IDENTIFYING THE SOURCE OF FOODBORNE DISEASE OUTBREAKS USING SPATIAL STATISTICS METHODS.

Main author: Sandra Rudeloff (Kuehne Logistics University - KLU)

Co-authors: Hanno Friedrich

INTRODUCTION

Foodborne diseases can spread widely and have a massive impact on public health. This phenomenon is exacerbated by the rising complexity and globalisation of the international food network. In the summer of 2011, an outbreak caused by Shiga toxin-producing Escherichia coli (STEC) O104:H4, spread by sprouts grown in Germany, caused 54 deaths and 4321 cases of illness in 16 countries over a nine-week period. Since conventional investigation processes require a high degree of manual work and given the high complexity of the international food trade network, authorities often identify the contamination source only weeks after the outbreak or not at all. As the damage caused increases with the duration of an outbreak, the rapid identification of the contaminated food item and its geographical source is vital to mitigating its proliferation.

Here, we present a spatial pattern recognition approach that has the potential to reduce the time needed to identify possible contamination outlets. Applied to different food outlet patterns, such as retail chains, restaurant chains or warehouse delivery regions, the algorithm helps to narrow down the list of food outlets most likely to be involved.

METHODOLOGY

Our goal is to identify the food outlets that are most likely to have distributed contaminated food products by comparing the spatial point pattern of the outbreak to different point patterns of food outlets (such as a retail chain or a restaurant chain). As a result, the algorithm delivers a ranked list of the food outlets whose spatial relationship to the outbreak pattern is the strongest.

For each food outlet pattern, we determine whether there is a spatial dependence between it and the outbreak pattern. Spatial statistics methods, namely bivariate point pattern analysis, can be used for this. The bivariate version of Ripley's K-function estimates the spatial dependence between two different types of location in a geographical region. It does this by identifying the distances within which spatial dependence is evident and testing the statistical significance of the observed clustering relative to the expected distribution of points if there were no spatial dependence. For each food outlet pattern, the bivariate K-function gives us the number of food outlet locations within a fixed distance of the outbreak locations standardised by intensity.

RESULTS

To measure the performance of the algorithm, we generate artificial outbreak scenarios based on different food outlet patterns that are simulated as having distributed the contaminated product. For each of these artificial outbreak patterns, two steps are taken:

Firstly, the algorithm detects whether there is a significant spatial dependence between the outbreak and each of the food outlet patterns by calculating the K -alue for each combination. This observed K-value is then compared to the K-values for 500 outbreak patterns that are randomly generated through Monte Carlo simulation and that depend only on the population density. If the observed K-value is larger than the maximum expected K-value under complete spatial randomness, the clustering is significant.

Secondly, the algorithm creates a ranked list of the most similar food outlet patterns that narrows down the options for further investigation.

The predictive power is defined as the capability of the algorithm to correctly identify the food outlet that is simulated as having distributed the contaminated product or include it in a shortlist of the food outlet patterns most likely to be involved given the artificially generated outbreak scenario.

DISCUSSION

In the case of an outbreak analysis, the algorithm shall help investigators to narrow down the options for further investigation in order to accelerate the process and prevent further illness. This analysis shows that, when information on outbreak case locations and food outlet locations is available, spatial statistics can help to identify the food outlets that are likely to be distributing the contaminated product. The algorithm takes into account that people exhibit a certain shopping behaviour, therefore, all cases within a certain distance are considered. However, the algorithm does not consider unregular shopping behaviour, such as traveling, which can only be revealed by interviewing sick people.

In conclusion, the proactive analysis of spatial data is intended to complement and guide outbreak investigations by identifying a shortlist of food outlets that are likely to be involved. Thus, the investigation process can be accelerated. As a result, measures such as the removal of the contaminated item from sale and the initiation of product recalls can be taken at an earlier stage of the outbreak and further proliferation of the illness can be prevented.