

# Identifying farms contributing the most to potential for transmission of infections in Scottish sheep network

Victoriya V. Volkova, Richard Howey, Nicholas J Savill and Mark E J Woolhouse  
Epidemiology Group, Centre for Infectious Diseases, University of Edinburgh, UK

## Introduction

Livestock movements connect farms into networks; properties of these networks impact on the potential for transmission of infections among the farms. We consider the network of sheep movements among Scottish farms, and the potential for transmission of an infection through this network. For a given year, the network can be represented by a contact matrix,  $A$ , where  $a(i,j)$  (the element in  $i$ th row and  $j$ th column) is 1 if there is a movement from farm  $j$  to farm  $i$  and 0 otherwise. For some disease scenarios the contact (and so  $a(i,j)$  entries) may need to be weighted by the number of batches or the number of animals moved from farm  $j$  to farm  $i$ . In either case, the expected number of secondary infected farms from an individual infected farm is proportional to the dominant eigenvalue of the contact matrix,  $\epsilon$  (Barbour 1978; Diekmann *et al.*, 1990). We therefore can use  $\epsilon$  to evaluate the contribution of the first, second and higher order moments of the contact network to the basic reproduction number,  $R_0$ , for an infection. Contact patterns of individual farms are highly heterogeneous; therefore targeting interventions at farms contributing the most to  $R_0$  is likely to be efficient. The contribution to  $R_0$  of a set of farms can be explored as the difference between  $\epsilon$  for the complete network and  $\epsilon$  for the resultant network when these farms and the contacts they make with the other farms are removed. Noting the '20-80' rule (at most 20% of farms contribute at least 80% of transmission potential – Woolhouse *et al.* 1997, 2005) we focus here on the sets of size  $0.2N$  from a network of  $N$  farms. Ideally we would compare all possible subsets of size  $0.2N = M$  from the total  $N$  to find the set targeting which achieves the greatest reduction in  $R_0$ . However, this would require  $N!/M!(N-M)!$  calculations of  $\epsilon$ , and is not feasible for large networks (the Scottish sheep movement network contains 15-16,000 farms each year). We evaluated several alternative methods for identifying the  $M$  farms contributing the most to  $R_0$ ; the two methods performing the best are presented below. We then compared the reductions in  $R_0$  when the  $M$  set of farms was identified using their contact information for the year of interest (current year) versus from the preceding year - the information most likely to be available in practice.

## Methods

We considered 4 one-year periods: 01/07/2003-30/06/2004, 01/07/2004-30/06/2005, 01/07/2005-30/06/2006 and 01/07/2006-30/06/2007. For each period, a contact matrix was constructed using each of the 3 weightings of contact between the farms: unweighted, by the number of batches moved and by the number of sheep moved.

**Method 1.** Starting from the complete year's contact matrix of size  $N$ , obtain  $\epsilon$  and identify the farm with the largest cross-product of the number of in-contacts and the number of out-contacts (both either unweighted or weighted). Remove the farm with the largest cross-product and its contacts. For the resultant network of size  $(N-1)$  calculate  $\epsilon$  and the cross-products for the farms; identify and remove the farm with the largest cross-product. Repeat until  $M$  farms have been identified. Calculate dominant eigenvalue for the resultant matrix,  $\epsilon'$ .

**Method 2.** Starting from the complete year's contact matrix, obtain  $\epsilon$  and the corresponding eigenvector. Identify the farm with the largest component in the eigenvector and remove this farm with its contacts. For the resultant network of size  $(N-1)$  obtain  $\epsilon$  and the corresponding eigenvector, identify the farm with the largest component in the eigenvector and remove this farm with its contacts. Repeat until  $M$  farms have been identified. Calculate  $\epsilon'$ .

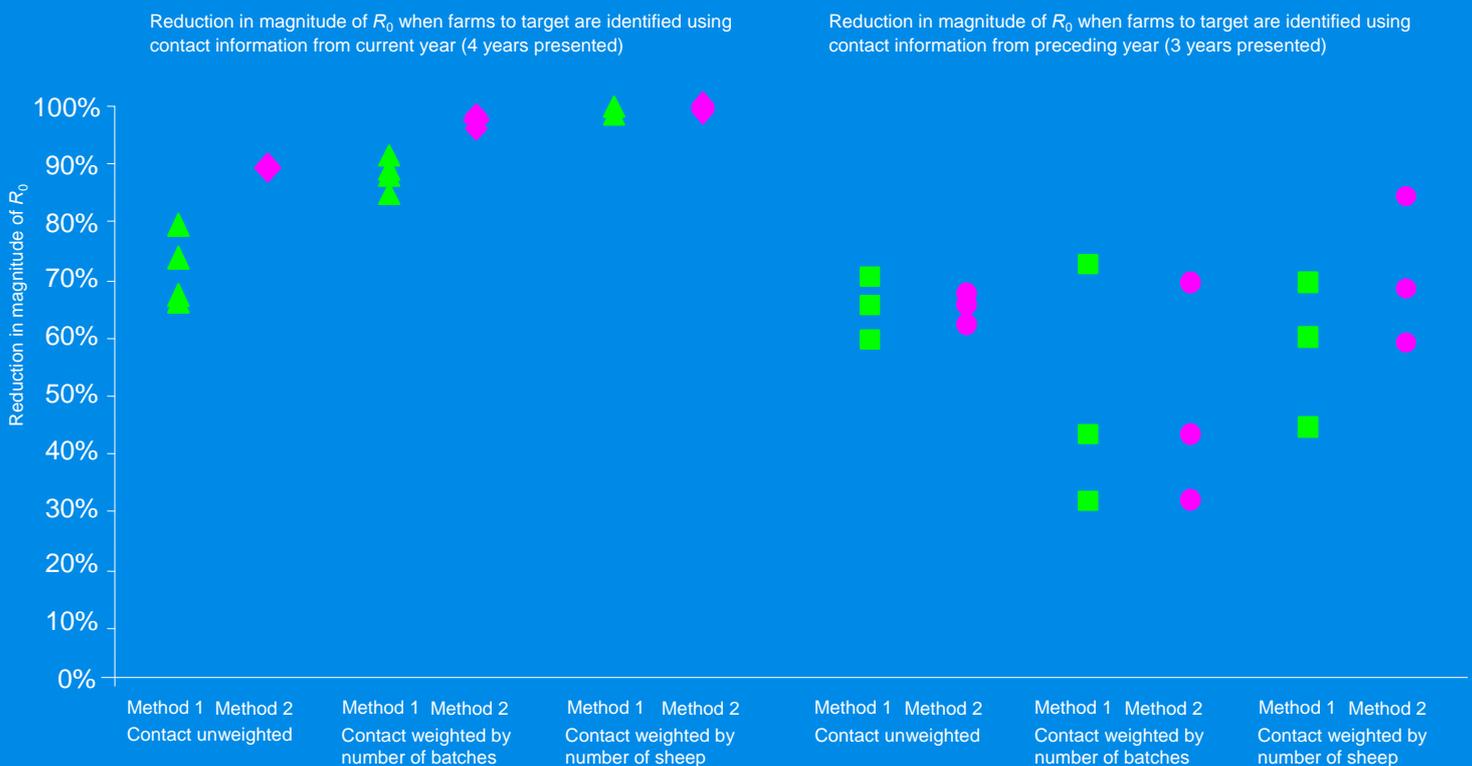
**Applying contact information from the preceding year.** Obtain  $\epsilon$  for the complete farm contact matrix for the year of interest. Identify the subset of  $M$  farms (by either Method 1 or 2) in the preceding year's network. Remove these farms and their contacts from the network of year of interest, and calculate  $\epsilon'$ .

**Reduction in the magnitude of  $R_0$ .** The reduction in the magnitude of  $R_0$  in each case was evaluated as  $(1 - \epsilon'/\epsilon)$ .

## Results and Conclusions

The figure below summarizes the results. When contact information from the year of interest is used, Method 2 outperforms Method 1 in identifying a 20% subset of farms contributing the most to  $R_0$  in the Scottish sheep network for all scenarios considered (left panel). However, in practice, the information available is likely to be that from the preceding year. In case of the Scottish sheep network, such information is much less valuable and its utility greatly varies from year to year, especially for disease scenarios where the numbers of batches or animals sent between the farms are important (right panel). Variation in contact patterns of individual farms, among other factors, affects the contribution of the network's properties to and the final magnitude of  $R_0$  in the network. As to the processes underlying this variation or change, investigating characteristics of farms consistently or intermittently appearing in the subset contributing the most to  $R_0$  may provide further insights.

### Comparing methods for identifying a 20% subset of farms contributing the most to $R_0$ in Scottish sheep farm network in one year



**Acknowledgements** This work was undertaken through the Scottish Government funded Centre of Excellence in Epidemiology, Population Health and Infectious Disease Control.

## References

- Barbour, A. D. (1978). Macdonald's model and the transmission of bilharzia. *Trans Roy Soc Trop Med Hyg* 72: 6-15.
- Diekmann, O., J. A. P. Heesterbeek. *et al.* (1990). On the Definition and the Computation of the Basic Reproduction Ratio  $R_0$  in Models for Infectious-Diseases in Heterogeneous Populations. *Journal of Mathematical Biology* 28:365-382.
- Woolhouse, M.E.J., C. Dye, *et al.* (1997). Heterogeneities in the transmission of infectious agents: implications for the design of control programs. *Proc Natl Acad Sci USA*, 94:338-42.
- Woolhouse, M.E.J., D.J. Shaw, *et al.* (2005). Epidemiological implications of the contact network structure for cattle farms and the 20-80 rule. *Biol Lett.* 1:350-352.