2008

Optimization of Epidemiologic Interventions: Evaluation of Spatial and Non-Spatial Methods That Identify Johne's Disease-Infected Subpopulations Targeted for Intervention

Ariel L. Rivas
University of New Mexico, alrivas@unm.edu

Marcelo Chaffer
Kimron Veterinary Institute, marcelos@moag.gov.il

Gerardo Chowell
Georgia State University, gchowell@gsu.edu

Daniel Elad
Kimron Veterinary Institute

Ori Koren
Israel Dairy Board, oriko@is-d-b.co.il

See next page for additional authors

Recommended Citation

Rivas AL, Chaffer M, Chowell G, Elad D, Koren O, Smith SD, Schwager SJ. Optimization of Epidemiologic Interventions: Evaluation of Spatial and Non-Spatial Methods That Identify Johne's Disease-Infected Subpopulations Targeted for Intervention. Israel Journal of Veterinary Medicine 63, 59-71, (2008).

OPTIMIZATION OF EPIDEMIOLOGIC INTERVENTIONS: EVALUATION OF SPATIAL AND NON-SPATIAL METHODS THAT IDENTIFY JOHNE'S DISEASE-INFECTED SUBPOPULATIONS TARGETED FOR INTERVENTION

Rivas A. L.[1], Chaffer M.[2], Chowell G[3], Elad D[2], Koren O.[4], Smith S. D.[1] and Schwager S. J.[1]

[1] College of Agriculture and Life Sciences, Cornell University, Ithaca, NY 14853, USA.
[2] Department of Bacteriology, Kimron Veterinary Institute, Bet Dagan, 50250, Israel.
[3] School of Human Evolution and Social Change, Arizona State University, Box 872402, Tempe, AZ 85282, USA
[4] Israel Johne's Disease Control Program, Israel Dairy Board, Rishon le Zion, 75054, Israel.
Correspondence: marcelos@moag.gov.il

Keywords: cluster, geographical information systems, network analysis, paratuberculosis

SUMMARY

The potential costs and/or benefits associated with two epidemiological methods were compared. Using the same epidemiologic dataset (74 Israeli dairy herds tested for bovine paratuberculosis of which 57 farms were regarded to be infected, and 619 non-tested herds), the efficacy associated with the identification of the target population where control or preventive measures could be applied was evaluated by: 1) A method that applied geographical information systems (GIS), spatial statistics, network analysis *(infective spatial links or ISL);* and 2) A method that only partially applied spatial techniques. Based on the herd size of tested and non-tested farms, the geographical area of influence of each infected farm was estimated. Using the Euclidean distance between tested farms (distances between 2701 farm pairs), the ISL method calculated two measures of spatial connectivity: the number of links/farm and the ISL index. These measures are analogous to the number of roads connecting a city (links/farm) and the width of a road (index). The more links and/or the greater the average index ("width"), the greater the chances of an infected farm to disseminate an infection (especially to neighboring farms). While not reaching statistical significance, positive indices of Moran's I test for some spatial lags prompted the additional investigation of a subset of 547 farm pairs. This subset included 33 farm pairs (16 individual farms) which displayed ≥ 2 links/farm, and ISL indices >7.5 times greater than average (high ISL farms). Regarding as "cost" the number of infected cows selected to receive an intervention, and as "benefit" the number of susceptible cows within the area of influence of an infected farm, hypothetical interventions implemented on the 16 high ISL farms yielded 39 % greater benefits and occupied a territory 9.5% smaller than decisions based on the 16 farms showing the highest prevalence. The analysis on spatial infective connectivity may lead to earlier, farm-specific and more beneficial, decisions than methods based only on outcomes (later data), such as prevalence.

INTRODUCTION

The detection of infected geographical units that highly influence disease spread ("super-spreaders"), and cost-benefit analysis of potential control measures, are permanent goals of epidemiological research. The recent development of Geographical Information Systems (GIS), contact network epidemiology, and spatial statistics provide new tools to advance these goals (1-8).

Two issues of general epidemiological interest are: 1)The identification of disease clusters, which may go undetected when these tests focus on global disease spatial autocorrelation (9) but do not emphasize the detection of local clustering (1), and 2) the measurement of the different ability for disease dispersal that two spatial points may possess (e.g., two infected farms), even when population sizes and disease prevalences are identical at both sites. While recent research has provided numerous concepts potentially applicable to address these issues (6-8), most reports have used simulated, rather than actual (geo-referenced) data.

An approach to addressing these issues is presented here. Using geo-coded data, contact network epidemiological concepts and spatial statistical methods are integrated. The analysis of infective spatial links (ISL) assesses aspects of the network that infectious diseases may use in their spatial dissemination. A *network* is defined as a collection of spatial *nodes* (e.g., farms) connected by infective *lines* (e.g., roads). The simplest network is composed of a line connecting two points. From network theory (6-8, 10-12), the index ("weight" or "width") of each line, and the number of lines associated with each node, characterize the connectivity of node pairs (e.g., farm pairs). The index can be regarded as analogous to the width of a road. When one farm is found in multiple links, but the other member of each pair is observed only once, it may be hypothesized that the farm

with multiple links (and/or with a higher link index) is the *disseminator* and the farm observed only once in the network is the *recipient* (13). If a farm possesses ≥2 links, it may be suspected that at least one outbound link may exist. Alternatively, the larger the link index ("width"), the greater are the chances of outbound connectivity (dissemination). Therefore, each farm can be characterized by at least 4 attributes: 1) The spatial coordinates of each point (latitude and longtitude), 2) Disease prevalence, 3) The number of links/farm, and 4) the link index.

Although ISL analysis has been applied to assess infected counties (13), infected farms have not yet been investigated using this approach. Records from Israeli dairy farms where paratuberculosis (Johne's disease) has been investigated offer an opportunity to evaluate ISL at the farm level. Paratuberculosis is an infectious, enteric disease, typically observed in adult cows (14, 15). Its causal agent (*Mycobacterium avium* subsp. *paratuberculosis*, or MAP), while not yet classified as a zoonotic agent, has been identified in biopsies of intestinal tissues of humans affected by Crohn's Disease (16).

Although the present study did not investigate the epidemiology of Johne's disease and while it is unknown whether ISL analysis may be applicable to control it (no longitudinal and prospective studies were performed), the availability of data on this disease provides a non-simulated spatial scenario for performing a methodological exercise. The goals of this study were to determine (a) whether infective spatial links (ISL, as described by their number, index, or both, and based on actual ,geo-referenced data) can detect infected sites suspected to be super-spreaders, and (b) whether ISL-based policy-making decisions result in similar or different costs/benefits than decisions based on prevalence alone.

MATERIALS AND METHODS

Disease data

The spatial locations and herd size of 693 Israeli dairy farms were recorded by the Department of Bacteriology, Kimron Veterinary Institute, which also tested 74 of the farms in 2004-5 for the presence of antibodies against Johne's disease (a voluntary, non-randomized sample). Cows ≥ 24-month old (n= 16,562) were tested. Farms with ≥ 3 percent seropositive cows were considered to be infected (17).

Spatial data

Available maps were geo-coded and new ones were built using Geographical Information Systems (GIS).[a-d] The Euclidean distances between every farm pair of the 74 tested farms (*point-to-point distances*) were generated as described previously (13), producing a matrix containing the between farm distances of 2701 farm pairs. *Buffers* (circular areas centered on the centroid of tested farms) were created by GIS using *features of a theme* and *merge*. The *Geo-processing wizard* command *(clip)* was used to determine the herd size of farms located within buffers.

Spatial statistical analysis of the Israeli national dairy farm structure

The structure of Israeli dairy farm spatial locations was investigated. Ripley's K function (18) was calculated with a commercial package.[d] The K function compares the spatial pattern of farms to the pattern expected from a homogenous Poisson point process, assessing whether farms are spatially distributed at random or not (i.e.,clustered).

Spatial statistical analyses of infected farms

Spatial disease autocorrelation (clustering) was investigated using Moran's I statistic, which evaluates the null hypothesis of no autocorrelation between spatial location and disease (i.e., seropositive farms) vs. the alternative hypothesis of disease spatial autocorrelation (i.e., farms closer to each other than average show greater disease prevalence than average) (9). Farms that belong to the same cluster display positive spatial autocorrelation. Moran's autocorrelation coefficient I is given by:

$$I = \frac{n\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} z_i z_j}{\left[\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\right]\left[\sum_{k=1}^{n} z_k^2\right]}$$

where n is the number of dairy farms, i and j are different dairy farms, z_i is the difference between the prevalence in farm i and the overall mean prevalence, z_j is the difference between the prevalence in farm j and the overall mean prevalence, and k is a farm index. The weights w_{ij} are given by: $w_{ij} = f(d_{ij}) = (d_{ij})^{-a}$ where d_{ij} is the Euclidean distance between farms i and j, where $i \neq j$. The parameter a is a diffusion coefficient that quantifies the ability of the infectious agent to spread into space (19). It measures the degree of disease distribution in relation to distance. Small values of α correspond to high dispersal ranges, while large values of α correspond to rather localized spatial distributions. The estimate α* of the parameter a was obtained by maximization of the spatial autocorrelation coefficient (20), such that $I(\alpha^*)\ I(\alpha) \geq$ for all $\alpha>0$.

Correlogram analysis was conducted to assess whether farms displayed any interaction, as expressed by spatial lags, g (distance between farm pairs)(13). A spatial lag g is a range of distances (e.g., g=0 is the range between 0-34 km, g=1 is the range between 35-69 km, g=2 is the range between 70-104 km,...g=10 is the range between 350-384 km). The spatial autocorrelation coefficient for spatial lag g is given by:

$$I(g) = \frac{n\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} z_i z_j}{\left[\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\right]\left[\sum_{k=1}^{n} z_k^2\right]}$$

where the weight wij is 1 for every farm pair (i, j) located within spatial lag g, and 0 for all farm pairs whose distance does not belong to spatial lag g. The correlogram is the graphic display of the spatial autocorrelation coefficient $I(g)$ plotted against spatial lags.

The contribution I_{ij} (g) of an individual spatial link between two nodes (e.g., farms i and j) indicates the strength of the spatial link between the farm pair (i, j). The index I_{ij} (g) considers the link connecting the pair of infected sites (i, j) located within a distance lag g, as indicated by

$$I_g(g) = \frac{m w_{ij} z_i z_j}{\left[\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}\right]\left[\sum_{k=1}^{n} z_k^2\right]}$$

where the weights w_{ij} are given by $(d_{ij})^{-\alpha}$ for every farm pair (i, j) located within spatial lag g, and 0 for all farm pairs whose distance does not belong to spatial lag g.

Spatial infective link measures

The spatial link index ("weight") attributed to each farm of a pair was estimated to be half ("average") of the farm pair index. When one farm was found in multiple farm pairs, the composite ISL index attributed to that farm was calculated as the sum of all averaged link indices. The total number of links per farm was determined by counting all farm pairs that shared a given farm. Farms with either one or ≥ two links were suspected to have inbound or outbound links, respectively.

Statistical analysis

The null hypothesis that the data (of any variable) was normally distributed was investigated by the Ryan-Joiner test, using commercial packages.[c,e] For all tests, $p<0.05$ was considered significant.

RESULTS

Geo-referenced data on the seroprevalence of tested farms (Table 1 and Figure 1A) were used to assess whether disease prevalence was spatially autocorrelated. For that purpose, Moran's *I* was calculated. This test, in its standardized form, measures global autocorrelation between disease and spatial location, yielding results that suggested, without reaching statistical significance, the presence of spatial clustering ($p= 0.0576$, Figure 1B). The between-farm distance assessment indicated that most pairs of tested farms were located within 100 km of each other (Figure 1C).

Analysis of infective spatial links (ISLs) of farm pairs that showed positive indices (those with Moran's *I* values above 0, n=1362 farm pairs) revealed one farm pair (farms *4* and *32*) markedly departing from the rest. While the median ISL index for all farm pairs was 0.0000586 (and the upper boundary of the mean ISL index [99% confidence] was 0.265), the index displayed by the pair including farms *4* and *32* was 101.8 (Figure 1D). The analysis of individual spatial lags (spatial correlogram) failed to identify any Moran's *I* values achieving statistical significance (Figure 1E). However, substantial spatial autocorrelation appeared to occur within the 70-175 km spatial lag.

Because > 14 farm pairs fell into each spatial lag, the 35 km intervals used in the analysis indicated that the sample size was adequate. While farms belonging to the spatial lag between 315 and 350 km showed greater indices, they were not considered for further investigation because (a) they did not show prevalence above 3 percent (the upper limit of background prevalence, [22]), and (b) the total number of observations falling within that lag was less than 100 (as opposed to more than 1400 farm pairs, observed in the 75-175km lag, Figure 1C).

The 547 farm pairs found within the 70-175 km spatial lag showed a median ISL index of 0.054. The ISL index was not normally distributed (Figure 1F), according to the Ryan-Joiner test. That was explained by the presence of a subset (n=33 farm pairs, composed of 16 individual farms) which displayed ISL indices ≥ 0.407, and ≥ 2 links per farm (high ISL farms, Table 2).

The 16 individual farms displayed a median composite ISL index of 1.123 (the averaged sum of all ISL link indices), a median number of links/farm of 3, a median composite ISL index per individual link of 0.39, and a median Johne's disease prevalence of 5.29 percent (Table 3). The 16 high ISL farms included farms *4* and *32* (Figure 2A), the only farm pair suspected to be clustered by the global Moran's *I* test ,although not reaching statistical significance, (Figure 1D). However, not all high ISL farms displayed the highest prevalence levels: 2 of the 16 farms that displayed the highest prevalence (farms *8* and *35*) were not included in the high ISL set (Figure 2B and Table 4).

Within the high ISL farms, a significant linear relationship was observed between the number of links per farm and the composite ISL index (R^2 *[adjusted]:* $r=0.88$, $p<0.001$, Figure 3A). While the correlations between prevalence and ISL indicators were also significant (Figures 3B, C), their coefficients were lower (R^2*[adjusted]:* 0.88 and 0.74, respectively). These findings indicated that ISL indicators were associated with, but not limited by, prevalence data.

To assess the vicinity of infected farms, the geo-referenced location of 619 non-tested Israeli dairy farms was examined. Geostatistical analyses indicated that Israel's national dairy herd was clustered as indicated by Ripley's *K* function, and on average, farms were more densely located within an 18 km distance (Figure 4). That information prompted the generation of circular regions of 18 km diameter (centered on each farm's centroid) as the likely neighborhood potentially influenced by infected farms (Figures 5A, B). Those circles or "buffers" indicated that in the event that a hypothetical control/preventive measure was applied to infected farms displaying high ISL values, the resulting benefit would be 39% higher (and it would occupy a territory 9.5% smaller) than if the same measure was applied to an equal number of farms displaying the highest prevalence. These results were produced by dividing the number of "protected" cows (benefits), located within the 18-km radius from infected farms, over the number of infected cows to which interventions would be applied (costs). Infected high ISL farms included 2696 cows, while 27243 cows were located in their area of influence (27243/2696 or a benefit of 10.11 protected cows per intervened cow). In contrast, the 16 farms showing the highest prevalence values included 3287 cows, and 23942 bovines were in their vicinity (23942/3287), resulting in a benefit/cost of 7.28. Hence, the ISL-based policy was 38.7% (10.11/7.28) more beneficial/less costly than the prevalence-based policy (Figures 5A, B; table 4).

DISCUSSION

This study was not conducted in order to determine paratuberculosis disease prevalence nor to compare intervention

measures applicable to Johne's disease. Rather its primary purpose was to explore two methods directed at identifying the population targeted to receive interventions. Five concepts were considered: 1) prevalence, 2) global disease spatial autocorrelation, 3) local disease spatial autocorrelation, 4) the number of links/site, and 5) the index of individual spatial links. The study was conducted in three phases. The first focused on tested farms, the second phase assessed the spatial structure of the Israeli national dairy herd, asking whether dairy farms were randomly distributed over space or not (clustered), and the third phase compared the two methods, based on information collected in the first and second phases, the area of influence of each infected farm and determining the hypothetical costs and benefits associated with each method.

Due to disease clustering observed in 7 farms (forming 3 clusters, Figure 2A), it was concluded that the Moran's I test may yield false negative results. Such a finding was expected since lack of statistically significant global autocorrelation may coexist with substantial local clustering (1).While correlogram analysis provided an approximation to a spatially explicit method, this method did not indicate exactly where disease clusters could be located, nor did it discriminate between farms of lower and greater prevalences. Prevalence-only oriented methods may differ from the ISL method in one aspect, at least: Prevalence does not necessarily assess connectivity (i.e., a means for disease dispersal). In contrast, GIS-based ISL may identify or calculate: 1) Farms showing connectivity-related indicators of a magnitude greater than average (greater number of links per farm and/or greater ISL index/link), 2) the exact spatial location of farms regarded as more connected (local disease clustering), and 3) cost-benefit like analyses (based on four geo-referenced variables, farm spatial location, herd size, farm composite ISL index, and the number of links/farm).

ISL analysis identified farms hypothesized to be super-spreaders. Both the number of infective links and the "weight" (index) of each link differentiated infected sites beyond the informative potential provided by prevalence. Everything else being equal (e.g., farms of identical herd size and identical prevalence), infected farms may differ in the number of links, the index of each link, or both, which may result in different risks for their (susceptible) neighbors. Spatial connectivity may be suspected to be more relevant in decision-making than prevalence alone because ISL focuses on dissemination which is an early factor or requisite for disease dispersal.. Regardless of intra-herd prevalence, no infectious disease may spread in the absence of a dissemination network. The ISL index and the number of links/farm were significantly and linearly related (Figure 3A). These findings confirmed for the first time to the best of our knowledge, in a spatially explicit, epidemiological study, a linear relationship between similar measures of connectivity predicted previously by the field of Network Theory (11).

Because ISL is relatively insensitive to geometrical shapes associated with target populations, it is suggested that this approach may be more robust than appraches that depend on the geometry of the area of study and/or that of the instrument itself (21). As a result, ISL analysis could detect both "non-fragmented" clusters (e.g., adjacent farms) and "fragmented" clusters (non-adjacent farms, also known as small-world connections [22]).

A greater "benefit" (a lower "cost") was achieved when the farms to be hypothetically intervened were selected on the basis of spatial connectivity. Although only future studies may determine whether the observed difference was due to unique (non-repeatable) conditions or whether greater benefits/lower costs are expected to be associated with the ISL approach, four considerations appear to support the latter alternative.

First, when an infected farm is located near the edge of a territory (e.g., a coastline), that farm, cannot threaten as many susceptible individuals on average, as a farm of identical population size and prevalence that is not located near an edge. When an infected farm is located near an edge, the size of the "benefit" is expected to be smaller, because fewer neighbors are expected.

Second, when the spatial distribution of the population at large is clustered, an infected site tends to be connected with a greater than average number of susceptible sites. For epidemiologic policy, the implication is that measures focusing on clustering (of both infected and susceptible populations) are likely to be more effective than those focusing only on infected farms (i.e., policies that do not consider clustering).

Third, when a cluster of infected sites is detected, by definition the location of infected farms does not vary at random since disease prevalence and spatial location are auto-correlated. As a result, the area of influence (neighborhood) of each infected site will tend to be smaller than the summation of neighborhood areas corresponding to each infected site. In other words, neighborhood areas will tend to coalesce resulting in smaller (merged) "buffers." Decisions applied to infected clusters are expected to result in a smaller "cost" ,when measured as square kilometers, than those applied to non-clustered infected areas. The greater the overlapping of "areas of influence" (due to clustering), the smaller will be the "cost." Control/preventive measures applied to infected clusters will tend to induce larger benefits/lower costs than measures applied to non-clustered (randomly disseminated) infected farms, even if they display high prevalence rates.

Fourth, although the ISL approach also considers disease prevalence, it is feasible to postulate that the nature of the prevalence detected by ISL may differ from that of the prevalence considered by non-spatial approaches. ISL focuses on early predictors of dispersion or earlier prevalence. In contrast, in non-spatial approaches, the prevalence being measured may reflect within-farm disease progression or later prevalence.

The data supported these propositions. The ISL-based selection detected one more cluster of infected farms, which reduced the "cost" of the area of intervention (a 3-farm cluster including farms 6, 40, and 56, Figure 2A). In contrast, the inefficacy of control measures based only on prevalence (increasing "costs" of control measures, without neighbors to be benefited) was suggested by the detection of two infected farms located near an edge (farms 8 and 35, Figure 2B).

The observation that high ISL farms revealed disease clustering should not be construed to mean that the spatial distribution of Johne's disease necessarily results in spatial clustering. While contagious diseases result in spatial clustering, non-contagious diseases may also reveal spatial clustering. Disease clustering may occur when conditions other than contagiousness are spatially clustered over space. For instance, Johne's disease may be associated with some soil conditions, which may be spatially clustered (23).

One limitation of approaches that assess susceptible populations using circular areas is that farm neighborhoods do not necessarily follow this pattern (21). Future studies may take advantage of GIS-generated data and consider additional data on trade, roads, farm size, animal density and/or traffic (24). Another potential source of bias (shared both by ISL and non-spatial [prevalence-based] methods) is the use of percentage data. Percentages are sensitive to population size. Small denominators (small herd size) could over-estimate the true global prevalence by giving false positive results, when only a few animals are infected. In contrast, large herds could underestimate the true global prevalence. The influence of herd size on prevalence was evident in the case of farm 32, where the hypothetical addition of just one more infected cow (in a herd of 149 cows) would have increased prevalence to 7.4 % (as opposed to the observed prevalence of 6.7 %, Table I), reducing the p value of Moran's I from 0.0578 to 0.0488 (Figure IB).

ISL analysis may inform policy in at least three ways: 1) by detecting local disease clusters (that are likely to be undetected by global spatial autocorrelation tests), 2) by estimating inter-site disease connectivity (a concept not considered by non-spatial prevalence-based studies), and 3) by quantifying the impact of measures potentially applied at specific locations and populations by use of a cost-benefit like analysis based on (spatially explicit) epidemiologic and demographic data.

REFERENCES

1. Anselin L. Local indicators of spatial association–LISA. Geogr Analysis 27:93-115, 1995.
2. Rivas AL, Tennenbaum SE, Aparicio JP, et al. Critical Response Time (time available to implement effective measures for epidemic control): model building and evaluation. Can J Vet Res 67:307-315, 2003.
3. Rivas AL, Smith S, Sullivan PJ, et al. Identification of geographical factors associated with early epidemic spread of foot-and-mouth disease. Amer J Vet Res 64:1519-1527, 2003.
4. Koopman J. Modeling infection transmission. Annu Rev Public Health 25: 303-326, 2004.
5. Kulldorf M. Tests of spatial randomness adjusted for an inhomogeneity: a general framework. J Amer Stats Assoc 101:1289-1305, 2006.
6. Pastor-Satorras R, Vespignani A. Epidemic spreading in scale-free networks. Phys Rev Lett 86:3200-3203, 2001.
7. Albert R, Barabási A-L. Statistical mechanics of complex networks. Rev Mod Phys 74:47-97, 2002.
8. Meyers LA, Pourbohloul B, Newman MEJ, et al. Network theory and SARS: predicting outbreak diversity. J Theoret Biol 232:71-81, 2005.
9. Moran PAP. Notes on continuous stochastic phenomena. Biometrika 37:17-23, 1950.
10. Chowell G, Hyman JM, Eubank S, Castillo-Chavez C. Scaling laws for the movement of people between locations in a large city. Phys Rev E 68: 066102, 2003.
11. Barrat A, Barthélemy M, Pastor-Satorras R, Vespignani A. The architecture of complex weighted networks. PNAS 101: 3747-3752, 2004.
12. Lasne A. Complex networks: from graph theory to biology. Lett Math Phys 78: 235-262, 2006.
13. Chowell G, Rivas AL, Smith SD, Hyman JM. Identification of case clusters and sites ofgreater infective connectivity in foot-and-mouth disease epidemics. Amer J Vet Res 67:102-113, 2006.
14. Chiodini RJ, Van Kruiningen HJ, Merkal RS. Ruminant paratuberculosis (Johne's disease): the current status and future prospect. Cornell Vet 74:281-262, 1984.
15. Chiodini RJ, Rossiter, CA. Paratuberculosis: a potential zoonosis? Veterinary Clinics of North America - Food Animal Practice 12:457-467, 1996.
16. Bull TJ, McMinn EJ, Sidi-Boumedine K, et al. detection and verification of Mycobacterium avium subsp paratuberculosis in fresh ileocolonic mucosal biopsy specimens from individuals with and without Crohn's disease. J Clin Microbiol 41: 2915-2923, 2003.
17. Chaffer M, Rivas AL, Elad D, Koren O, Garazi S, Chowell G, Schwager SJ. Receiver operating characteristic-based accuracy assessment of a serological test used to detect Johne's disease in Israeli dairy herds. Can J Vet Res 72:18-26, 2008.
18. Ripley BD. Modelling spatial patterns. J Roy Stat Soc B 39:172-212, 1977.
19. Tobler WR. Geographical filters and their inverses. Geogr Analysis 1: 234-253, 1969.
20. Cliff AD, Ord JK. Measures of autocorrelation in the plane; and distribution theory for the join-count, I, and c statistics. In: Cliff AD, Ord JK, eds. Spatial processes: models and applications. London: Pion Ltd, 34-50, 1981.
21. Assunçao R, Costa M, Tavares A, Ferreira S. Fast detection of arbitrarily shaped disease clusters. Stat Med 25:723-742, 2006.
22. Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. Nature 393:440-442, 1998.

23. Ward MP, Pérez AM. Association between soil type and paratuberculosis in cattle herds.Am J Vet Res 65:10-14, 2004.

24. Rivas AL, Schwager SJ, Smith S, Magri A. Early and cost-effective identification of high risk/priority control areas in foot-and mouth disease epidemics. J Vet Med B 51:263-271, 2004.

FOOTNOTES

a Arc View GIS 3.3, ESRI, Redlands, Calif.
b Arc View 8.0, ESRI, Redlands, Calif.
c Matlab 7.0.1, Mathworks Inc., Natick, Mass.
d S+ Spatial Stats 6.2, Springer, NY
e Minitab 14, Minitab, State College, PA

Table 1. Herd prevalence (percentage of MAP-seropositive cows) in 57 Israeli dairy herds

Rank	Farm ID	Prevalence (%)	Rank	Farm ID	Prevalence (%)
1	*4*	9.195	30	*28*	2.273
2	*1*	7.778	31	*11*	2.174
3	*15*	7.453	32	*2*	2.083
4	*29*	6.818	33	*43*	2.020
5	*32*	6.711	34	*21*	1.880
6	*24*	6.667	35	*60*	1.845
7	*74*	6.667	36	*13*	1.695
8	*68*	5.556	37	*9*	1.465
9	*64*	5.405	38	*20*	1.429
10	*35*	5.022	39	*59*	1.266
11	*22*	5.021	40	*18*	1.254
12	*50*	4.867	41	*10*	1.132
13	*5*	4.400	42	*67*	1.119
14	*31*	4.132	43	*62*	0.969
15	*37*	3.731	44	*52*	0.763
16	*16*	3.659	45	*70*	0.696
17	*56*	3.636	46	*48*	0.692
18	*39*	3.620	47	*34*	0.685
19	*40*	3.584	48	*17*	0.680
20	*8*	3.571	49	*45*	0.592
21	*27*	3.571	50	*53*	0.515
22	*6*	3.297	51	*19*	0.446
23	*71*	3.279	52	*14*	0.426
24	*33*	3.245	53	*47*	0.413
25	*12*	3.165	54	*46*	0.405
26	*30*	2.632	55	*58*	0.397
27	*51*	2.620	56	*72*	0.313
28	*49*	2.295	57	*66*	0.302
29	*7*	2.273			

Table 2. Israeli farm pairs potentially influencing Johne's disease spread (with ≥ 2 ISL links)

Farm pair rank	Members of the pair (farm IDs		ISL index contributed by the farm pair
1	*1*	*15*	2.32654
2	*15*	*29*	2.31513
3	*15*	*24*	2.23895
4	*4*	*29*	2.23804
5	*4*	*24*	2.16440
6	*1*	*74*	1.79358
7	*29*	*32*	1.42056
8	*24*	*32*	1.37382
9	*5*	*15*	1.02990
10	*4*	*40*	0.90417
11	*22*	*68*	0.82982
12	*4*	*68*	0.78917
13	*4*	*5*	0.76082
14	*15*	*27*	0.68273
15	*1*	*35*	0.67639
16	*4*	*27*	0.66000
17	*4*	*6*	0.65569
18	*15*	*31*	0.59154
19	*1*	*29*	0.58301
20	*32*	*40*	0.57007
21	*1*	*24*	0.56383
22	*4*	*56*	0.50030
23	*32*	*68*	0.49914
24	*15*	*56*	0.49568
25	*5*	*32*	0.48953
26	*4*	*50*	0.48459
27	*22*	*50*	0.46619
28	*22*	*29*	0.42159
29	*27*	*32*	0.41892
30	*31*	*74*	0.41808
31	*6*	*32*	0.41645
32	*29*	*37*	0.41202
33	*22*	*24*	0.40772

The following 16 farms displayed ≥ 2 links each (they were found in 2 or more pairs): 1 (5 links), 15 (7 links), 29 (6 links), 24 (6 links), 4 (9 links), 74 (2 links), 32 (7 links), 5 (3 links), 40 (2 links), 22 (4 links), 68 (2 links), 27 (3 links), 6 (2 links), 31 (2 links), 56 (2 links), and 50 (2 links).

Table 3. Israeli dairy farms with high ISL indicators and/or high prevalence

Farm ID (16 farms with high ISL links)	Composite (averaged) ISL index*	Number of links per farm	Median index/ link	Farm ID (16 farms with highest prevalence)	Prevalence (%) in 16 farms of highest ISL links **	Prevalence (%) in 16 farms of highest prevalence
15	4.8402	7	0.5087	*4*	7.453	9.195
4	4.5785	9	0.6914	*1*	9.195	7.778
29	3.6951	6	0.5943	*15*	6.818	7.453
24	3.3743	6	0.3707	*29*	6.667	6.818
1	2.9716	5	0.5623	*32*	7.778	6.711
32	2.5950	7	0.6158	*74*	6.711	6.667
22	1.8110	4	0.4527	*24*	5.021	6.667
5	1.1401	3	0.5529	*68*	4.400	5.556
74	1.1058	2	0.3800	*35*	6.667	5.022
27	0.8808	3	0.3707	*22*	3.571	5.021
68	0.8095	2	0.2936	*50*	5.556	4.867
40	0.7371	2	0.4047	*5*	3.584	4.400
6	0.5360	2	0.3685	*31*	3.297	4.132
31	0.5048	2	0.2680	*56*	4.132	3.636
56	0.4979	2	0.2524	*40*	3.636	3.584
50	0.4753	2	0.2489	*8*	4.867	3.571

* Calculation of composite (averaged) ISL index (example from farm 4, based on Table 1):
(2.23804 + 2.16440 + 0.90417 +0.78917 +0.76082 + 0.66000 + 0.65569 + 0.50030 + 0.48459)/2= 4.57859.

** While both lists include the same number of farms, two farms differ between the ISL-based and the prevalence-based lists: farms 6 and 27 are included in the ISL list while farms 8 and 35 are part of the prevalence list.

Table 4. ISL-based vs. prevalence-based cost-benefit like ratio

Farms with high ISL composite indices		**Farms with high prevalence**	
Farm ID	Herd size (cows)	Farm ID	Herd size (cows)
29	44	*29*	44
24	45	*24*	45
56	55	*56*	55
27	56	*4*	87
4	87	*1*	90
1	90	*32*	149
6	91	*15*	161
32	149	*68*	216
15	161	*50*	226
68	216	*22*	239
50	226	*5*	250
22	239	*40*	279
5	250	*8*	280
40	279	*74*	345
74	345	*31*	363
31	363	*35*	458
Cows within infected herds: 2696 *Untested cows within vicinity* *(18km-diameter circle): 27243 (266 herds)* *Benefit/cost ratio (27243/2696): 10.105*		*Cows within infected herds: 3287* *Untested cows within vicinity* *(18km-diameter circle): 23942 (193 herds)* *Benefit/cost ratio (23942/3287: 7.284*	

*Added benefit of the ISL-based over the prevalence-based selection: 10.105/7.284=**38.73%***

LEGENDS

Figure 1.

Prevalence of paratuberculosis (Johne's disease) in 74 Israeli dairy farms and assessment of global disease spatial autocorrelation. A:

Location and prevalence (percent of seropositive cows) of 74 tested and 619 non-tested farms. **B:** *p* value of Moran's I, expressed as a function of parameter *a* (see text for definition). The minimal *p* value observed was 0.0576, corresponding to $a = 2.79$ (a non-statistical significant value, although approaching significance). **C**: Number of farm pairs per spatial lag (km). Notice that, except the highest spatial lag, all other spatial lags included > 65 farm pairs (i.e., the number of farm pairs included in spatial lags < 350 km was sufficiently large). **D:** Distribution of all positive infective spatial links (those > 0, median= 0.0000586, n=1362 farm pairs). The highest link index (farms *4* and *32*) was 101.08. **E:** Spatial correlogram (each spatial lag=35 km). Notice that no Moran's *I* value (in standard form) reached 1.66 (the minimal value associated with statistical significance). **F:** Distribution of (log-transformed) positive infective spatial links observed within 70-175 km spatial lags (n=547 farm pairs, median: 0.054). The Ryan-Joiner test failed to indicate a normal data distribution.

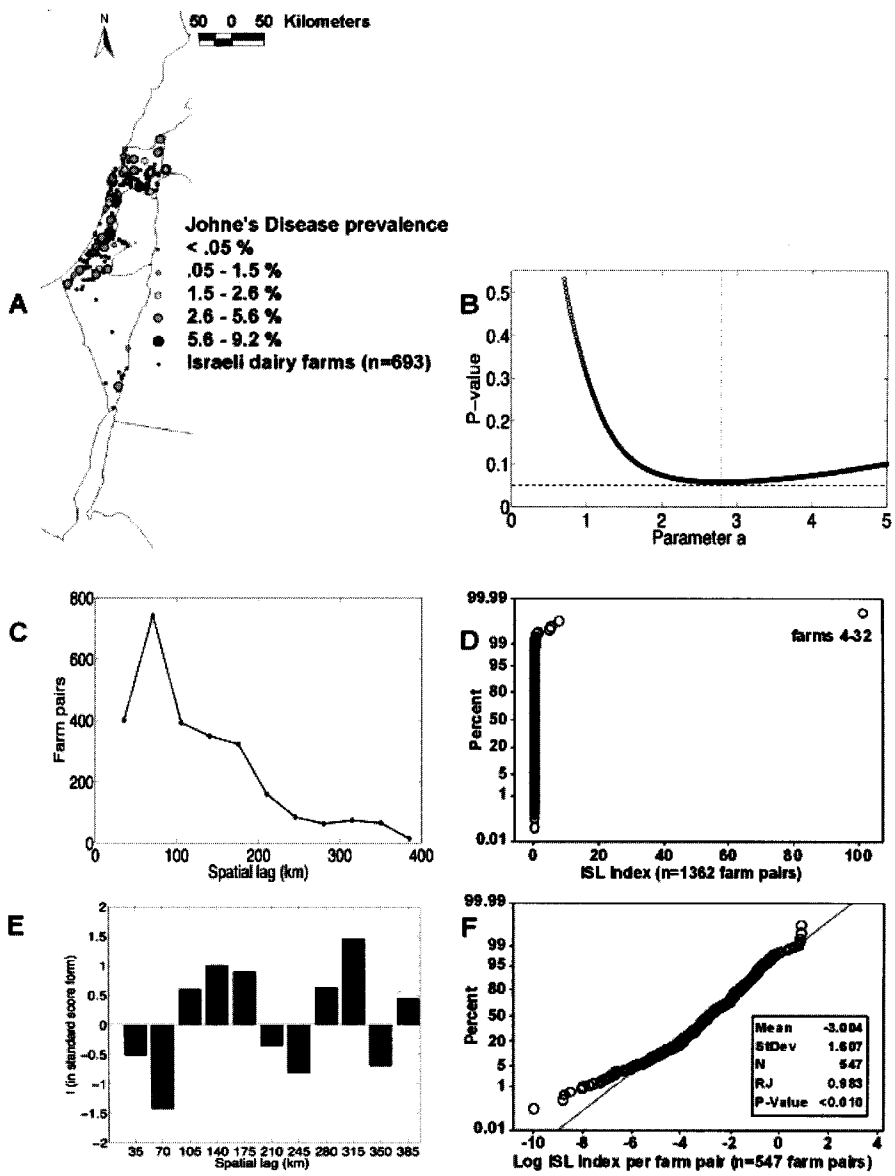


Figure 2.

Infective spatial links and prevalence in Israeli dairy farms displaying the highest values.

The 16 farms that showed the highest (composite) ISL link indices (those with ≥ 2 links/farm) are shown in **A**. For illustrative purposes, the link index generated by the first 25 of the 33 farm pairs is shown (the whole list is reported in Table II). To facilitate comparisons, the same number of farms displaying the highest prevalence is shown in **B**. Notice that the close distance between farms *24* and *29* and between farms *4* and *32* gives the appearance of overlapping when in fact, they did not overlap. Polygons indicate 3 and 2 clusters of infected farms in **A** and **B**, respectively.

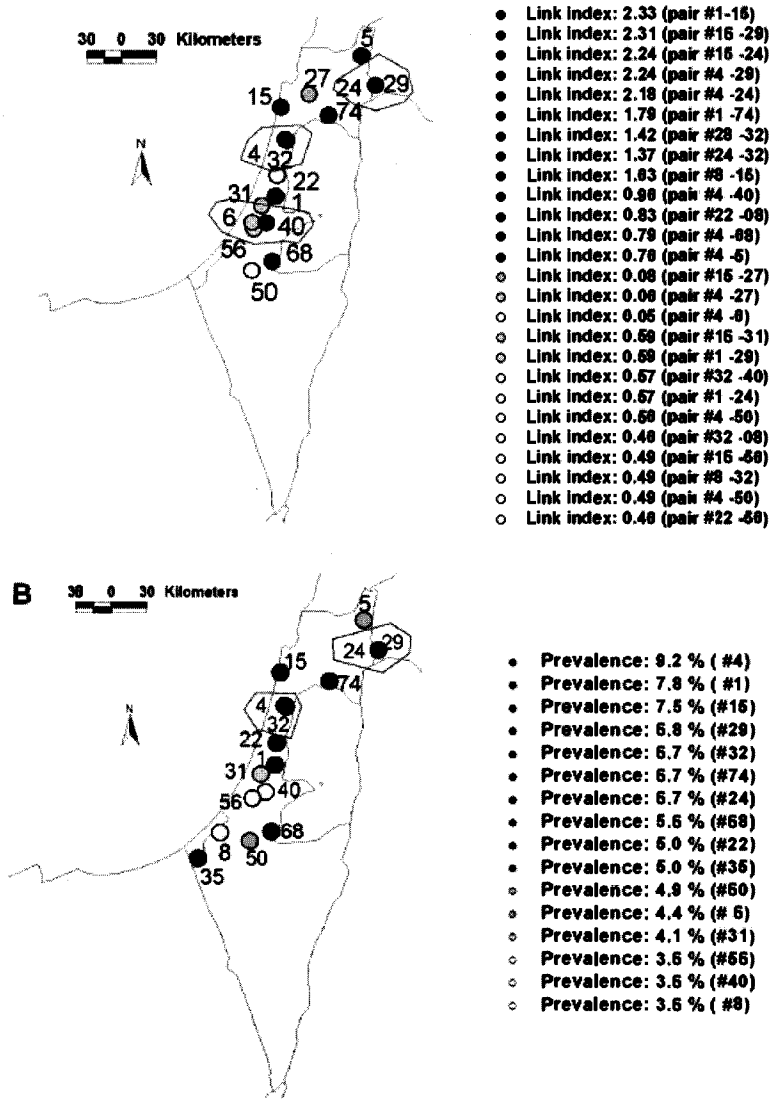


Figure 3.

Relationships within ISL indicators and between ISL indicators and prevalence. A:

Regression analysis between: (**A**) the ISL index and number of links/farm (those with ≥2 links per farm, (**B**) the ISL index and prevalence, and (**C**) the number of links/farm and prevalence.

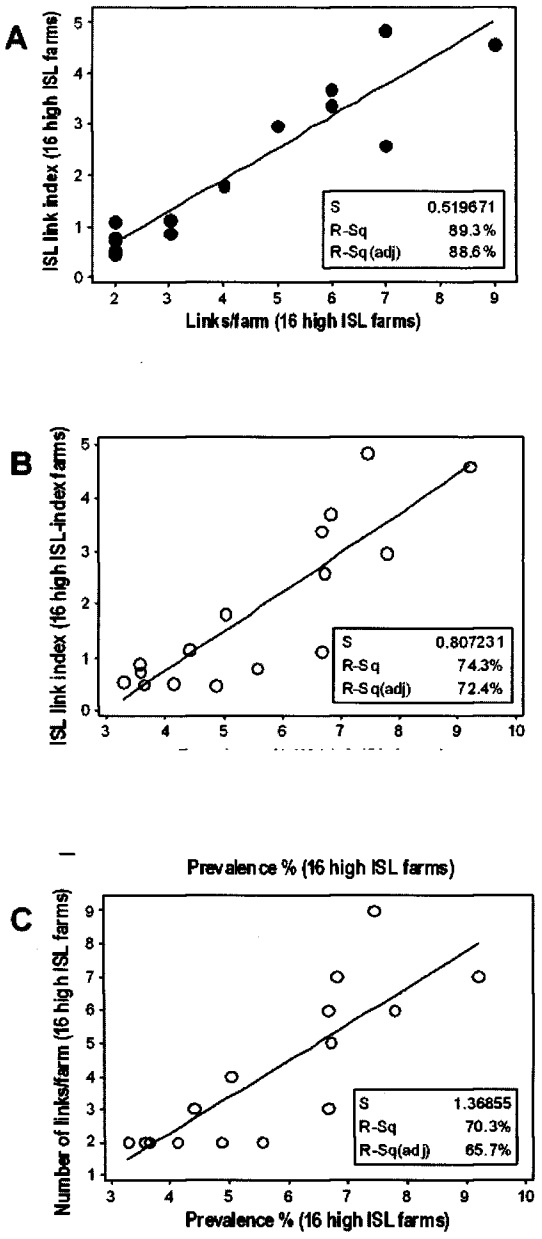
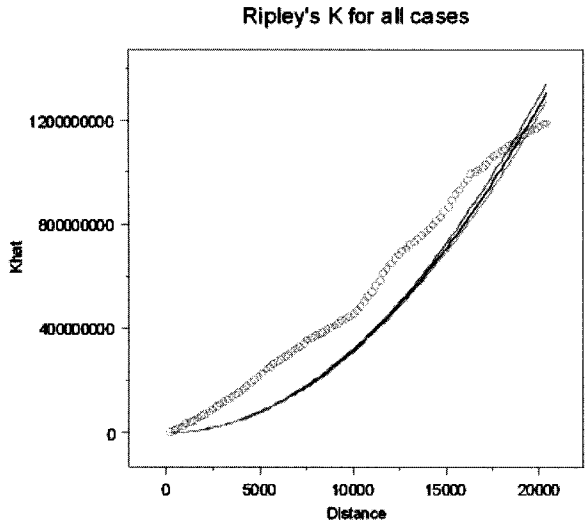


Figure 4. Cluster assessment of non-tested Israeli dairy farms.

Ripley's *K* analysis of the distance (meters) between Israeli dairy herds, based on 100 simulations. Lines represent the maximum, mean and minimum values of *K*hat for simulated data showing no spatial clustering. Points above the lines indicate statistically



significant clustering. Results indicate farm clustering within 18,000 meters (18 km).

Figure 5.

Cost-benefit analysis of ISL-based and prevalence-based methods.

Determination of denominator data (susceptible herds within each infected farm's neighborhood): herd size of susceptible herds located within a 18-km diameter from each infected farm. **A**: those located within the vicinity of the 16 high-ISL farms. **B**: those located within the vicinity of the 16 high-prevalence farms.