Original Research Article

A geospatial analysis of pertussis and its risk factors in southern Ontario from 2005–2016

Taha Elghamudi[a], Olaf Berke[b]

[a] Michael G. DeGroote School of Medicine, McMaster University, Hamilton, Canada
[b] Department of Population Medicine, Ontario Veterinary College, Guelph, Canada

Abstract

Introduction: Pertussis, commonly known as whooping cough, is a bacterial respiratory tract infection caused by *Bordetella pertussis*. Pertussis affects more than 48 million people worldwide annually, most of whom are under the age of 5.

Hypothesis & Objectives: The hypothesis being investigated is that pertussis incidence, between 2005 and 2016, is not equally distributed across public health units in southern Ontario. We aim to identify disease cluster locations and associate geospatial fluctuations in incidence rates with putative risk factors.

Materials and Methods: Data was sourced from Public Health Ontario on pertussis incidence in southern Ontario for all ages, specifically for each public health unit's geographical area. A choropleth map was generated using data smoothed by empirical Bayesian estimation in a spatial analysis context. Following the creation of an incidence map for southern Ontario, the spatial scan test was applied to elucidate the existence of any disease clusters at a public health unit level. Moran's I was used to determine whether there was evidence of any spatial dependence in pertussis incidence. Finally, putative risk factors were assessed in Poisson regression models and spatial Poisson regression models as potential predictor variables.

Results and Discussion: The flexible spatial scan test identified three spatial clusters where incidence rates of pertussis were higher than expected. A spatial Poisson regression model was fit that included predictor variables of socioeconomic status and population density. For every 100 people/km2 increase in population density there was a significant 6% increase in pertussis incidence ($p=0.03$). Interestingly, vaccination rates were not found to be predictive of pertussis incidence nor did the variable improve the model. This epidemiological study identifies where pertussis incidence is clustered and what variables it is associated with, both of which are valuable for public health purposes and as a reference for future research into pertussis.

Keywords: pertussis, spatial cluster, mapping, Ontario, Canada, regression, vaccination rates, socioeconomic status, population density.
Corresponding author: taha.elghamudi@medportal.ca

Introduction

Background on pertussis

Pertussis, commonly known as whooping cough, is an infectious disease that affects more than 48 million people worldwide and accounts for approximately 300,000 deaths annually (1-4). In 2017, Ontario saw 584 cases (5) while in the US there are over 48,000 cases of pertussis annually, with a high rate of underreporting due to mild symptoms being common (6).

Pertussis is caused by the bacteria *Bordetella pertussis* which infects the respiratory tract and releases toxins, particularly the pertussis exotoxin, that irritate the airways and cause paroxysmal coughing spasms (7,8). Persistent coughing contributes to the highly contagious nature of pertussis. Pertussis is mainly a childhood disease, with 38% of reported cases being infants younger than 6 months, and close to three quarters of cases being children younger than 5 years old (9).

Vaccines have been crucial in the decrease of pertussis incidence in the 20th century. In the United States, the incidence of pertussis decreased from 250,000 cases per year before vaccine introduction in the mid-20th century to as low as 1,010 reported cases in 1976. However, pertussis has slowly re-emerged, with pertussis outbreaks now being seen even in vaccinated populations. Reasons for this re-emergence may include a possible decreased effectiveness of the pertussis vaccine (4,10,11), or parents delaying or forgoing vaccination (12–14).

Moreover, the lack of robust reporting in the past contributes to the more recent rise in reported cases. Studies on cases of prolonged coughing suggest that despite the increase in reporting, pertussis is still more prevalent than was previously thought (4). It is likely that up to a million cases of respiratory illness due to pertussis infection exist in the US per year (4).

Spatial Epidemiology

The spatial pattern of a disease's incidence is related to the agent, host, and environment, otherwise known as the "epidemiological triangle." These factors are essential to learn about a disease and how to control it. Spatial epidemiology can provide clues about these factors and their interrelations. For example, testing for the level of clustering helps to determine the nature of the disease agent, finding clusters can identify populations at risk, and disease mapping helps in visualizing the environment (15).

Clustering refers to the spatial dependence in the data. A disease can have a high degree of clustering where cases are found in close spatial proximity, a high degree of regularity where the disease is equally spaced out, or it can have neither of these patterns and be spatially independent. Disease clusters are regions where the number of cases is too high to occur randomly and therefore those regions can be identified as high-risk. Thus, clustering and disease clusters refer to different spatial patterns and can be independent of one another.

General linear models are a useful approach for describing patterns and relationships in spatial data. One of the assumptions of general linear models is that observations are independent of one another. However, when analyzing spatial data of a contagious disease this assumption may not hold true. If a disease is rare then it is often considered to follow a Poisson distribution. Thus, the analysis of risk factors for rare diseases often employs a spatial Poisson regression model (16). This type of model is part of a larger group called generalized linear mixed models (16). These models are random effect models, where the random effects are spatially correlated (16).

In Ontario, pertussis is a "reportable disease," meaning physicians and other health-care practitioners are required to report cases of pertussis to their Medical Officer of Health. The rates of pertussis incidence are then collected by each Public Health Unit (PHU) and reported on an annual basis. It is expected in this analysis that pertussis incidence and the identified risk factors vary in space, thus prompting the use of spatial epidemiology to relate them. To the best of the authors' knowledge no research has been conducted on the geospatial fluctuations in pertussis incidence in southern Ontario.

Hypothesis

This study seeks to investigate whether pertussis incidence is equally distributed across PHUs in Ontario between 2005 and 2016. We aim to identify possible disease clusters and associate geospatial fluctuations in incidence rates with putative risk factors.

Materials and Methods

Materials

Incidence rates of pertussis and population size between 2005 and 2016 in each of the PHUs in southern Ontario were retrieved from Public Health Ontario (14). These annual data were aggregated so that each PHU had a cumulative value for both population and pertussis incidence. The base map used to visually represent the data was a boundary file of the 29 PHUs in southern Ontario retrieved from Statistics Canada (17). PHU data on vaccination rates were taken from the *Technical Report of Immunization coverage for school pupils in Ontario* (18). PHU data on population density and "low income" status were taken from the 2013 Census found on the Statistics Canada website (19). "Low income" was used as a measure for socioeconomic status, representing the percentage of families or individuals spending 20% more than average of their before-tax income on food, shelter, and clothing (19).

Methods

Before the data was visualized or analyzed, it was adjusted in order to produce more statistically appropriate results. Given that population sizes differ between different PHUs and that pertussis is a relatively rare disease, there was a difference in the level of error in our estimates since the standard error increases with decreasing sample size. Essentially, regions with smaller populations tend to have less accuracy and are likely to have more extreme results. Regional data were therefore standardized with a method called shrinkage estimation prior to visualizing or analyzing the data (20). This is also called "map smoothing", or empirical Bayesian estimation, in a spatial analysis context. Shrinkage estimation can be done based on iterative likelihood estimates for prior moments (21) or based on methods-of-moments estimators (22). Both methods for producing estimators are valid for rare events, which follow an approximate Poisson distribution.

A choropleth map was generated using the smoothed pertussis incidence. This was done under many considerations. The smoothed incidence by PHU over the entire study period was classified using the quintiles of the empirical distribution. Five colours were used in the map, as it was found that best differentiation between regions was found with 5 colours (23). The scale of the colours was generated based on the quintiles of the distribution of incidence rates (23). The colour scheme chosen was colour-blind adjusted, to accommodate those with colour-blindness (24).

After the data was smoothed and the choropleth map generated, clustering was tested for and specific disease clusters were identified. Clustering was tested for using Moran's I correlation coefficient (25). This test is the spatial equivalent of Pearson's correlation coefficient and values for it generally range between -1 and +1. Large positive values indicate strong clustering whereas large negative values indicate regularity in the distribution, and values close to zero indicate spatial independence (26).

Pertussis clusters, or regions with higher incidence rates than expected through random chance, indicating increased risk in those areas, were identified. This was done with a spatial scan test (27) which is based on scanning windows of fixed area size to identify locations of excess risk. Specifically, flexibly-shaped scanning windows were used because the circle, which is normally used as the scanning window in spatial scan tests, does not always represent the practical shape of disease clusters (28,29). The primary cluster (i.e. the most statistically significant) was identified, as well as any secondary clusters with less significance but potentially greater relative risk for pertussis infection.

A Poisson regression model was fitted using vaccination rates (at 7 and 17 years of age), low income, and population density as predictors. Since this type of model ignores spatial effects, a spatial Poisson regression model was also fitted to see if it held more explanatory value. These models start with the inclusion of all potential predictor variables and sequentially remove the least explanatory variable. Interaction effects between potential predictor variables were also investigated by testing whether the addition of interaction terms improved the model.

The model that was eventually chosen with this method maximized explanatory value while minimizing the number of variables by removing the ones that had the least significant effect.

All mapping and data analysis were done using R Statistical Software as a Geographic Information System (30,31). All statistical tests were also done in R and were evaluated at a significance level of $\alpha=0.05$.

Results and Discussion

Maps of pertussis incidence in southern Ontario

The visualization of pertussis incidence rates (Figure 1) is a geospatial representation of the average annual incidence of the disease over 2005–2016 in southern Ontario. Important to note in the legend is the difference in the length of each quintile because a long 1st and/or 5th quintile would indicate the potential for a geospatial outlier in pertussis incidence.

Clusters

There were three clusters identified by the flexible spatial scan test (Figure 2). The primary cluster was comprised of Toronto and York PHUs. The high population density of these two regions is the chief reason for this being the primary cluster. The number of new cases of a disease is strongly related to the number of susceptible individuals, with population density being a very important factor (32). Additionally, the large populations in these regions results in the largest power or amount of evidence for the presence of a cluster as identified by the flexible scan test.

The other two clusters consisted almost entirely of PHUs made up of rural counties in Southwestern Ontario. Interestingly, the secondary cluster (consisting of Elgin-St Thomas, Huron, Oxford, and Perth PHUs) has a greater relative risk (RR=4.39, p=0.01) and standardized incidence ratio (SIR=4.02, p=0.01) than the other two clusters (Table 1). However, the smaller population means the analysis for this specific cluster is underpowered, which is likely why it is not the primary cluster. Rural southwestern Ontario regions, including Wellington and Dufferin counties, which make up part of the third cluster, have some of the highest incidence rates in southern Ontario. These counties are home to many of Ontario's Mennonite communities, which tend to have lower vaccination rates than the general population (33). It is possible that there is an association between these lower vaccination rates and the increased incidence of pertussis.

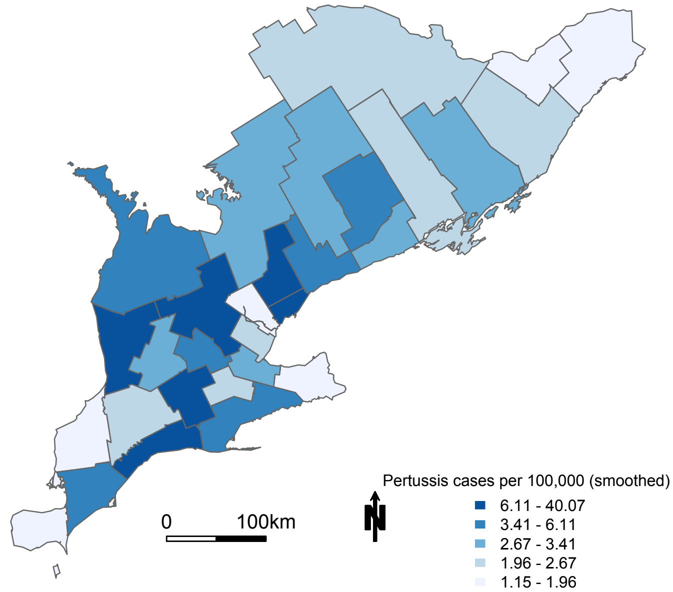

Figure 1. Choropleth map of the smoothed pertussis incidence data in southern Ontario Public Health Units, from 2005 to 2016.

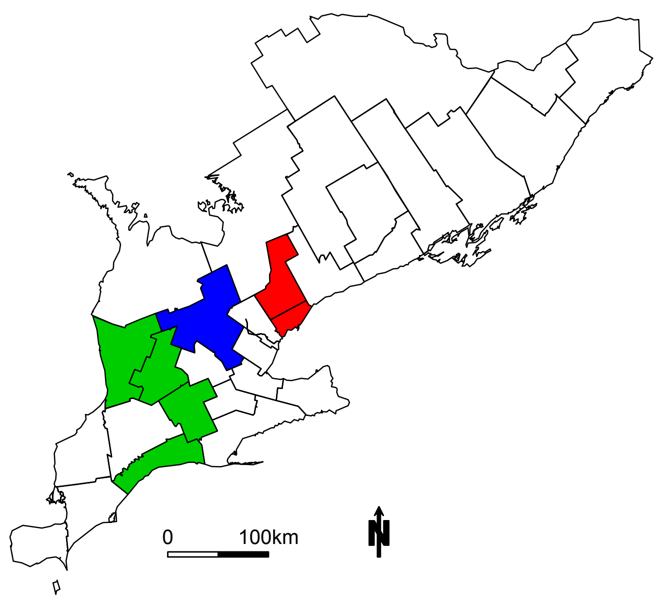

Figure 2. Map of the three clusters of pertussis incidence in southern Ontario Public Health Units, from 2005 to 2016. The primary cluster is represented in red, the secondary cluster is represented in green, and the tertiary cluster is represented in blue.

Table 1. Clusters of pertussis incidence in southern Ontario and their supplementary data.

Public Health Units	**Toronto, York**	**Elgin-St Thomas, Huron, Oxford, Perth**	**Wellington-Dufferin-Guelph**
Cases	3290	733	246
Expected Cases	2033	182	148
SIR (Standardized Incidence Ratio)	1.61	4.02	1.65
RR (Relative Risk)	2.21	4.39	1.68
p-value	0.01	0.01	0.01

Finally, although three clusters of regions at higher risk of pertussis were detected, there was no evidence for an overall pattern of clustering in the pertussis incidence data for southern Ontario. Moran's I was very close to 0 (-0.0345, $p<0.001$) indicating spatial independence. This spatial independence may be explained by the large geographic size of the regions (PHUs) used in our spatial analysis relative to the transmission radius of pertussis among the human population. This transmission area is relative to the activity area of humans, which should be on average much smaller than the PHUs.

Regression model

The model that was used was a spatial Poisson regression model that included low income and population density as predictor variables (Table 2). The presence of overdispersion in the general linear model motivated the use of a spatial model. Overdispersion is when the variance in the data is significantly greater than the mean, when they should be the same in a Poisson model. This was checked by looking at the ratio of deviance residual sum over degrees of freedom, which should be approximately equal to 1. Vaccination rates were not included because no relationship was found between vaccination rates and pertussis incidence. Population density was found to have a statistically significant effect with a relative risk of 1.06 ($p=0.03$). Given this relative risk, for every 100 people/km^2 increase in population density there was a 6% increase in pertussis incidence. Low income was kept in the model despite not being significant (RR=0.35, $p=0.08$) because upon its removal, the size of the effect (i.e. RR) of population density changes dramatically. Thus, low income was identified as a confounding variable that was necessary to maximize explanatory value of the model. It has a relative risk of 0.35, therefore, for every 10%

increase in the proportion of families and individuals categorized as low income, pertussis incidence decreased by a factor of 0.35.

Table 2. Spatial Poisson regression model for pertussis incidence in southern Ontario with Low Income and Population Density as predictor variables.

Variable	**Estimate**	**Standard Error**	**Relative Risk**	**p-value**
(Intercept)	-8.936	0.717	N/A	N/A
Low Income	-0.105	0.058	0.35	0.08
Population Density	0.001	0.0002	1.05	0.03

The finding that vaccination rates were not significant was unexpected. A possible explanation for this finding is that since pertussis is a rare disease in southern Ontario and rates of vaccination are high, the differences between PHUs in vaccination rates are not substantial enough to cause any significant difference in incidence. Low income, defined as the percentage of families or individuals spending 20% more than average of their before-tax income on food, shelter and clothing, was inversely related to pertussis which was also unexpected. However, this finding was not significant and low income is a confounding variable with population density. This confounding is likely the result of the fact that the Toronto PHU is an outlier with the highest population density, and it is also the PHU with the highest value for low income. Thus, this result should be viewed with caution.

Limitations

Although pertussis is a reportable disease in Ontario, there is still a significant underreporting problem that affects the data for pertussis incidence (4). Underreporting of pertussis is likely due to pertussis infections sharing symptoms with other respiratory diseases, coupled with the fact that when symptoms are not overly severe patients often elect to not see a doctor for their cough.

Another very important limitation of this research is the lack of controlling for age when conducting tests or creating maps. Age is a crucial factor in pertussis incidence, since it is a childhood disease. It is possible that different PHUs have varying age distributions in their populations which influences the overall incidence. Age may have played an essential role in creating the clusters that were found and would be a confounding variable when associating these clusters with external factors such as vaccination rates or socio-economic status.

Toronto PHU being an outlier in population density is a limiting factor for making meaningful conclusions based on the regression model. However, removing this PHU from the dataset leaves no significant variables in the regression model. Toronto is such a different environment from the rest of southern Ontario that classifying it alongside the other PHUs should restrict how the model is interpreted. Further, more detailed, analysis of Toronto would be necessary, but the respective data are not available for this project.

Conclusions and prospective research

Pertussis incidence in southern Ontario for the years 2005–2016 has a spatial pattern that has now been mapped. There are three clusters of higher-than-expected incidence in both rural and highly metropolitan PHUs. Vaccination rate was not found to be a significant predictor variable for pertussis incidence. Population density was identified as a significant predictor variable and was included in the spatial regression model. Socioeconomic status (SES), here represented by low income percentage, was a confounder between population density and pertussis incidence. However, it was included to improve the experimental model.

Knowledge of the spatial pattern of pertussis in Ontario as well as its predictor variables is valuable for public health workers who can use it to augment their strategy in reducing the incidence of this infectious disease. Primary care physicians who care for children in the higher-risk clusters can also be better informed of the increased risk their young patients face.

Future research should be directed towards other potential predictor variables of pertussis incidence that were not included in this article. These variables include the different strains of *Bordetella pertussis*, different types of pertussis vaccines, household structure, climate, and other demographic factors. Future research should also be done in other geographic locations (such as another Canadian province, or a US state) to see whether the results here can be replicated elsewhere. Of particular importance is the question of whether vaccination rates have an effect on pertussis incidence in other developed areas.

References

1. Cotter PA, Miller JF. Genetic analysis of the Bordetella infectious cycle. Immunopharmacology. 2000 Jul 25;48(3):253–5.

2. Mattoo S, Cherry JD. Molecular pathogenesis, epidemiology, and clinical manifestations of respiratory infections due to Bordetella pertussis and other Bordetella subspecies. Clin Microbiol Rev. 2005 Apr;18(2):326–82.

3. Crowcroft NS, Stein C, Duclos P, Birmingham M. How best to estimate the global burden of pertussis? Lancet Infect Dis. 2003 Jul 1;3(7):413–8.

4. Cherry JD. The Present and Future Control of Pertussis. Clin Infect Dis. 2010 Sep 15;51(6):663–7.

5. Public Health Ontario. Pertussis (Whooping Cough) [Internet]. Toronto ON: Public Health Ontario. Available from: https://www.publichealthontario.ca/en/diseases-and-conditions/infectious-diseases/vaccine-preventable-diseases/pertussis

6. Centers for Disease Control and Prevention. Pertussis (Whooping Cough) [Internet]. Atlanta: Centers for Disease Control and Prevention; 2017. Available from: https://www.cdc.gov/pertussis/outbreaks/about.html.

7. Cotter PA, Jones AM. Phosphorelay control of virulence gene expression in Bordetella. Trends Microbiol. 2003;11(8):367–73.

8. Melvin JA, Scheller E V., Miller JF, Cotter PA. Bordetella pertussis pathogenesis: Current and future challenges. Nat Rev Microbiol. 2014;12(4):274–88.

9. Lauria AM, Zabbo CP. Pertussis (Whooping Cough) [Internet]. StatPearls. Available from: https://www.ncbi.nlm.nih.gov/books/NBK519008/

10. Andrews R, Herceg A, Roberts C. Pertussis notifications in Australia, 1991 to 1997. Commun Dis Intell. 1997;21(11):145–8.

11. De Melker HE, Schellekens JFP, Neppelenbroek SE, Mooi FR, Rümke HC, Conyn-Van Spaendonck MAE. Reemergence of pertussis in the highly vaccinated population of the Netherlands: Observations on surveillance data. Emerg Infect Dis. 2000;6(4):348–57.

12. Hellenbrand W, Beier D, Jensen E, Littmann M, Meyer C, Oppermann H, et al. The epidemiology of pertussis in Germany: Past and present. BMC Infect Dis. 2009 Feb 25;9.

13. Raguckas SE, VandenBussche HL, Jacobs C, Klepser ME. Pertussis resurgence: Diagnosis, treatment, prevention, and beyond. Pharmacotherapy. 2007 Jan;27(1):41–52.

14. Tanaka M, Vitek CR, Pascual FB, Bisgard KM, Tate JE, Murphy T V. Trends in Pertussis among Infants in the United States, 1980-1999. J Am Med Assoc. 2003 Dec 10;290(22):2968–75.

15. Berke O. Exploratory spatial relative risk mapping. Prev Vet Med. 2005 Oct 12;71(3–4):173–82.

16. Mohebbi M, Wolfe R, Jolley D. A poisson regression approach for modelling spatial autocorrelation between geographically referenced observations. BMC Med Res Methodol. 2011;11(1).

17. Statistics Canada. Health region boundary files. In: Health Indicators. 2015. Available from https://www150.statcan.gc.ca/n1/pub/82-402-x/2011001/reg-eng.htm.

18. Public Health Ontario. Immunization Coverage Report for School Pupils in Ontario: 2013–14, 2014–15 and 2015–16 School Year. Toronto, ON: Queen's Printer for Ontario; 2017.

19. Statistics Canada. Health Profile, December 2013 – City of Toronto Health Unit [Internet]. Ottawa, ON: Statistics Canada; 2013 Dec. Available from: www12.statcan.gc.ca/health-sante/82-228/details/page.cfm?Lang=E&Tab=1&Geo1=HR&Code1=3595&Geo2=PR&Code2=35&Data=Rate&SearchText=City of Toronto Health Unit&SearchType=Contains&SearchPR=01&B1=All&Custom=&B2=All&B3=All.

20. Berke O. Exploratory disease mapping: Kriging the spatial risk function from regional count data. Int J Health Geogr. 2004 Aug 26;3(1).

21. Clayton D, Kaldor J. Empirical Bayes Estimates of Age-Standardized Relative Risks for Use in Disease Mapping. Biometrics. 1987 Sep;43(3):671.

22. Marshall RJ. Mapping disease and mortality rates using empirical Bayes estimators. J R Stat Soc Ser C Appl Stat. 1991;40(2):283–94.

23. Brewer CA, Pickle L. Evaluation of methods for classifying epidemiological data on choropleth maps in series. Ann Assoc Am Geogr. 2002;92(4):662–81.

24. COLORBREWER 2.0 [Internet]. 2018. Available from: http://colorbrewer2.org/#type=sequential&scheme=OrRd&n=3

25. Moran PA. Notes on continuous stochastic phenomena. Biometrika. 1950;37(1–2):17–23.

26. Waller LA, Gotway CA. Applied Spatial Statistics for Public Health Data. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2004. (Wiley Series in Probability and Statistics).

27. Openshaw S, Charlton M, Wymer C, Craft A. A mark 1 geographical analysis machine for the automated analysis of point data sets. Int J Geogr Inf Syst. 1987 Jan 1;1(4):335–58.

28. Takahashi K, Kulldorff M, Tango T, Yih K. A flexibly shaped space-time scan statistic for disease outbreak detection and monitoring. Int J Health Geogr. 2008 Apr 11;7(1):14.

29. Yao Z, Tang J, Zhan FB. Detection of arbitrarily-shaped clusters using a neighbor-expanding approach: A case study on murine typhus in South Texas. Int J Health Geogr. 2011 Mar 31;10(1):23.

30. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria; 2019.

31. Brunsdon C, Comber L. (2015). Introduction to R for Spatial Analysis and Mapping. London: Sage; 2015. 360 p.

32. Tarwater PM. The effects of population density on the spread of disease. Texas Med Cent Diss. 1999 May; Available from: https://digitalcommons.library.tmc.edu/dissertations/AAI9929469.

33. Williamson G, Ahmed B, Kumar PS, Ostrov BE, Ericson JE. Vaccine-preventable diseases requiring hospitalization. Pediatrics. 2017 Sep 1;140(3).

34. Berke O. Choropleth mapping of regional count data of Echinococcus multilocularis among red foxes in Lower Saxony, Germany. Prev Vet Med. 2001 Dec 3;52(2):119–31.