Paulo increased from 84.3 days (1981) to 109.3 days (1984–5) to 145.8 days (1987). Similar trend was detected in relation to exclusive breastfeeding: from 26.6 days (1981) to 63.8 days (1984–5) to 66.8 days (1987). The present study shows that the median duration of exclusive breastfeeding amongst the clientele of public health centres in Santo André remains close to that estimated in São Paulo before the breastfeeding programme began. Therefore, by using a simplified method of data collection and analysis, it was possible to detect a key feature of breastfeeding practices in this health-service-based sample, namely the early introduction of supplementary foods. This finding indicates the need for improved strategies of breastfeeding promotion to be addressed to this population. In conclusion, we suggest that probit analysis of current status breastfeeding data could be further explored as a method for rapid epidemiologic assessment to support decision-making at the local level, in the context of primary health care.

References