A Bayesian network approach to the study of historical epidemiological databases: modelling meningitis outbreaks in the Niger

A Beresniak,‡ E Bertherat,§ W Perea,∥ G Soga,∥ R Souley,∥ D Dupont∥ & S Hugonnet∥

Objective
To develop a tool for evaluating the risk that an outbreak of meningitis will occur in a particular district of the Niger after outbreaks have been reported in other, specified districts of the country.

Methods
A Bayesian network was represented by a graph composed of 38 nodes (one for each district in the Niger) connected by arrows. In the graph, each node directly influenced each of the “child” nodes that lay at the ends of the arrows arising from that node, according to conditional probabilities. The probabilities between “influencing” and “influenced” districts were estimated by analysis of databases that held weekly records of meningitis outbreaks in the Niger between 1986 and 2005. For each week of interest, each district was given a Boolean-variable score of 1 (if meningitis incidence in the district reached an epidemic threshold in that week) or 0.

Findings
The Bayesian network approach provided important and original information, allowing the identification of the districts that influence meningitis risk in other districts (and the districts that are influenced by any particular district) and the evaluation of the level of influence between each pair of districts.

Conclusion
Bayesian networks offer a promising approach to understanding the dynamics of epidemics, estimating the risk of outbreaks in particular areas and allowing control interventions to be targeted at high-risk areas.

Introduction

Throughout the African “meningitis belt”, epidemics of meningococcal disease have been reported since the disease was first described, in 1912.1 Every year, western African countries within the Sahelo–Sudanian band experience major outbreaks of meningococcal meningitis, each of which can affect up to 200000 people, most of them young children.2 Burkina Faso, Mali and the Niger, for example, are regularly hit by meningitis epidemics.3–7 Meningitis is characterized by high levels of seasonal endemicity, with large epidemics of meningococcal meningitis occurring cyclically.8 Each such epidemic typically starts in January and ends in late May.9 The three main pathogens causing bacterial meningitis in Africa are Neisseria meningitidis, Haemophilus influenzae type b and Streptococcus pneumoniae.4,9,10

As a major cause of morbidity and mortality in sub-Saharan Africa, meningitis merits specific control measures.11–13

Burkina Faso, Mali and the Niger have each already established a broad surveillance network for collecting data on cases of the disease.4,12,13 Since 1986, peripheral health centres in each of these countries have routinely collected data on suspected cases of meningitis and reported each such case to health districts for subsequent analysis at the national health ministry. The information collected in this surveillance now forms important epidemiological databases that could improve our understanding of the nature of meningitis epidemics and allow high-risk areas to be identified. Current strategies for controlling meningitis epidemics in sub-Saharan Africa are based on epidemiological, immunological and logistical considerations. The strategy currently advocated by the World Health Organization (WHO) consists of the early detection of epidemics, the treatment of cases with antibiotics, and mass vaccination to halt the outbreak (if possible, within 4–6 weeks of the epidemic threshold being reached).3,11,13,14 The early detection of meningitis epidemics is based mainly on the observation of the trends in weekly incidence.3,11,12,14,16–18

The effectiveness of mass vaccination as a strategy for the control of meningitis epidemics has been questioned.19 Given the sporadic nature of the outbreaks, the optimal use of vaccines to control both short-term epidemic and endemic meningococcal disease has been the subject of much debate.20 In particular, the results of several studies in Africa have shown that vaccination during outbreak situations is suboptimal, mainly because populations in resource-poor areas cannot be immunized rapidly enough.21 In addition, although rapid laboratory diagnosis is an essential component in the surveillance of meningococcal epidemics, as it allows decision-makers to select the most appropriate vaccine for mass vaccination,22–24 the resource-poor countries most affected by such epidemics struggle to achieve such diagnosis.25 A new conjugate vaccine that protects against serogroup A meningitis (the form involved in most outbreaks of meningococcal disease in Africa) was approved by the United States Food and Drug Administration in February 2010. Conjugate vaccines prevent both carriage and the transmission of bacteria from person to person.

Extensive descriptive analyses that have been performed on meningitis surveillance databases have provided important information on epidemic cycles, seasonality and the correlations between morbidity and certain co-factors, such as climatic parameters.2,24–26 While the seasonal and spatial patterns of the disease appear to be linked to climate, the mechanisms

---

* Data Mining International, Route de l’Aéroport 29–31, CP 221, 1215 Geneva 15, Switzerland
* Epidemic and Pandemic Alert and Response, World Health Organization, Geneva, Switzerland
* World Health Organization, Niamey, Niger
* Department of Statistics, Surveillance and Response to Epidemics, Ministry of Health, Niamey, Niger
* Data Mining America, Montreal, Canada.

Correspondence to A Beresniak (e-mail: aberesniak@datamining-international.com).

Submitted: 10 January 2011 – Revised version received: 5 December 2011 – Accepted: 5 December 2011 – Published online: 20 January 2012

A Bayesian network consists of a graph showing the probabilistic relationships between one or more variables of interest. Bayesian methods are particularly valuable whenever there is a need to extract information from data that are uncertain or subject to any kind of error or noise.\textsuperscript{27,28} When applied to the forecasting of epidemics, a Bayesian network can allow potential "dependence relationships", such as the probability that an outbreak will occur in one district after it has occurred in certain other districts, to be explored.

A Bayesian network is represented by a graph composed of nodes connected by arrows. Each arrow begins at a "parent" node and ends at a "child" node, with the "parent" directly influencing the connected "child" in some way. The degree of influence between each "parent" and "child" is a conditional probability that can usually be computed from the data. Each node of a Bayesian network has an associated conditional probability table. In an epidemiological analysis, the nodes might represent the districts of a country and the arrows and conditional probability table might show how a disease outbreak in one district is linked to the probability of a disease outbreak in another district. In the example shown (Fig. 1), if district A experienced an outbreak, the probabilities that districts B and C experienced outbreaks would be 85% and 15%, respectively.

Since January 1986, WHO has supervised the weekly collection of data on meningitis incidence in each district of the Niger and these data now form a substantial historical database. The data recorded include district name and population, week number and the reported numbers of suspected cases of meningitis and of deaths attributed to meningitis in that district and week. For the present study, the data from the database for the 14 calendar years from January 1986 to December 1999 were used to construct a Bayesian network, with a node for each of the 38 districts in the Niger. The weekly incidence thresholds set by WHO for epidemic meningitis (i.e. at least 10 cases per 100,000 inhabitants in a district with a population of at least 30,000 or at least five cases per 100,000 in a district with a smaller population) were used to give each district a weekly score of 1 (if an epidemic had occurred) or 0 (if no outbreak had occurred). The corresponding weekly incidence thresholds for "epidemic alert", which could be used for the early detection of a meningitis epidemic, are lower (at least five cases per 100,000 inhabitants and at least two cases per 100,000, respectively).

Two types of analysis were then performed, one (the "first-level" analysis) taking no account of the time taken for an epidemic in one district to influence the development of an epidemic elsewhere (ignoring the timing of the epidemics) and the other (the "second-level" analysis) assuming that an epidemic in one district would influence the development of an epidemic elsewhere 1, 2 or 3 years later. A maximum lag of 3 years was explored because this was both the median period in the cycles of meningitis outbreaks detected in the historical database and the estimated duration of protection resulting from a vaccination programme.

The results of the "first-level" analysis, performed with version 1.0 of the Discoverer software package (Bayesware, Milton Keynes, United Kingdom of Great Britain and Northern Ireland), were used to create a graph of the Bayesian network and included the relevant conditional probability table, showing the probabilities of meningitis outbreaks. Because of the limited capability of the Discoverer software, the analysis was restricted to the first 14 years of the meningitis database for the Niger. The "minimum description length" learning algorithm was used, with network

---

**Fig. 1.** Simple Bayesian network composed of nodes, arrows and conditional probabilities for the occurrence of a meningitis epidemic in each of a country's five hypothetical districts

**Fig. 2.** Bayesian network graph showing how each of the Niger's 38 districts is influenced by, and/or influences, the probability of a meningitis outbreak in at least one other district
Research

Bayesian modelling of meningitis outbreaks in the Niger

A Beresniak et al.

414

Table 1. Districts of the Niger, their populations and the numbers of districts influenced by them or influencing them in terms of meningitis epidemics between 1986 and 1999

<table>
<thead>
<tr>
<th>District</th>
<th>Population*</th>
<th>Code no.</th>
<th>No. of districts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Influencing</td>
</tr>
<tr>
<td>Agadez</td>
<td>93433</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Agueie</td>
<td>332106</td>
<td>2</td>
<td>12</td>
</tr>
<tr>
<td>Arlit</td>
<td>112705</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Bilma</td>
<td>21903</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Birni</td>
<td>416949</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>Boboye</td>
<td>299775</td>
<td>6</td>
<td>8</td>
</tr>
<tr>
<td>Bouza</td>
<td>328327</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>Dakoro</td>
<td>532006</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>Diffa</td>
<td>190890</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>Dogon</td>
<td>587137</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>Dosso</td>
<td>406357</td>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td>Filingue</td>
<td>464232</td>
<td>12</td>
<td>6</td>
</tr>
<tr>
<td>Gaya</td>
<td>299561</td>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>Goure</td>
<td>258540</td>
<td>14</td>
<td>2</td>
</tr>
<tr>
<td>Guidan</td>
<td>421752</td>
<td>15</td>
<td>4</td>
</tr>
<tr>
<td>Illela</td>
<td>308835</td>
<td>16</td>
<td>5</td>
</tr>
<tr>
<td>Keita</td>
<td>245826</td>
<td>17</td>
<td>2</td>
</tr>
<tr>
<td>Kollo</td>
<td>370308</td>
<td>18</td>
<td>3</td>
</tr>
<tr>
<td>Loga</td>
<td>157534</td>
<td>19</td>
<td>3</td>
</tr>
<tr>
<td>Madagoua</td>
<td>372043</td>
<td>20</td>
<td>2</td>
</tr>
<tr>
<td>Madarounfa</td>
<td>334562</td>
<td>21</td>
<td>2</td>
</tr>
<tr>
<td>Magarua</td>
<td>564915</td>
<td>22</td>
<td>5</td>
</tr>
<tr>
<td>Maine</td>
<td>176995</td>
<td>23</td>
<td>0</td>
</tr>
<tr>
<td>Maradi</td>
<td>165840</td>
<td>24</td>
<td>3</td>
</tr>
<tr>
<td>Matameye</td>
<td>288542</td>
<td>25</td>
<td>4</td>
</tr>
<tr>
<td>Mayahia</td>
<td>483998</td>
<td>26</td>
<td>4</td>
</tr>
<tr>
<td>Miriah</td>
<td>688910</td>
<td>27</td>
<td>4</td>
</tr>
<tr>
<td>N'goumi</td>
<td>70591</td>
<td>28</td>
<td>2</td>
</tr>
<tr>
<td>N'namiy</td>
<td>882236</td>
<td>29</td>
<td>1</td>
</tr>
<tr>
<td>Ouallam</td>
<td>328297</td>
<td>30</td>
<td>1</td>
</tr>
<tr>
<td>Say</td>
<td>265583</td>
<td>31</td>
<td>1</td>
</tr>
<tr>
<td>Tahoua</td>
<td>421399</td>
<td>32</td>
<td>1</td>
</tr>
<tr>
<td>Tanout</td>
<td>420242</td>
<td>33</td>
<td>0</td>
</tr>
<tr>
<td>Tchintabaraden</td>
<td>91261</td>
<td>34</td>
<td>0</td>
</tr>
<tr>
<td>Tera</td>
<td>488873</td>
<td>35</td>
<td>0</td>
</tr>
<tr>
<td>Tessoua</td>
<td>412240</td>
<td>36</td>
<td>0</td>
</tr>
<tr>
<td>Tilaberi</td>
<td>245934</td>
<td>37</td>
<td>0</td>
</tr>
<tr>
<td>Zinder</td>
<td>195831</td>
<td>38</td>
<td>0</td>
</tr>
</tbody>
</table>

* In 2006.

* See Fig. 3 (available at: http://www.who.int/bulletin/volumes/90/6/11-086009).

Table 2. Conditional probability table for a meningitis outbreak in the Boboye district of the Niger, 1986–1999

<table>
<thead>
<tr>
<th>Situation in:</th>
<th>Probability of outbreak in Boboye (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agadez: No outbreak</td>
<td>98.8</td>
</tr>
<tr>
<td>Agadez: Outbreak</td>
<td>67.1</td>
</tr>
<tr>
<td>Birni: No outbreak</td>
<td>90.0</td>
</tr>
<tr>
<td>Birni: Outbreak</td>
<td>3.8</td>
</tr>
</tbody>
</table>

Parameters estimated via the maximum-likelihood technique. The “second-level” analysis was performed using the data collected on meningitis in the Niger between January 1986 and December 2005, a β version of the Bayesia package (Bayesia SAS, Laval, France) and the Bayes Net and BNLAT toolboxes from the Matlab package (Mathworks Inc., Natick, United States of America).

Results

Between January 1986 and December 2005, 182 244 suspected cases of meningitis and 14 859 meningitis-related deaths were recorded in the Niger. The cumulative attack rate over the 20 annual cycles of meningitis that occurred over this period was 1919.6 (suspected) cases per 100 000 population. The corresponding mean annual attack rate was 96.0 cases per 100 000 population. The Niger experienced six major national-level epidemics of meningitis between January 1986 and December 2005, with, at times, more than 50 cases recorded per 100 000 population in a single week.

First-level analysis

The results of the first-level analysis, which ignored any time lag between epidemics, are summarized in Fig. 2, with arrows running from the influencing districts to the influenced districts. From this graph, the number of districts influenced by a given district and the number of districts influencing a specific district (Table 1), can be determined at a glance.

Conditional probabilities were calculated for all districts. As an example, Table 2 presents the conditional probabilities of an outbreak and of no outbreak in the Boboye district, according to the outbreak status of the two influencing districts: Agadez and Birni. These results indicate that, when epidemics occur in both Agadez and Birni, Boboye should be included in any vaccination strategy because the probability of an outbreak in this district is high (96.2%).

Second-level analysis

Fig. 3 (available at: http://www.who.int/bulletin/volumes/90/6/11-086009) illustrates the results of the second-level analysis. The large number of links between the nodes in the uppermost plot in this figure, which shows the links when no lag (“Year 0 to Year 0”) or a lag of 1, 2 or 3 years is used, make this plot difficult to interpret. The links become
clearer when they are plotted separately for lags of 1, 2 and 3 years, as in the other plots in Fig. 3. Unfortunately, this analysis did not allow the accurate calculation of conditional probabilities, mainly because observations with “no outbreak” status were too few.

Fig. 4 summarizes the results of the analysis of longitudinal surveillance data for those districts in which meningitis incidence reached the epidemic threshold in 2003 and 2004. Arrows again indicate the influencing relationships.

Discussion

The control of epidemic meningitis remains an unresolved problem in Africa, partly because the location of major outbreaks, which might be affected by relatively rare events, is so difficult to predict. The data analysed in the present study were collected over such a long period (up to 20 years) that they were probably affected by unpredicted rare events, such as unusual migratory patterns, armed conflicts and catastrophic dry seasons. In general, valid predictive statistics on such rare events cannot be derived from mathematical models, but data mining methods, when applied to clinical and epidemiological data, may allow previously unpredictable or unknown trends to be made apparent.10 In the Niger, such an approach may well help to elucidate the factors that influence or trigger meningitis outbreaks. Unfortunately, historical epidemiological databases like the one used in the present study are very rare in the context of epidemic diseases, especially in developing countries. While descriptive analyses constitute a standard approach, they do not allow optimal responses to future major public health issues to be planned.

In the present study, although the first-level analysis allowed the conditional probabilities between the influencing and influenced districts to be estimated, it made no allowance for the time needed for the inter-district influences to reveal themselves as changes in meningitis incidence. Unfortunately, when lags of 0, 1, 2, or 3 years were considered in the second-level analysis, too few “no-outbreak” observations were available for an accurate estimation of conditional probabilities. Nevertheless, both methodological approaches appear promising in the planning and timely deployment of vaccination programmes, especially given the limited duration of the protection offered by current vaccines.

A validation test of the Bayesian network model was recently carried out using a 6-year time horizon, with observed surveillance data from 2006, 2007 and 2008 compared with expected model predictions from 2004, 2005 and 2006 (data not shown). Districts reaching epidemic alert or epidemic thresholds in 2004, 2005 and 2006 were considered as potentially influencing other districts in 2006, 2007 and 2008. The proportion of the outbreaks observed in 2006–2008 that were successfully predicted using the data collected in 2004–2006 was 63% when the epidemic threshold was used for the influencing districts and 91% when the alert threshold was used. This promising performance justifies the continued interest in such modelling for guiding immunization planning, including the extension of vaccination campaigns to “influenced” districts when an “influencing” district has reached the epidemic threshold.

More sophisticated approaches should be explored. For example, it would be interesting to analyse the collected data year by year and to build a Bayesian network for each year. Re-sampling simulation techniques could be used to simulate and replicate observations, therefore allowing the construction of a Bayesian network for each simulation, together with a conditional probability table.

In conclusion, a Bayesian network approach offers an innovative and promising technique for extracting meaningful and clinically useful information from a large historical epidemiological database. As suggested by the present study, this method can improve our understanding of the dynamics of epidemic outbreaks, help make health interventions more effective and optimize resource allocations. Unfortunately, the method relies on large volumes of longitudinal records that are rarely available in the context of epidemic diseases. In this respect, the historical database on meningitis in the Niger provided a rare data set that made the testing of this promising approach possible. Unfortunately, no operational conditional probabilities regarding meningitis outbreaks could be determined when a time lag was included. However, further research allowing Bayesian networks to be combined with re-sampling techniques, would allow this issue to be addressed, to the potential benefit of those countries where relevant large epidemiological databases are unavailable.

Acknowledgements

We thank Michel Lamure, from Claude Bernard Lyon 1 University (France), for his important contributions.

Competing interests: None declared.

Fig. 4. Longitudinal surveillance data listing the districts of the Niger where meningitis reached an epidemic threshold in 2003 and/or 2004

Malzakh

تبيِج شبكة بايزي لدراسة قواعد بيانات الأمراض الوبائية التاريخية: نمذجة حالات تفشي التهاب السحايا في النجیر كان لكل عقدة تأثير مباشر على كل العقدة التابعة للموضوعة في نهاية الأسهم المجرفة من تلك العقدة، وفق الاحتياطات الشرطية، و تتضمن الاحتياطات بين المناطح “المؤثرة” و “المتأثرة” للتحليل بواسطة قواعد البيانات التي احتوت على سجلات أسبوعية حالات تفشي التهاب السحايا في النجیر في عامي 2003 و 2004. في الرسم البياني، تم تمثيل شبكة بايزي برسم بواين مكون من 38 عقدة (واحدة لكل منطقة في النجیر) موصولة بأسهم. وفي الرسم البياني،
Bayesian modelling of meningitis outbreaks in the Niger

A Beresniak et al.

目的
制定一种评估尼日尔在指定的其他各个地区报告发生脑膜炎之后特定地区脑膜炎疫情的风险的工具。

方法
贝叶斯网络由以箭头连接的38个节点(一个节点表示尼日尔的一个地区)组成的图形表示。根据条件概率,在每个图形中,每个节点直接影响从此节点发出的箭头末端的每个“子”节点。通过分析保存尼日尔在1986至2005年之间每周脑膜炎疫情记录的数据库来评估“影响”和“受影响”地区之间的概率。对于所关注的每一周,为每个地区给定布尔型变量分数1(如果该周内该地区脑膜炎发生率达到流行病阈值)或0。

结果
贝叶斯网络方法能提供重要的原始信息,可用于识别影响其他地区(和受特定地区影响的地区)脑膜炎风险的地区以及评估每一对地区之间的影响等级。

结论
贝叶斯网络作为了解流行病动态、评估特定区域流行病疫情的风险并实现针对高风险区域的干预控制的方法前景广阔。
Resumen

Un enfoque de red bayesiana para el estudio de bases históricas de datos epidemiológicos: modelo de brotes de meningitis en Niger

Objetivo
Desarrollar una herramienta para evaluar el riesgo de la aparición de un brote de meningitis en un distrito determinado del Niger después de que se haya informado acerca de otros brotes en otros distritos específicos del país.

Métodos
Se representó una red bayesiana con un gráfico compuesto por 38 nodos (uno para cada distrito en el Niger) conectados mediante flechas. En el gráfico, cada nodo influye directamente en cada uno de los nodos de "niños" que se encuentran en los extremos de las flechas que surgen de dicho nodo, de acuerdo con las probabilidades condicionales. Se calcularon las probabilidades entre distritos "influyentes" e "influidos" mediante el análisis de bases de datos que recogían registros semanales de los brotes de meningitis en Niger entre 1986 y 2005. Por cada semana de interés, se adjudicaba a cada uno de los distritos un valor de variable booleana de 1 (si la incidencia de meningitis en el distrito alcanzaba un umbral epidémico en esa semana) o de 0.

Resultados
El enfoque de red bayesiana proporciona información importante y original, lo que permite identificar cuáles son los distritos que influyen en el riesgo de meningitis de otros distritos (y cuáles están bajo la influencia de cualquier distrito determinado) y evaluar el nivel de influencia entre cada par de distritos.

Conclusión
Las redes bayesianas ofrecen un enfoque prometedor para entender las dinámicas de las epidemias, permiten calcular el riesgo de brotes en áreas determinadas y concentrar las intervenciones de control objetivo en las áreas de alto riesgo.

Referencias

Fig. 3. Diagrams illustrating how a meningitis outbreak in a given district of the Niger influenced a similar outbreak in one or more other districts.

Note: The diagrams are based on weekly incidence data collected between 1986 and 2005. The numbers 1 to 38 indicate the districts in the Niger. The top plot illustrates the links with time lags of 0, 1, 2 or 3 years while the other plots show, separately, the links with no time lag (“Year 0 to Year 0”) and lags of 1 year (“Year 0 to Year 1”) and 3 years (“Year 0 to Year 3”).