SPATIAL-TEMPORAL ANALYSIS OF ROSS RIVER VIRUS DISEASE PATTERNS IN QUEENSLAND, AUSTRALIA

MICHÈLE L. GATTON, LOUISE A. KELLY-HOPE, BRIAN H. KAY, AND PETER A. RYAN
Australian Centre for International and Tropical Health and Nutrition, Queensland Institute of Medical Research, Herston, Queensland Australia; University of Queensland, Brisbane, Queensland, Australia

Abstract. Ross River virus is the most common vector-borne disease in Australia, with the majority of notifications being in Queensland. This study describes a retrospective spatial analysis of Queensland Ross River virus disease notifications spanning a 10-year period. Notifications were mapped to the local government area (LGA) of the residence of the patient. Ross River virus disease outbreaks within each LGA were detected by applying a Poisson model. Estimates of the seasonal incidence rates indicated wide variation between seasons and LGAs. Positive spatial autocorrelation between LGAs experiencing outbreaks indicated that LGAs within the same region often experience outbreaks at the same time. A hierarchical cluster analysis of the outbreak data was used to group LGAs with similar spatial and temporal outbreak patterns. This analysis highlights the variability in Ross River virus disease notification rates across Queensland, and provides a robust method for identifying disease outbreaks.

INTRODUCTION

Ross River (RR) virus is the most common mosquito-borne virus in Australia, and accounts for approximately 70% of all human mosquito-borne disease cases each year.1 Between 1991 and 2000, the number of notifications of RR virus disease in Australia ranged between 3,151 and 7,783, with corresponding disease notification rates between 14.6 and 42.5 per 100,000.1 Approximately 60% of the notifications in Australia occur in Queensland.2

The ecology of RR virus is complex, involving transmission cycles between various vertebrate species and a variety of mosquito vectors.3 In the human host, RR virus infections may result in clinical disease characterized by polyarthritis, rash, myalgia, and lethargy. The virus is able to multiply in a variety of non-human vertebrate hosts,4 with different hosts being identified as important in rural and urban areas.5,6 Along with regional variations in the prevalence of the vertebrate hosts, different mosquito species appear important in different areas.7,8 These factors suggest that the ecology of the virus and resultant pattern of human infection differ between regions. Studies in two major population centers in the southeast region of Queensland confirm this with considerable heterogeneity in RR virus disease risk.9,10

Analysis of the incidence of RR virus disease across all of Queensland has never been comprehensively assessed, although disease incidence has been compared between selected locations.11−13 Generally, tropical locations have tended to have higher RR virus disease incidence than subtropical regions.12 Similarly, coastal locations have been reported to have higher RR virus disease notification rates compared with inland towns.13 Numerous investigators have explored the relationship between the incidence of RR virus disease and variables such as rainfall, temperature, humidity, sea level, Southern Oscillation Index, and Quasi-Biennial Oscillation.13−19 Some of these studies demonstrated a relationship between RR virus disease incidence and climatic factors, but development of predictive regional models16,18 is likely to be of greater use to public health authorities, provided they have good sensitivity and specificity. Here we report the first regional examination of the incidence of RR virus disease throughout Queensland over a 10-year period. We explore the spatial distribution of the disease and provide a robust method for identifying epidemics in local areas. This is the first step in identifying conditions favorable for outbreaks of RR virus disease in Queensland.

METHODS

Study area. Queensland, the second largest state in Australia, is located in the northeastern corner of the country. This state has an area of 1,727,200 km², with widely varying environments ranging from tropical rainforest to drought-prone inland plains. Most of the population of Queensland inhabit the southeast corner of the state and eastern coastline. During the 10-year period considered in this study (1991–2001), the population increased by approximately 20% from 2.96 million to 3.57 million.20 Queensland is separated into 125 administrative local government areas (LGAs), with populations ranging in size from 236 to more than 882,000.

Notification data. Ross River virus disease notification data from June 1991 to May 2001 were obtained from the Communicable Diseases Section of Queensland Health. A notification was reported if serologic testing indicated a four-fold change in antibody titer between paired acute and convalescent sera, or if IgM and IgG antibody levels against RR virus were consistent with acute infection. Each complete notification included place of residence (location plus street/road), date of onset, and age and sex of the patient. Notification data was geocoded by the Queensland Department of Local Government and Planning into Statistical Local Areas using the 2001 Australian census boundaries, and later grouped into LGAs. Patient data were grouped into three age categories; 0–29 years, 30–59 years, and ≥ 60 years. Information relating to the number of people residing in each LGA in each year from 1991 and 2001 was provided by the Australian Bureau of Statistics.

All annual incidence rates (IRs) were based on the year starting in June and ending in May of the next calendar year. This classification was used to avoid splitting the peak transmission season that occurs from December to April. Incidence rates were calculated for winter (June to August), spring (September to November), summer (December to February), and autumn (March to May).

Long-term incidence rates. Due to the high variability and skewed distribution of seasonal IRs of some LGAs over the study period, a long-term IR that was representative of nor-
nal transmission was estimated for each season using either the crude IR or median, dependent on the LGA population size and distribution of the data in the 10-year period. For LGAs with less than 2,500 residents, the crude seasonal IR for the data of the 10-year period was used. For LGAs with larger populations, the crude IR was used if 1) the maximum number of notifications in the LGA for the season of interest was less than five, or 2) if the crude IR and trimmed IR differed by less than 20% (that is, (crude IR – trimmed IR)/trimmed IR < 0.2). The trimmed IR was calculated by omitting the years having the smallest and largest number of notifications from the calculation of the IR. If neither of these conditions were met, then the median IR was used. If the median was zero, the trimmed IR was substituted.

Seasonal age/sex standardized rates for each LGA were calculated using direct standardization,21 with the 10-year state population (1991–2001) acting as the base population. Where the long-term LGA IR was represented by the median IR, the age/sex rates for each LGA were calculated as IR_{i,j} = (c_{i,j}/p_{i,j}) × median IR (crude IR), where c_{i,j} and p_{i,j} are the total number of notifications and population in age group i and sex group j respectively for the 10 years considered. The correction factor (median IR/crude IR) ensures that applying the age/sex rates back to the base population results in the median IR being obtained.

Determining extreme events. To identify periods with an abnormally high number of RR virus disease notifications within LGA k, the number of notifications in time period m (n_{k,m}) was assumed to follow a Poisson process: n_{k,m} ∼ Poisson(E_{k,m}), where E_{k,m} is the expected number of cases within LGA k applying the long-term age/sex specific rates for the LGA and season of interest (internal standardization). Seasons within each year were defined as extreme if there was a < 1% chance of obtaining the observed number of notifications based on the long-term IR for the specific LGA. This definition serves as a mathematically robust method of identifying extreme events. However, when the expected number of notifications is small (< 1), LGAs may be declared as extreme if they have two or three notifications. Although these situations are mathematically extreme, it is unlikely that they would be considered as outbreaks by public health officials. Therefore, LGAs having less than one expected notification within a season were only declared as extreme if at least five notifications were actually received.

Identifying outbreaks. An outbreak was defined by the occurrence of at least one season with an extreme number of notifications within an LGA. The outbreak was deemed to have started with the first season recognized as extreme; the duration was defined by the number of consecutive seasons classified as extreme. Chi-square analysis was used to test the relationship between the season the outbreak started and the length of the outbreak.

Spatial autocorrelation of disease notification. ArcView Geographic Information System (GIS), version 3.2a (Environmental Systems Research Institute, Inc., Redlands, CA) was used to visualize the data and also to perform spatial data analysis. The spatial autocorrelation between LGAs for the long-term IR was assessed for each season using the Moran’s I statistic.22 This statistic is a weighted correlation coefficient used to detect departures from spatial randomness. Joint count statistics, as described by Lee and Wong,23 were used to assess the spatial autocorrelation between LGAs classified as extreme in each season and year, and also between LGAs having outbreaks in each season and year. For each method, randomized sampling was assumed, and joint count statistics were only performed if there were more than five LGAs classified as extreme in the specified season and year.

Grouping LGAs with similar temporal patterns of outbreaks. The LGAs having similar temporal patterns of RR virus disease outbreaks in the summer and autumn were grouped using hierarchical cluster analysis (SPSS version 9.0.1; SPSS, Inc., Chicago, IL). This analysis was conducted using complete linkage and binary squared euclidean distances, and allowed the frequency/regularity of RR virus disease outbreaks to be broadly classified across the state.

RESULTS

A total of 27,561 notifications of RR virus disease were received by Queensland Health from January 1991 to May 2001. Of these, 25,135 (91.2%) occurred within the study period (June 1991 to May 2001) and had complete patient data regarding place of residence, age, and sex. These 25,135 notifications formed the data for this study.

The crude IR for Queensland was 76.6 per 100,000 person-years (py), with the IRs for individual LGAs varying between 0 and 1,271 per 100,000 py. One LGA had no reported cases of RR virus disease in the 10 years analyzed. Disease notifications peaked in the 30–59-year-old age group (127.7 per 100,000 py), with comparable levels of 44.5 and 46.4 per 100,000 py for the 0–29- and ≥ 60-year-old age groups, respectively. There was a slight sex bias with a notification rate of 83.7 per 100,000 py for females compared with 70.1 per 100,000 py for males.

There was high seasonal variability, with 85% of notifications occurring in the summer and autumn; notifications were more than twice as likely in the autumn than in the summer (odds ratio = 2.12; 95% confidence interval = 2.06–2.18). More than 85% of the LGAs typically have some RR virus disease notifications during the year, although 50–74% experience little or no transmission during the winter and spring. The seasonal crude IRs were 25.7, 20.3, 85.5, and 176.8 per 100,000 py for winter, spring, summer, and autumn, respectively.

Along with strong seasonality, there were large variations in the number of notifications between years (Figure 1). In general, the number of notifications in winter and spring re-

![Figure 1](image-url)
mained fairly constant, whereas the number of cases in summer and autumn showed large fluctuations.

Applying the specified criteria to determine the long-term IR for an LGA resulted in the crude IR being used for between 104 and 121 LGAs, depending on the season. The long-term IRs for the remaining LGAs were determined using the median IR, except for one LGA in summer. In this instance, the median was zero so the trimmed mean was used. In all instances when the median was used, the resulting IR was less than the corresponding crude IR, and one or two years with exceptionally high numbers of notifications could be identified. These exceptional years had IRs at least twice that of the year with the next highest IR.

The long-term IRs showed significant positive spatial autocorrelation for all seasons with Moran’s I statistics ranging from a minimum of 0.2513 in winter to a maximum of 0.4414 in autumn (expected Moran’s I = −0.0081, variance = 0.0026, P < 0.01). This indicates clustering of LGAs with similar long-term IRs (Figure 2). Although there was no obvious pattern in IR across the entire state, it is evident that the southeast region of the state had comparably low incidence all year round (Figure 2).

Five thousand extreme indicators were calculated (125 LGAs × 4 seasons × 10 years). Approximately 50% of these indicators occurred in LGAs where the expected number of notifications was less than one. This value was highest (71%) in winter and lowest (21%) in autumn. Only 6.3% of the indicators were classified as extreme (Table 1). The number of LGAs classified as having extreme numbers of RR virus disease notifications closely following the variations in the total number of notifications within the state.

Using the extreme indicators, we identified 235 separate RR virus disease outbreaks in 90 LGAs. No abnormally high numbers of notifications were detected in the remaining 35 LGAs. Eighty-eight percent of outbreaks started in the summer (46%) or autumn (42%). More than 95% of the outbreaks lasted less than six months; 72% lasted one season, while 25% lasted two seasons. Chi-square analysis indicated that for outbreaks starting in winter, spring or summer, there was no significant association between season of outbreak and length of outbreak (P > 0.1). However, similar analysis of outbreaks starting in autumn compared with the other seasons indicated that short outbreaks of one season were more prevalent in autumn compared with other seasons (92.9% versus 57.4%; P < 0.01).

Spatial analysis of the extreme indicators for LGAs in each season of each year showed a positive autocorrelation between LGAs classified as extreme for all of the seasons and years analyzed. Of the 17 season/year combinations tested (the remainder of the combinations had less than 5 LGAs classified as extreme), 13 showed significant positive autocorrelation with z-values ranging from −2.35 to −8.16 (P < 0.01; Table 1). There was no correction made for the effect of multiple tests. The four combinations classified as not significant had z-values in the range of −1.16 to −1.62 (P > 0.1). Similar results were obtained when the spatial autocorrelation of outbreaks was assessed.

Assessment of the temporal pattern of RR disease outbreaks in summer and autumn indicated that the LGAs could be grouped into five clusters (Figures 3 and 4). Each of these clusters had different patterns and frequencies of outbreaks over the 10 years of data (Figure 3). The first cluster consisted of eight LGAs and was characterized by frequent outbreaks every 2–3 years. The six LGAs forming cluster 2 were also prone to regular outbreaks similar to cluster 1, but differed from the LGAs in cluster 1 by a delay from summer to autumn in the 1998-1999 outbreak (Figure 3). The five LGAs in cluster 3 were the only LGAs within the state to experience an outbreak in the summer of 1999-2000, with 4 (80%) of the 5 LGAs also experiencing an outbreak in the autumn of 1996–1997. Cluster 4, consisting of 35 LGAs, was characterized by only one outbreak in the summer of 1995–1996. The majority of LGAs in the fifth and largest cluster (71 LGAs) had no outbreaks over the study period.

**DISCUSSION**

Queensland has the highest reported number of RR virus disease cases in Australia, prompting several previous studies into disease incidence in selected cities and towns. However, this is the first study to examine patterns of RR virus disease notifications across the entire state, and the first to define a statistically sound method of separating unusual RR virus disease activity from endemic transmission. We have calculated age/sex standardized IRs for the 125 LGAs within Queensland and shown that LGAs with similar disease notification rates tend to be grouped together. We were also able to group the LGAs into five clusters based on the temporal pattern of their RR virus disease outbreaks.

To determine if an outbreak of RR virus disease has occurred, it is first necessary to know the IR associated with endemic transmission. Many studies have used the mean IR or mean number of cases as an indicator of usual transmission. However, since the number of cases during an epidemic can be much higher than during normal endemic transmission, inclusion of data from epidemic years overestimates the usual incidence of a disease. Additionally, mosquito-borne diseases that experience seasonality can show a skewed distribution in time causing a violation in the assumptions of data normality. For this reason we considered the median IR a better measure of the usual incidence compared with the mean. However, many LGAs had small populations at risk that influenced the potential number of notifications. It was determined that any LGA expecting less than 0.5 notifications in a season would be likely to have a median of zero notifications for that season over the 10 years included in the study. In these instances, the mean IR provided a better indicator of notification rates compared with the median. Thus, the determination of a long-term measure of the LGA-specific IR was, by necessity, dependent on the population size of the LGA.

The spatial analysis of the standardized long-term seasonal IRs indicated that LGAs with similar IRs tended to be clumped together. However, caution needs to be used when comparing the IRs between LGAs since it has long been recognized that rates from areas with small populations are more unstable than those with larger populations, and that large differences in population sizes can effect the sensitivity of the Moran’s I statistic. We have tried to overcome this problem by developing an average IR from 10 years of data. However the results presented should still be viewed as a broad guide of disease incidence across the state.

In winter, the tropical LGAs on the east coast tended to
have higher IRs than either inland LGAs or LGAs in the southeast corner of the state. The pattern of LGAs in the southeast corner having lower IRs than most of the other LGAs was maintained in all seasons, as previously noted for a selection of regions. The reasons for this difference require further investigation, but may be partly due to the highly developed and broadscale saltmarsh mosquito control programs in place. The pattern of tropical coastal LGAs
having higher IRs than inland regions was not evident in summer and autumn. During these two seasons, many inland LGAs had IRs that were higher than their coastal equivalents. This contradicts a previous study\textsuperscript{13} that concluded that monthly mean incidence between 1985 and 1996 in coastal districts was higher than inland regions. Importantly, one of the two inland regions used by Tong and Hu\textsuperscript{13} was within the southeast corner of the state, a region with consistently low IRs. This is likely to have had a major impact on the prediction of the inland region incidence since only two inland towns were included.

Many investigators have referred to RR virus disease outbreaks or epidemics throughout Australia.\textsuperscript{1,16,18,19} However, a definitive definition of an outbreak is rarely provided. We used the Poisson distribution to identify periods with high numbers of notifications. This technique provided a statistically robust and consistent method of detecting outbreaks.

### TABLE 1

<table>
<thead>
<tr>
<th></th>
<th>Winter</th>
<th>Spring</th>
<th>Summer</th>
<th>Autumn</th>
</tr>
</thead>
<tbody>
<tr>
<td>1991–1992</td>
<td>0</td>
<td>0</td>
<td>11\textsuperscript{†} (11)</td>
<td>28\textsuperscript{†} (18)</td>
</tr>
<tr>
<td>1992–1993</td>
<td>6\textsuperscript{†} (4)</td>
<td>9\textsuperscript{†} (5)</td>
<td>2 (0)</td>
<td>7\textsuperscript{†} (5)</td>
</tr>
<tr>
<td>1993–1994</td>
<td>8 (5)</td>
<td>11 (6)</td>
<td>28\textsuperscript{†} (22)</td>
<td>15\textsuperscript{†} (8)</td>
</tr>
<tr>
<td>1994–1995</td>
<td>0</td>
<td>3 (3)</td>
<td>0</td>
<td>17\textsuperscript{†} (17)</td>
</tr>
<tr>
<td>1995–1996</td>
<td>2 (0)</td>
<td>1 (1)</td>
<td>45\textsuperscript{†} (44)</td>
<td>42\textsuperscript{†} (17)</td>
</tr>
<tr>
<td>1996–1997</td>
<td>2 (0)</td>
<td>0</td>
<td>11\textsuperscript{†} (11)</td>
<td>13 (12)</td>
</tr>
<tr>
<td>1997–1998</td>
<td>2 (2)</td>
<td>0</td>
<td>2 (2)</td>
<td>15 (14)</td>
</tr>
<tr>
<td>1998–1999</td>
<td>2 (1)</td>
<td>1 (1)</td>
<td>9\textsuperscript{†} (9)</td>
<td>12\textsuperscript{†} (8)</td>
</tr>
<tr>
<td>1999–2000</td>
<td>0</td>
<td>0</td>
<td>5 (5)</td>
<td>0</td>
</tr>
<tr>
<td>2000–2001</td>
<td>0</td>
<td>0</td>
<td>4 (4)</td>
<td>1 (1)</td>
</tr>
</tbody>
</table>

* Season/year combinations that were tested for statistically significant spatial autocorrelation between local government areas (LGAs) classified as extreme are in bold with \textsuperscript{†} indicating the presence of a significant positive autocorrelation (\(P < 0.01\)). Values in parentheses indicate the number of LGAs that experienced the start of an outbreak in the specified season and year.

---

**FIGURE 3.** Proportion of local government areas (LGAs) within each cluster in Queensland, Australia that experience the start of an outbreak of Ross River virus disease in the summer and autumn of the years 1991–1992 to 2000–2001. Cluster 1 consists of 8 LGAs, while clusters 2–5 contain 6, 5, 35, and 71 LGAs, respectively.
within LGAs. Use of aggregated data and a fixed number of notifications as the criterion for an outbreak readily identifies outbreaks in major population centers, but can overlook outbreaks in smaller populations. Conversely, when outbreaks are detected within aggregated datasets, it is assumed that all areas within the region experienced an outbreak, possibly resulting in an “ecologic fallacy.”

Assessing the number of notifications at the LGA level provides the opportunity to identify outbreaks that would otherwise go unnoticed if only state wide notifications were used. It also provides a specific database for health personnel to identify and respond to disease outbreaks.

The use of notification data as an indicator of disease incidence or transmission has inherent problems, such as reporting bias. Across Queensland, it is probable that different regions have different reporting rates and biases based on education, proximity to diagnostic facilities, and the perceived benefits of obtaining a diagnosis given that no specific treatment for the virus disease exists. Several of these factors have been previously identified, and may partly explain the observed differences in the notification rates across the state. The advantage of our proposed method is that disease outbreaks are identified relative to the norm of the local region. Thus, reporting bias is not likely to influence the identification of extreme events, provided that reporting trends are consistent over time within the region.

The grouping of LGAs into clusters based on the temporal patterns of outbreaks indicated that the LGAs within Queensland appear to have varying susceptibilities to RR virus disease outbreaks. An interesting, but not surprising, result was the tendency of LGAs within the same cluster to be located in close proximity. This suggests that specific regional factors may be important in stimulating RR virus disease outbreaks because the five clusters did not exhibit outbreaks at the same time.

The 14 LGAs grouped into clusters 1 and 2 represent 61% of the population of Queensland. The LGAs within these clusters were located solely in the southeast corner of the state, and were prone to regular outbreaks at ~2-year intervals from 1991 to 1999. Since the majority of the population resides in these LGAs, this population center drives the fluctuations in the total state notifications. The regular patterns in state notifications have been reported to coincide with changes in the phase of the Quasi-Biennial Oscillation (QBO), thus prompting the hypothesis that RR virus transmission and the QBO may be linked. However, as demonstrated here, this cyclic pattern is restricted to a small region in the southeast corner of the state. Given that the QBO is a large-scale pattern of climate fluctuation, it is not immediately obvious why only a small part of the state exhibits a cyclic pattern in RR virus disease notifications. Other studies relating large-scale climatic phenomenon to RR virus disease outbreaks have also been based on aggregated data for Queensland. However, our results highlight the importance of the information obtained from a detailed geographic analysis of disease notifications, particularly in areas that consist of diverse environments, or highly clustered populations.

Several investigators have tried to relate RR virus disease incidence to environmental factors with varying success. However, climatic factors only form one part of the complex virus-mosquito-human interaction. Modeling the incidence of RR virus disease also requires recognition of other factors that can potentially affect transmission such as human behavior, the mosquito species present, control measures, the availability of mosquito breeding sites, and abundance of non-immune vertebrate hosts. Assuming that all, or some, of these factors contribute in some way to the number of notifications, it would be difficult to accurately predict the incidence of RR virus disease using environmental or landscape factors alone. An alternative to modeling RR virus disease incidence is to focus on deviations from the usual level of notifications for a region, as we have done here. By considering deviations from normal, many of the underlying factors contributing to RR virus transmission need not be considered.

The techniques adopted in this study are equally relevant to other diseases and/or locations provided good quality notification data exist. The examination of such data at a local level can identify areas where diseases are endemic, or have consistently high infection rates relative to surrounding areas. The identification of disease outbreaks provides the opportunity to explore conditions or factors that may be responsible for stimulating outbreaks, and allows for the development of early warning systems. Early outbreak detection provides the opportunity to implement control measures and/or education campaigns targeting the appropriate audience when it is most required. Given the recent advances in GIS and spatial analysis and their increasing popularity, our study has highlighted the importance of using basic notification data to provide an adequate depiction of disease incidence within a community.

Received December 15, 2003. Accepted for publication January 28, 2004.

Financial support: This work was supported by a Queensland Health Arbovirus Prevention Grant.

Authors’ addresses: Michelle L. Gatton, Australian Centre for International and Tropical Health and Nutrition, Queensland Institute of
Medical Research, PO Royal Brisbane Hospital, Herston, Queensland 4029, Australia; Telephone: 61-7-3362-0416, Fax: 61-7-3362-0104, E-mail: michellG@qimr.edu.au. Louise A. Kelly-Hope, Vector Research Division, Liverpool School of Tropical Medicine, Pembroke Place, Liverpool, L3 5QA, United Kingdom; Telephone: 44-151-705-3107, Fax: 44-151-705-3369, E-mail: lakhope@liverpool.ac.uk. Brian H. Kay, Australian Centre for International and Tropical Health and Nutrition, Queensland Institute of Medical Research, PO Royal Brisbane Hospital, Herston, Queensland 4029, Australia; Telephone: 61-7-3362-0350, Fax: 61-7-3362-0106, E-mail: brianK@qimr.edu.au. Peter R. Ryan, Australian Centre for International and Tropical Health and Nutrition, Queensland Institute of Medical Research, PO Royal Brisbane Hospital, Herston, Queensland 4029, Australia; Telephone: 61-7-3362-0351, Fax: 61-7-3362-0106, E-mail: peterR@qimr.edu.au.

REFERENCES


