

## Clustering of equine grass sickness cases in the United Kingdom: a study considering the effect of position-dependent reporting on the space–time $K$ -function

N. P. FRENCH<sup>1</sup>\*, H. E. MCCARTHY<sup>2</sup>, P. J. DIGGLE<sup>3</sup> AND C. J. PROUDMAN<sup>2</sup>

<sup>1</sup> Institute of Veterinary, Animal and Biomedical Sciences, Massey University, Palmerston North, New Zealand

<sup>2</sup> Department of Veterinary Clinical Science and Animal Husbandry, University of Liverpool Faculty of Veterinary Science, Leahurst, Neston, Cheshire, UK

<sup>3</sup> Department of Mathematics and Statistics, Lancaster University, Lancaster, UK

(Accepted 4 October 2004)

### SUMMARY

Equine grass sickness (EGS) is a largely fatal, pasture-associated dysautonomia. Although the aetiology of this disease is unknown, there is increasing evidence that *Clostridium botulinum* type C plays an important role in this condition. The disease is widespread in the United Kingdom, with the highest incidence believed to occur in Scotland. EGS also shows strong seasonal variation (most cases are reported between April and July). Data from histologically confirmed cases of EGS from England and Wales in 1999 and 2000 were collected from UK veterinary diagnostic centres. The data did not represent a complete census of cases, and the proportion of all cases reported to the centres would have varied in space and, independently, in time. We consider the variable reporting of this condition and the appropriateness of the space–time  $K$ -function when exploring the spatial-temporal properties of a ‘thinned’ point process. We conclude that such position-dependent under-reporting of EGS does not invalidate the Monte Carlo test for space–time interaction, and find strong evidence for space–time clustering of EGS cases ( $P < 0.001$ ). This may be attributed to contagious or other spatially and temporally localized processes such as local climate and/or pasture management practices.

### INTRODUCTION

Equine grass sickness (EGS) is a largely fatal, pasture-associated dysautonomia of unknown aetiology. The acute disease is characterized primarily by failure of normal alimentary function. The highest incidence of EGS is believed to occur in Scotland where the disease was first identified in 1907, although the disease also occurs throughout England and Wales (for a review see [1]). The temporal pattern of EGS has also been well described, with most cases occurring between April and July [2–4]. However, to date, little has been

done to explore the spatial patterns of EGS in England and Wales. In addition, no studies have considered the joint distribution of EGS cases in space and time.

The well-described spatial distribution might arise as a result of regional variation in the population at risk, reporting bias and other static variables associated with the onset of disease (e.g. soil type, habitat and altitude). Likewise, the temporal pattern might be attributed to seasonal variation in risk factors such as temperature and pasture growth. However, we were interested in exploring the possibility that these spatial and temporal processes interact to provide space–time clustering. Such a situation can arise if there is a contagious element or spatially and temporally localized variation in other ‘driving’ variables.

\* Author for correspondence: Professor N. P. French, Institute of Veterinary, Animal and Biomedical Sciences, Massey University, Palmerston North, New Zealand.  
(Email: n.p.french@massey.ac.nz)

Estimating the extent of clustering might assist in the formulation of disease-prevention strategies in the face of a confirmed case.

There are many approaches to the exploration of space–time clustering in human and animal diseases. The Mantel test, Barton’s method, nearest-neighbour test and Knox’s test are techniques useful for investigating space–time interaction. These and other methods are reviewed in refs [5–7]. Most of these tests concentrate solely on testing the null hypothesis of no space–time interaction and do not consider the scale or nature of space–time interaction. Diggle et al. [8] suggested an alternative approach, which provided a formal, edge-corrected test for space–time interaction and informal graphical methods for describing the nature and scale of space–time clustering. This method extends the use of  $K$ -functions [9] (previously used to describe purely spatial point patterns) to spatial–temporal processes. The technique has been used to describe sporadic cases of human disease but has rarely been used in veterinary epidemiology [10, 11]. We use  $K$ -function analysis to examine space–time clustering of EGS in England and Wales and consider the effects of position-dependent under-reporting on the validity of the space–time  $K$ -function.

## METHODS

### Data on EGS

Data from histologically confirmed cases of EGS in England and Wales were collected from UK veterinary diagnostic centres. Cases that were diagnosed from clinical signs only were excluded from the study due to the possibility of misdiagnosis. The date used was the date the horse first showed clinical signs or the date the horse was admitted to a veterinary hospital (if applicable). The former was used in preference. Because this is an acute-onset disease the difference between the two was likely to have been less than 2 days. The geographical location where the case occurred was identified as the nearest village or town to the premises where the horse originated. This would have been within 1–2 km of the premises (client confidentiality restricted the authors access to exact premises locations). Each location was converted into a four-figure Ordnance Survey grid reference using the *Ordnance Survey Gazetteer of Great Britain* (third edition). Information was collected on 133 histologically confirmed cases of grass sickness in the years 1999 and 2000. Two datasets were used in the analysis. The

first comprised all reported cases including multiple cases from the same premises. The second was a slightly smaller (119 cases) database comprising only the first case to be reported on each of the premises during the study period. This smaller dataset was analysed to examine space–time clustering in the absence of very local (i.e. same premises) effects.

### Space–time $K$ -function analysis

The technique is fully described by Diggle et al. [8]. Briefly, the observed spatial–temporal point pattern is compared with a pattern that has the same temporal and the same spatial properties as the original data, but no space–time interaction. Like the Knox test this method does not require knowledge of the underlying population at risk. This involves the estimation of three component processes, i.e.  $K$ -functions:

$K(s, t) = \lambda^{-1} E$  [number of further events occurring within distance  $s$  and time  $t$  of an arbitrary event of the process],

$K_1(s) = \lambda_1^{-1} E$  [number of further events occurring within distance  $s$  of an arbitrary event of the process],

$K_2(t) = \lambda_2^{-1} E$  [number of further events occurring within time  $t$  of an arbitrary event of the process],

where  $E$  refers to the edge-corrected expected value [e.g. for  $K_1(s)$  it is the mean number of events within distance  $s$  of an arbitrary event, adjusted for edge-effects],  $\lambda$  is the expected number of events per unit space per unit time (intensity of the disease process), and  $\lambda_1$  and  $\lambda_2$  are the spatial and temporal intensities respectively.

The product of the spatial and temporal  $K$ -functions  $K_1(s)K_2(t)$  is the expected  $K$ -function under the hypothesis of no space–time interaction and is used as a benchmark for comparison with the observed space–time  $K$ -function:  $K(s, t)$ .

In addition to providing an edge-corrected Monte Carlo test for space–time interaction (see [8] for full details), the method can be used to estimate and describe proportional and absolute increases in disease risk, attributable to space–time interaction, for a range of spatial and temporal separations. This is achieved by graphing three diagnostic functions. The edge-correction and the additional interpretation of the nature of space–time interaction are advantages of this method over the Knox test.

$$\hat{D}(s, t) = \hat{K}(s, t) - \hat{K}_1(s)\hat{K}_2(t). \quad (1)$$

This function is proportional to the increased numbers of cases within distance  $s$  and time  $t$  by comparison with a process with the same spatial and temporal structures but no space–time interaction. It is analogous to the risk difference in epidemiology and is plotted against  $s$  and  $t$ .

$$\hat{D}_0(s, t) = \hat{D}(s, t) / \{\hat{K}_1(s)\hat{K}_2(t)\}. \quad (2)$$

This is proportional to the relative increase in cases within distance  $s$  and time  $t$  compared with a process with the same spatial and temporal structures but no space–time interaction. It is analogous to the relative risk and is plotted against  $s$  and  $t$ .

$$R(s, t) = \hat{D}(s, t) / \sqrt{\{V(s, t)\}}, \quad (3)$$

where  $V(s, t)$  is the variance of  $\hat{D}(s, t)$  [see [8] for the calculation of  $V(s, t)$ ]. When plotted against  $K_1(s)K_2(t)$  this is analogous to a plot of the standardized residuals against fitted values. If the spatial and temporal processes were independent (i.e. there was no space–time interaction) then we would expect approximately 95% of the values to lie between the values  $\pm 2$ . A disadvantage of this plot is that the spatial and temporal scales are no longer explicit and the ‘residuals’ are highly interdependent. Rather than a ‘cloud’ of points, this lack of independence may produce obvious patterns (e.g. lines of points) within the plot. We suggest that for most applications, the function  $\hat{D}_0(s, t)$  will be of most interest.

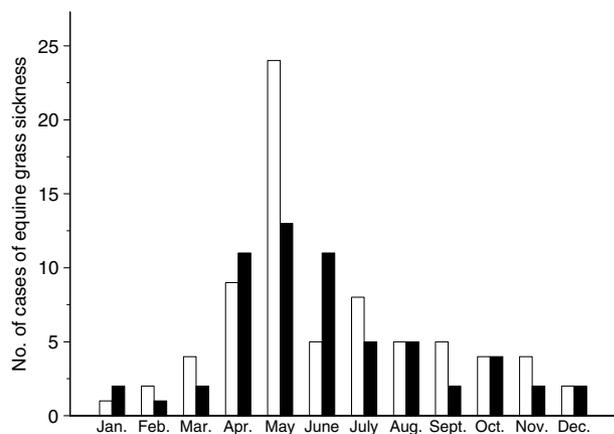
### Considering the effects of position-dependent thinning (systematic under-reporting)

The data used in this analysis were not a complete census of all cases in England and Wales. Furthermore, the proportion of all cases of EGS included in the analysis varied according to geographical location and may also have varied over time. This is termed ‘position-dependent thinning’ and arises from the variation in case reporting throughout the region studied (for example it is likely that proportionally more cases were ascertained in areas closer to the contributing diagnostic centres). This means that the spatial distribution of the cases used in this analysis does not represent the true underlying spatial distribution of EGS cases in England and Wales. Likewise there may be times of the year when proportionally fewer cases are reported. Although tests for purely spatial or purely temporal clustering would be affected by position-dependent thinning or

temporal-dependent thinning respectively, we show that the Monte Carlo test for space–time interaction used in this analysis is not invalidated by such thinning, provided the spatial and temporal thinning operate independently.

For a more formal consideration of the effect of position-dependent thinning on the test of no space–time interaction we consider a spatio-temporal dataset  $(x_i, t_i): i = 1, \dots, n$  generated by a process with no space–time interaction. Hence the complete set of spatial locations,  $X = (x_1, \dots, x_n)$  is independent of the complete set of temporal locations,  $T = (t_1, \dots, t_n)$ . A spatio-temporal thinning is defined by a function  $p(x, t)$  denoting the probability that a point survives in the thinned process, given that it is at space–time location  $(x, t)$ . If spatial and temporal thinnings operate independently, then  $p(x, t) = p_1(x)p_2(t)$ . If we assume that spatial and temporal thinnings are independent we can imagine the thinning to take place in two stages: the first according to the spatial thinning function  $p_1(x)$ ; the second thinned from the survivors of the first stage according to the temporal thinning function  $p_2(t)$ . If we call the temporal locations which survive the first stage of thinning  $T^*$ , and the surviving spatial locations  $X^*$ , then under the stated assumptions,  $T^*$  consists of a random sample from the original set  $T$ . Since  $T$  is independent of  $X$ , it follows that  $T^*$  is independent of any subset of  $X$  chosen without reference to  $T$ , and in particular is independent of  $X^*$ . Hence  $X^*$  and  $T^*$  are independent and a test of no space–time interaction in the data  $X^*$ ,  $T^*$  is a valid test of no space–time in the original data  $X, T$ . If we repeat the argument, applying a second thinning to  $X^*$ ,  $T^*$  according to the temporal thinning function  $p_2(t)$ , and reverse the roles of the spatial and temporal locations, we produce thinned data  $X^{**}$ ,  $T^{**}$ . A test of no space–time interaction in  $X^{**}$ ,  $T^{**}$  is also a valid test of no space–time interaction in  $X^*$ ,  $T^*$  and hence in the original data  $X, T$ .

However, if we reject the null hypothesis of no space–time interaction, we cannot always interpret the estimated space–time interaction in the thinned process as an estimate of the space–time interaction in the original process. To do this we need to make the stronger assumption that  $p(x, t) = p$ , as  $K(s, t)$  is invariant under this form of ‘homogeneous’ thinning. However, preliminary simulation studies have shown that for the mechanism by which position-dependent thinning of EGS takes place, both the absolute and relative differences between the estimated and theoretical  $K$ -functions were small [12].



**Fig. 1.** Seasonal distribution of 133 histologically confirmed cases of equine grass sickness during 1999 (□) and 2000 (■) in England and Wales.

All analyses were performed using the SPLANCS library [13] in S-PLUS (Insightful Corporation, Seattle, WA, USA).

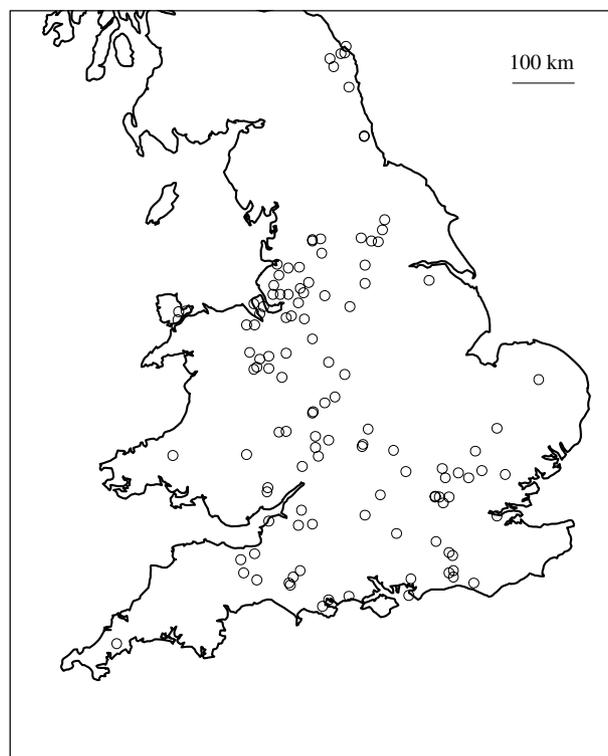
## RESULTS

### Temporal and spatial patterns of EGS

Figure 1 shows a peak in May in the temporal pattern of the 133 EGS cases used in this study. Figure 2 shows the spatial distribution of the 133 histologically confirmed cases of EGS used in this study. The apparent spatial clustering of EGS cases in several parts of England and Wales is potentially misleading. This is most likely due to uneven, clustered reporting of cases, particularly around the major contributing diagnostic centres. Hence purely spatial thinning  $p(x)$  is a plausible feature of the problem.

### Space-time K-function analysis

When all cases were considered together there was strong evidence of space-time clustering of EGS cases. The Monte Carlo test for space-time interaction computed a  $P$  value of  $<0.001$  providing formal evidence for space-time interaction. The residual plot (Fig. 3); referring to spatial separations up to 30 km (in 5-km increments) and temporal separations up to 90 days (in 4-day increments), revealed that almost all of the residuals were  $>2$ . This indicated space-time interaction over wide spatial and temporal separations, but particularly for lower values of  $K_1(s)K_2(t)$  corresponding to shorter spatial and temporal distances. This is confirmed by the  $\hat{D}(s, t)$  plot which showed a sharp rise over the first 10 km and 20 days



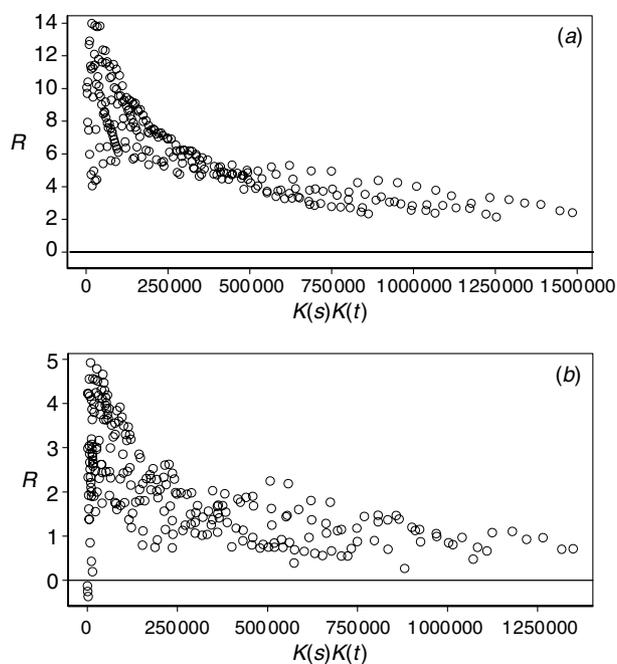
**Fig. 2.** The spatial distribution of 133 histologically confirmed cases of equine grass sickness during 1999 and 2000 in England and Wales.

(not shown) and demonstrated further in the  $\hat{D}_0(s, t)$  plot (Fig. 4) which showed the greatest proportional increase within these distances. High values on the  $z$  axis of the  $\hat{D}_0(s, t)$  plot indicates that there are many more outbreaks within the given spatial and temporal separation than would be expected if there were no space-time clustering.

When multiple cases were excluded from the analysis this left 119 EGS outbreaks (incident cases). The  $R(s, t)$  and  $\hat{D}_0(s, t)$  plots of the 119 EGS outbreaks (Figs 3 and 4) also showed that a large proportion of incident cases within 10 km and 20 days of each other could be attributed to space-time clustering. Thus the exclusion of multiple cases still revealed significant evidence of clustering. The Monte Carlo test for space-time interaction computed a  $P$  value of 0.016.

## DISCUSSION

By providing both an edge-corrected test for space-time interaction and an estimate of the nature and extent of the process, the  $K$ -function has advantages over the Knox test. Furthermore, this method complements the many other approaches to the analysis

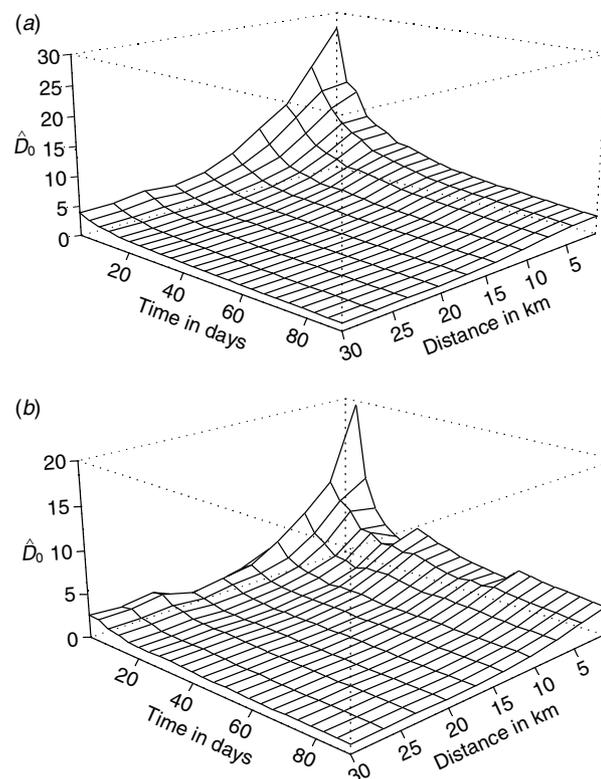


**Fig. 3.** Residual plots of  $R(s, t)$  against  $K_1(s)K_2(t)$  for (a) all 133 cases and (b) 119 incident cases of equine grass sickness reported during 1999 and 2000 in England and Wales.

of data with both a spatial and temporal dimension. For example, the space–time scan statistic [14] is useful for identifying the location of individual clusters or ‘galaxies’ of disease in space–time cylinders. In contrast,  $K$ -function analysis helps to describe the more general properties of a space–time process.

In this study it was highly likely that the proportion of cases reported varied across and within regions due to variable degrees of under-reporting by veterinary surgeons and pathologists. Temporal variation in the degree of under-reporting could also occur if, for example, a diagnosis was less likely to be made outside the high-risk months of April, May and June.  $K$ -function analysis is not invalidated by random under-reporting or ‘thinning’. Furthermore, the test for space–time interaction is not invalidated by position-dependent under-reporting, provided the spatial and temporal thinning were independent. Therefore, unless there was an interaction between spatial and temporal variation in reporting (e.g. a veterinary surgeon reported all cases within a small area over a short time period and only a fraction of other cases encountered) under-reporting would not have invalidated the test for space–time clustering.

This study provided strong evidence for space–time clustering of grass sickness cases in the United Kingdom. Space–time clustering was evident both



**Fig. 4.** A 3-dimensional plot of the  $\hat{D}_0(s, t)$  function for (a) all 133 cases and (b) 119 incident cases of equine grass sickness reported during 1999 and 2000 in England and Wales.

when individual EGS cases and individual outbreaks of EGS were investigated. The presence of a space–time interaction on such a local scale is consistent with either an infectious aetiology, and/or risk factors that localize both spatially *and* temporally.

The causal agent of EGS has yet to be determined but there is increasing evidence for the role of *Clostridium botulinum* type C [15, 16]. It is possible, therefore, that the observed space–time clustering is due to contagious spread of this organism. However, it has been reported that only one horse in a field of many develops the disease and the occurrence of disease does not follow a typical spread of an infectious disease. It seems more likely that the onset of EGS is attributable to risk factors that are localized in both space and time. Tocher et al. [17] mapped the spread of the disease in Scotland and reported that no theory could explain the spread of disease and was undecided whether the distribution of cases resembled that of a contagious epidemic disease or a disease arising from the spread of a sporing organism. We hypothesize that environmental factors, possibly meteorological, in particular geographical areas, trigger the development

of disease in susceptible individuals. It is possible that *Cl. botulinum* type C is more prevalent in certain areas of the United Kingdom and that disease is triggered in susceptible animals under conditions that stimulate toxin production *in vivo*. This might explain the space-time clustering of EGS cases found in this study.

#### ACKNOWLEDGEMENTS

This study was funded by The Home of Rest for Horses and forms part of a larger study of EGS. The authors thank Dr Paulo Ribeiro, Dr Helen Clough and Sarah Fenton for their help with the study, and the many veterinary surgeons and pathologists who made data available for the study.

#### REFERENCES

1. McCarthy HE, Proudman CJ, French NP. Epidemiology of equine grass sickness: a literature review (1909–1999). *Vet Rec* 2001; **149**: 293–300.
2. Guthrie WJ. Grass sickness in horses. *J Roy Agric Soc* 1940; **100**: 50–59.
3. Doxey DL, Gilmour JS, Milne E. The relationship between meteorological factors and equine grass sickness (dysautonomia). *Equine Vet J* 1991; **23**: 370–373.
4. Wood JL, Doxey DL, Milne EM. A case-control study of grass sickness (equine dysautonomia) in the United Kingdom. *Vet J* 1998; **156**: 7–14.
5. Bailey TC, Gatrell AC. Interactive spatial data analysis. Longman, Harlow, 1995.
6. Ward MP, Carpenter TE. Analysis of time-space clustering in veterinary epidemiology. *Prev Vet Med* 2000; **43**: 225–237.
7. Carpenter TE. Methods to investigate spatial and temporal clustering in veterinary epidemiology. *Prev Vet Med* 2001; **48**: 303–320.
8. Diggle PJ, Chetwynd AG, Haggkvist R, Morris SE. Second order analysis of space-time clustering. *Stat Methods Med Res* 1995; **4**: 124–136.
9. Ripley BD. The second order analysis of stationary point processes. *J Appl Prob* 1976; **13**: 255–266.
10. French NP, Berriatua E, Wall R, Smith K, Morgan KL. Sheep scab outbreaks in Great Britain between 1973 and 1992: spatial and temporal patterns. *Vet Parasitol* 1999; **83**: 187–200.
11. French NP, Clough HE, Berriatua E, McCarthy HE, Proudman CJ, Hillyer MH. The use of *K*-function analysis to detect and describe space-time clustering of animal diseases. In: Proceedings of the Ninth International Society on Veterinary Epidemiology and Economics, Breckenridge, CO (Abstract ID 314) 2000.
12. Fenton SE, Clough HE, Diggle PJ, French NP. An Investigation of the effect of under-reporting of disease cases on the space-time *K*-function. International Conference on Spatial Point Process Modelling and its Applications (SPPA). Benicassim, Castellón (Spain), 4–8 April 2004.
13. Rowlingson BS, Diggle PJ. SPLANCS: spatial point pattern analysis code in S-plus. *Comput Geosci* 1993; **19**: 627–655.
14. Kulldorff M, Athas WF, Feurer EJ, Miller BA, Key CR. Evaluating cluster alarms: a space-time scan statistic and brain cancer in Los Alamos, New Mexico. *Am J Public Health* 1998; **88**: 1377–1380.
15. Hunter LC, Poxton IR. Systemic antibodies to *Clostridium botulinum* type C: do they protect horses from grass sickness (dysautonomia)? *Equine Vet J* 2001; **33**: 547–553.
16. McCarthy HE, French NP, Edwards GB, et al. Equine grass sickness is associated with low antibody levels to *Clostridium botulinum*. Findings from a matched case-control study. *Equine Vet J* 2004; **36**: 123–129.
17. Tocher JF, Tocher JW, Brown W, Buxton JB. Grass Sickness Investigation Report. *Vet Rec* 1923; **3**: 37–45 and 75–89.