
SHORT REPORT

Lessons from Nigeria: the role of roads in the geo-temporal progression of avian influenza (H5N1) virus

A. L. RIVAS^{1,2*}, G. CHOWELL^{3,4}, S. J. SCHWAGER², F. O. FASINA^{5,6},
A. L. HOOGESTEIJN⁷, S. D. SMITH², S. P. R. BISSCHOP⁶, K. L. ANDERSON¹
AND J. M. HYMAN⁸

¹ *Population Health and Pathobiology, College of Veterinary Medicine, North Carolina State University, Raleigh, NC, USA*

² *Biological Statistics and Computational Biology, and Institute for Resource Information Sciences, Cornell University, Ithaca, NY, USA*

³ *School of Human Evolution and Social Change, Arizona State University, Tempe, AZ, USA*

⁴ *Division of Epidemiology and Population Studies, Fogarty International Center, National Institutes of Health, Bethesda, MD, USA*

⁵ *National Veterinary Research Institute, PMB 01, Vom, Plateau, Nigeria*

⁶ *Poultry Reference Centre, Faculty of Veterinary Science, University of Pretoria, South Africa*

⁷ *Human Ecology, CINVESTAV, Mérida, México*

⁸ *Mathematical Modeling and Analysis, Los Alamos National Laboratory, Los Alamos, NM, USA*

(Accepted 1 July 2009)

SUMMARY

The daily progression of the 2006 (January–June) Nigerian avian influenza (AI H5N1) epidemic was assessed in relation to both spatial variables and the generation interval of the invading virus. Proximity to the highway network appeared to promote epidemic dispersal: from the first AI generation interval onwards >20% of all cases were located at <5 km from the nearest major road. Fifty-seven per cent of all cases were located ≤31 km from three highway intersections. Findings suggest that the spatial features of emerging infections could be key in their control. When the spatial location of a transmission factor is well known, such as that of the highway network, and a substantial percentage of cases (e.g. >20%) are near that factor, early interventions focusing on transmission factors, such as road blocks that prevent poultry trade, may be more efficacious than interventions applied only to the susceptible population.

Key words: Avian flu, emerging infections, geographical information systems, surveillance.

Historically, epidemiological control policy has focused on the population of hosts more often than it has addressed transmission factors. Yet, factors like the highway network may be important to epidemic dispersal. For instance, the prevalence of some human infectious diseases may be higher in municipalities

crossed by or next to major roads than in inhabited territories not linked by interstate highways [1]. In viral diseases affecting animals, proximity to interstate highways and/or location in areas of high road density may promote epidemic dispersal [2]. These concepts were considered in the analysis of the daily progression of the 2006 Nigerian highly pathogenic avian influenza (HPAI) virus epidemic (subtype H5N1), where spatial and bio-temporal factors were explored. The spatial component analysed was the

* Author for correspondence: Dr A. L. Rivas, College of Veterinary Medicine, North Carolina State University, 4700 Hillsborough St, Raleigh, NC, USA.
(Email: alrivas@ncsu.edu or alr4@cornell.edu)

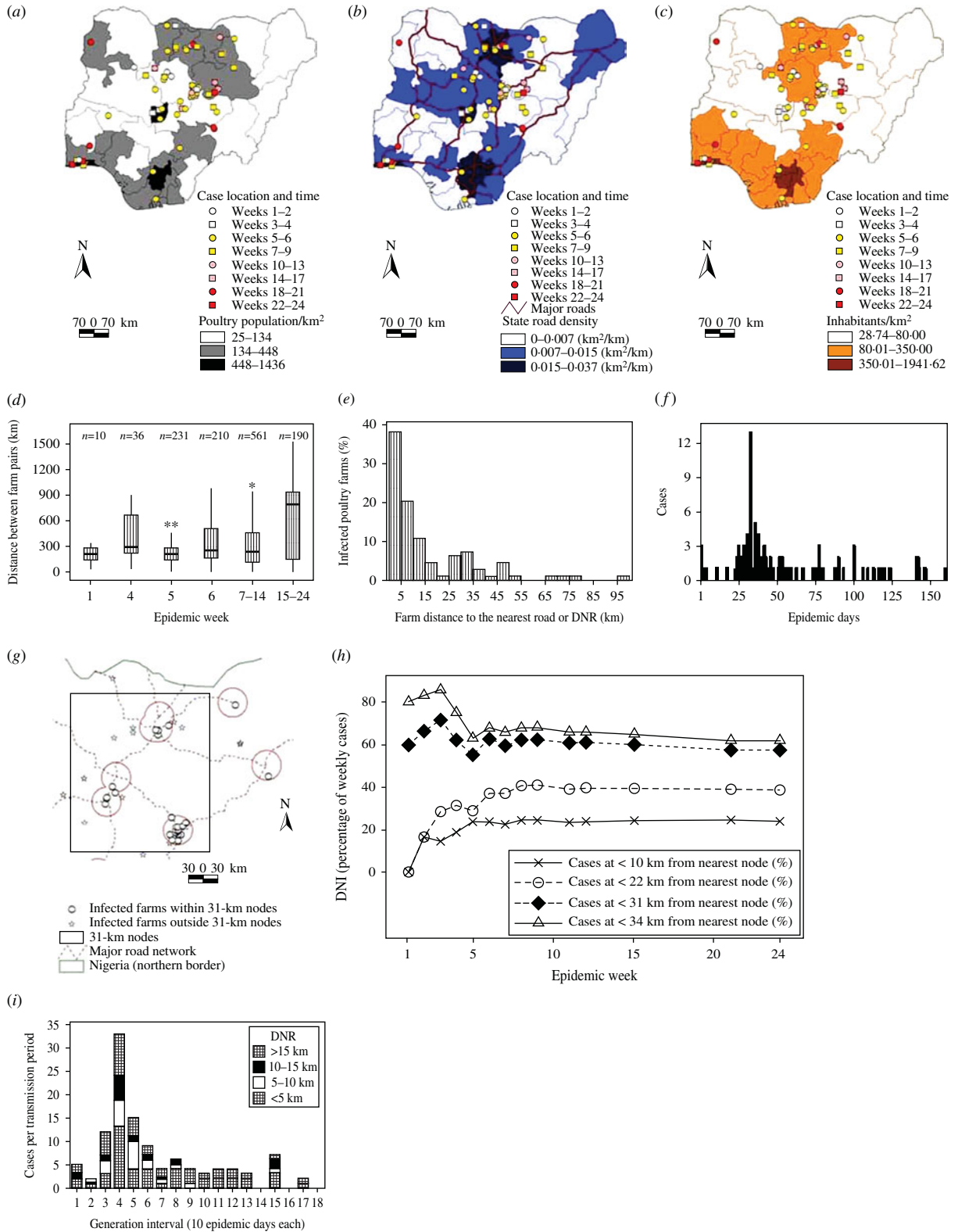


Fig. 1. Spatial-temporal progression of the 2006 Nigerian avian influenza virus (subtype H5N1) epidemic. Spatial location of weekly cases (infected farms) in relation to (state) poultry population density (*a*), road density (*b*), and human population density (*c*). Note that some farms located at a short distance are represented by a single point. (*d*) Median distance between pairs of infected farms. Box plots describe the median distance (central horizontal line) as well as 25th and 75th percentiles (whiskers). Asterisks denote observations above the 75th percentile. Note that no calculations were made at epidemic weeks 2 and 3 because only one case was reported in each of these weeks. (*e*) Percentage of cases ($n = 113$ farms, January–June 2006)

distance from each infected farm to the nearest major road (DNR), major road intersection (DNI), or other infected farms. The bio-temporal component investigated was the generation interval (transmission period) of the pathogenic agent [3], which includes but exceeds the viral replication period. While the replication period (for HPAI, estimated as ~ 2 days) refers to a single host [4], the HPAI generation interval (estimated as ~ 10 days) denotes the time period between the infection reports for two individuals, located in different farms or flocks, when one of these individuals is the primary case and the other is the secondary case [5, 6]. Because the HPAI virus may survive outside the host (e.g. in faeces, water [7]), its generation interval is usually longer than its replication period. The purpose of this study was to utilize the data collected in the 2006 Nigerian HPAI H5N1 epidemic in order to generate hypotheses on transmission factors associated with emerging infections of rapid dissemination.

The unit of study was the group of Nigerian cases (infected poultry farms, expressed as counts, percentages, or case density/km²). Any farm reporting at least one infected animal was defined as a case. Each case was characterized by its (a) latitude and longitude, (b) reporting time, and (c) Nigerian state of affiliation. Additional variables were (d) state road density (km of roads per km² of state area), (e) state human population density (inhabitants/km²), and (f) state poultry population density (birds/km²), as well as the Euclidean distance (km) from the centroid of each infected farm to (g) the nearest major road (DNR), (h) the nearest major road intersection (DNI), or (i) any other infected farm. Cases were analysed as counts or percentages of (a) all cases (located at specified DNR or DNI), (b) new weekly cases, or (c) in relation to 10-day periods (assumed to reflect the HPAI H5N1 inter-farm generation interval).

Both the case geo-temporal data (113 poultry farms reported as infected by HPAI H5N1 between January and June, 2006) and the human and poultry population data of all Nigerian states were provided by the National Veterinary Research Institute of Nigeria [8, 9]. The Nigerian major road network map was

derived from country situational base maps produced by the World Health Organization [10]. The estimated HPAI generation interval considered inter-farm, not intra-farm, infections [5, 6].

Farm-related distances were calculated using Geographical Information Systems (GIS, ArcView GIS 3.3 and ArcGIS Desktop 9.0, both from ESRI, USA). To generate the DNR, the GIS `NEAR` command identified the road segment nearest to each infected farm, the latitude and longitude values of the nearest point on this road segment, and the distance to the nearest point. These attributes were added to the infected farm *layer*. To calculate the DNI, a GIS *point layer* of all road intersections was created. Using the `POINTDISTANCE` command, a table was generated which contained (a) farm identifier, (b) nearest intersection identifier, and (c) distance. The same procedure calculated the distance between every pair of infected farms at a particular time unit; for instance, if 10 farms reported infections at a certain week, the median inter-farm distance at that time was that of 45 farm pairs ($10 \times 9/2$). To empirically determine (a) whether disease clustering occurred, (b) if so, whether clustering was associated with major road intersections, and (c) if so, the critical radius of clusters (the smallest radius of circles that, earlier and over time, contained the highest percentage of weekly cases), using intersection identifiers, circles of various (arbitrarily chosen) radii were created, which were centred at road intersections ('nodes'). Counts or percentages of weekly cases falling inside nodes were compared to the values found outside nodes. Density layers (road, human population, and poultry population) were produced by *intersecting* the variable of interest (e.g. 'road network') with the appropriate spatial scale (e.g. 'states').

Relationships between (state) density-related variables were explored with regression analysis. Median weekly inter-farm distances among infected farms were calculated with the Mann-Whitney test. Data were processed with Minitab 15 (Minitab Inc., USA).

Relationships between (state) case density/km² and other (state) variables were explored (Fig. 1*a-c*). Neither poultry density nor human population

vs. distance to the nearest road (DNR). (f) Number of cases per epidemic day. (g) Cases within and outside 31-km radius circles (nodes), centred at major road intersections. (h) Percentage of new weekly cases *vs.* distance to the nearest road intersection (DNI), reported within 10-, 22-, 31-, and 34-km nodes. The smallest radius yielding the highest percentage of new cases earlier and for longer was ≤ 31 km DNI. The difference in percentage of cases found between ≤ 21 and ≤ 31 km DNI was substantial (\geq twofold at ≤ 31 km DNI than at ≤ 21 km DNI); in contrast, the additional percentage of cases found at ≤ 34 km than at ≤ 31 km DNI was negligible. (i) Count-based, composite (epidemic-temporal-spatial) number of cases per generation interval (10 days each) and per DNR class.

Table 1. Case (infected farm) Euclidean distance to the nearest road (DNR) by epidemic day ($n=113$ farms)

Case DNR (km)	Epidemic day	Case DNR (km)	Epidemic day	Case DNR (km)	Epidemic day	Case DNR (km)	Epidemic day	Case DNR (km)	Epidemic day
4.23	1	78.38	32	30.78	37	8.30	52	0.23	94
29.32	1	48.49	33	1.07	38	1.83	53	4.23	101
1.59	1	14.30	33	0.81	38	19.30	54	1.41	101
13.12	3	5.47	33	3.23	38	12.93	54	30.78	101
70.26	4	1.96	33	2.37	38	9.98	55	27.10	107
2.37	11	47.13	33	48.57	39	0.46	57	0.73	110
8.47	18	12.48	33	8.47	40	42.75	59	0.73	114
19.34	23	12.48	33	0.81	40	15.57	60	36.35	117
10.60	24	0.81	33	7.94	41	8.47	63	27.10	119
28.68	24	8.30	33	5.58	41	0.23	67	0.73	121
9.98	25	46.46	33	1.47	42	36.35	68	9.12	124
2.61	25	25.38	33	8.47	42	4.23	74	50.79	125
0.81	25	2.37	33	0.81	42	0.12	76	11.75	142
5.44	27	8.37	33	17.98	43	11.41	78	0.05	142
31.80	27	22.56	34	28.02	43	2.60	78	6.59	143
33.83	28	67.37	36	8.47	44	1.10	78	15.90	143
8.62	29	9.98	36	28.68	45	8.37	79	2.29	144
0.55	29	2.32	36	95.62	46	30.78	80	0.05	147
33.83	29	8.47	36	9.98	46	36.35	81	11.75	148
12.67	31	0.811	36	2.37	47	8.47	88	45.27	161
4.23	31	2.94	37	12.67	48	33.83	89	0.09	162
5.84	31	1.33	37	0.73	51	0.73	91		
2.61	31	11.92	37	1.83	52	33.83	91		

density predicted case density. While road density approached significance as a predictor of case density ($P=0.06$), its validity was questionable because results were highly influenced by the data from one state, likely to be an outlier (not shown).

The median Euclidean distance between pairs of newly infected farms per week was higher at the end of the period under study (weeks 15–24, median: 785.1 km, $n=190$ farm pairs) than before week 15 (median: 233.4 km, $n=1048$ farm pairs, Fig. 1*d*), a statistically significant difference as determined by the Mann–Whitney test ($P<0.001$). However, because in emerging infections epidemic data cannot be considered to be independent random samples, the biological significance of this calculation is unknown.

The data collected on the infected farm distance to the nearest road (DNR) supported the analysis of four spatial classes: <5 km, 5 to <10 km, 10–15 km, and >15 km DNR. Farms at a <5 km DNR accounted for 38% of all cases (43/113, Fig. 1*e*, Table 1). When the number of cases per epidemic day was plotted, several gaps before the epidemic peak (day 33) indicated that even in the period of greater case growth, cases were not always reported on a daily basis (Fig. 1*f*).

The median DNI was 24.2 km in the first week. Farms located at ≥ 100 km DNI only became infected after epidemic week 4. Case clustering was observed: 57% of all cases (65/113) were within 31 km of three road intersections (Fig. 1*g*). From the second epidemic week onwards, cases at ≤ 10 km DNI accounted for a substantial percentage of new weekly cases ($\sim 20\%$, Fig. 1*h*).

When, instead of days or weeks, epidemic data were analysed with a double descriptor [number of cases per estimated HPAI generation interval, per spatial (DNR) class], an exponential growth phase was observed between the second and fourth generation intervals with cases reported in each of the generation intervals until the fourteenth generation interval of the epidemic (Fig. 1*i*). Two epidemic phases were observed: (1) the one comprising first generation interval, and (2) the phase that included the remaining intervals (Fig. 1*i*). While both <10 km DNR and >15 km DNR cases were observed in the first generation interval, <10 km DNR cases predominated in the second interval. Cases located at <5 km DNR represented $>20\%$ of all cases in each of the first eight generation intervals (Fig. 1*i*). By epidemic day 4

(generation interval I), cases at <5 km DNR represented 40% of all infected farms.

These data support the hypothesis that the Nigerian major highway network promoted epidemic spread. In the first half of 2006, HPAI H5N1 cases could be categorized into two major spatial classes: (1) the predominant class, which included cases close to roads, intersections, or other infected farms, and (2) a secondary class, including cases located at >15 km DNR, >31 km DNI, or long inter-farm distances. Because not all road intersections were equally associated with cases and >20% of cases reported at any generation interval were at <5 km DNR, both a non-random (clustered) case distribution hypothesis and Pareto or power law distributions (the '20:80 rule') were supported by the data [11]. The '20:80' rule refers to a high percentage of epidemic size (e.g. ~80%) associated with a few (e.g. ~20%) highly influential cases.

While the early cases were reported in the centre and in the periphery of Nigeria (near Lagos), only those near three major highway intersections (those in the centre of the country) were predominantly associated with viral dispersal. The data did not appear to support the hypothesis that migratory birds or wind might have disseminated H5N1 within Nigeria. In contrast, anthropogenic-mediated dispersal was likely. Possible behaviours that may spread HPAI H5N1 include subsistence agricultural practices, such as live bird markets [12] and early re-population of infected premises with susceptible birds.

These hypotheses are unlikely to be contradicted by non-reporting or delayed reporting. Because (1) the unit of the outcome was not the number of infected birds but the number of farms reporting infections, and (2) HPAI is associated with high mortality [8, 9], the magnitude of non-reporting in this epidemic could not be as high as that of a subclinical disease or an epidemic where the unit of the measured outcome is the individual animal.

The use of temporal units expressed as (inter-farm) HPAI generation intervals appeared to prevent false-negative results. If expressed in days, the epidemic might have seemed to cease when no new cases were reported in three consecutive days – the time equivalent to a replication period, as observed several times before the epidemic peak (Fig. 1*f*). In contrast, when measured as generation intervals, no gaps were observed before the epidemic peak (Fig. 1*i*). Caution is warranted in relation to the generation interval considered here: the estimates used were based on studies

conducted in other countries where other HPAI subtypes (not H5N1) were isolated [5, 6].

Within three epidemic weeks, the Nigerian scenario showed infected farms separated by Euclidian distances >900 km. Later, pairs of infected farms were up to 1500 km apart. This situation did not support the hypothesis that post-outbreak policies limited to the susceptible population (e.g. vaccination, de-population, quarantine [3]) would be effective. Instead, this epidemic dataset provided an opportunity to revise two aspects of control policy traditions: (1) the time when control measures are chosen as well as the focus of such measures, and (2) the feasibility and information value of assessments that compare infected and susceptible populations.

In emerging infectious diseases, decisions should be made as early as possible [13]. Even if an early decision to use an inexpensive control measure is less certain than a decision that requires waiting for definitive data, the early intervention may produce much larger benefits and may therefore be the better choice. However, control measures have classically been based on the ratio between the number of secondary cases and those generated in the primary generation interval [3, 4]. The calculation of this ratio requires a waiting time of at least two generation intervals (~20 days in this scenario). At such a late time in the epidemic progression, measures focusing on susceptible hosts necessarily become costly, complex, involve a large geographical scale, and may require a lengthy period to induce effects. If, in addition to control policies that focus on the susceptible host, early dissemination factors are emphasized, more beneficial results might be achieved.

In emerging infectious diseases, all cases, except primary cases, are dependent, i.e. they are generated from some of the primary cases [3, 14]. In this scenario, not all secondary cases showed the spatial features revealed by all primary cases. Only <10 km DNR primary cases seemed to trigger the epidemic spread: they predominated in 12/13 consecutive generation cycles (Fig. 1*i*). Epidemic cases do not have identical weight either: secondary cases generate more cases than tertiary ones, tertiary cases create more cases than quaternary ones, and the last cases – if the epidemic stops – produce none.

Because comparisons between infected and susceptible populations are sensitive to statistical power issues and emerging diseases are typically characterized by only a very few cases at the onset, such

comparisons may be unfeasible in scenarios like the one described. Even if feasible, comparisons between infected and non-infected individuals may be inconsequential: if the putative dispersal mechanism associated with cases differs in magnitude in relation to susceptible individuals, e.g. if the DNR of cases measured at an early epidemic time is shorter than the DNR of susceptible individuals, that difference is not evidence that later changes (in any direction) may occur. If, instead, no difference in DNR is found between infected and susceptible individuals, such a finding would support an urgent intervention: it would suggest that the mechanism already shown to promote epidemic spread could soon reach individuals still not affected. These considerations, together with the focus on early decision-making, support the view that comparisons between infected and susceptible populations may not apply to the early phase of emerging infections. Instead, in early phases of infectious epidemics, the priority, as John Snow showed a century and a half ago, may be the identification of plausible transmission factors associated with cases [15].

A simple assessment of the association between the percentage of cases and a transmission factor, if conducted during the first generation interval of the invading microbe, may improve decision-making. In the situation under analysis, such a decision could have been made at epidemic day 4 – one fifth of the time required by classical models, i.e. two generation intervals or 20 epidemic days. Because the percentage of cases located a short distance from the nearest road was known to be $\geq 20\%$ before the first generation interval concluded, a dispersal mechanism could have been postulated: epidemic spread mediated by roads. This data-driven proposition could have supported the implementation of control measures at specific spatial points, such as road blocks that prevent poultry trade.

If the actual spread had been mediated by other means, the cost of such decision would have been negligible. However, if correct, this decision could have stopped the epidemic spread because no secondary cases could occur. If (a) the percentage of primary cases associated with a factor likely to act as a dispersal mechanism is $> 20\%$ [11], (b) the spatial structure of this factor is known and measurable (such as the location of major roads), and (c) the spatial location of infected sites is also known (if the disease is clinically observable, such as HPAI H5N1), then (d) early decisions can be produced.

In emerging infections, if rapid dissemination occurs, a pre-existing contact network may be suspected. However, connectivity alone may not suffice to explain epidemic dissemination. If connectivity is too high (if space is completely occupied by a network of highways or rivers), increased connectivity can only occur at the expense of population density. Hence, the interaction that involves connectivity, infected and susceptible subpopulations is unlikely to be linear: while epidemic diffusion requires a minimum of connectivity, maximal epidemic diffusion may decrease, if not cease, when increases in connectivity result in decreases of population density (e.g. a road density so high that no land is available for housing or farms). Hence, a practical application of these lessons is the anticipatory generation of matrices that include spatially explicit data on 100% of the susceptible population. Such data may identify the specific network nodes, e.g. highway intersections, that reveal not only high connectivity values (e.g. shorter distances between pairs of nodes), but also high demographic values (e.g. high farm density). When such data are available before emerging epidemics occur, then they may help to allocate resources earlier, at the nodes suspected to be critical. The effectiveness of control measures meant to disrupt pre-existing networks can only be effective if implemented earlier, not later. Such early decisions could also increase the effectiveness of later measures, reducing the cost or coverage of mass vaccination, isolation, and other measures focusing on the host.

These considerations may be of interest for policy-makers of countries not yet affected by HPAI H5N1 as well as those of countries susceptible to emerging infectious diseases. Information on variables such as farm location, inter-farm distance, and the DNR and DNI of each farm, if available before an emerging infection occurs, could inform early implementation of control policies.

ACKNOWLEDGEMENTS

We are grateful for the assistance of the Executive Director and the staff of the National Veterinary Research Institute of Nigeria, Mrs. Celia Abolnik (ARC-OVI, South Africa), and the support of the Centro de Investigaciones Avanzadas (CINVESTAV, Mérida, YUC, México) and the Center for Non-Linear Studies (Los Alamos National Laboratory, Los Alamos, NM, USA).

DECLARATION OF INTEREST

None.

REFERENCES

1. **Cook RL, et al.** What's driving an epidemic? The spread of syphilis along an interstate highway in rural North Carolina. *American Journal of Public Health* 1999; **89**: 369–373.
2. **Rivas AL, et al.** Identification of geographical factors associated with early epidemic spread of foot-and-mouth disease. *American Journal of Veterinary Research* 2003; **64**: 1519–1527.
3. **Grassly NC, Fraser C.** Mathematical models of infectious disease transmission. *Nature Reviews Microbiology* 2008; **6**: 477–487.
4. **Das A, et al.** Detection of H5N1 high-pathogenicity avian influenza virus in meat and tracheal samples from experimentally infected chickens. *Avian Diseases* 2008; **52**: 40–48.
5. **Stegeman A, et al.** Avian influenza A virus (H7N7) epidemic in The Netherlands in 2003: course of the epidemic and effectiveness of control measures. *Journal of Infectious Diseases* 2004; **190**: 2088–2095.
6. **LeMenach, et al.** Key strategies for reducing spread of avian influenza among commercial poultry holdings: lessons for transmission to humans. *Proceedings of the Royal Society of London Series B: Biological Sciences* 2006; **273**: 2467–2475.
7. **Brown JD, et al.** Persistence of H5 and H7 avian influenza viruses in water. *Avian Diseases* 2007; **51**: 285–289.
8. **Fasina FO, et al.** Molecular characterization and epidemiology of the highly pathogenic avian influenza H5N1 in Nigeria. *Epidemiology and Infection* 2009; **137**: 456–463.
9. **Adene DF, Oguntade AE.** The structure and importance of the commercial and village based poultry industry in Nigeria (http://www.fao.org/docs/eims/upload//214281/poultrysector_nga_en.pdf). Accessed 1 May 2007.
10. **World Health Organization.** ([http://www.reliefweb.int/rw/fullMaps_Af.nsf/luFullMap/B6887A6EB2E41E1FC125723400298710/\\$File/who_REF_nga060208.pdf?OpenElement](http://www.reliefweb.int/rw/fullMaps_Af.nsf/luFullMap/B6887A6EB2E41E1FC125723400298710/$File/who_REF_nga060208.pdf?OpenElement)). Accessed 15 July 2007.
11. **Woolhouse MEJ, et al.** Heterogeneities in the transmission of infectious agents: implications for the design of control programs. *Proceedings of the National Academy of Sciences USA* 1997; **94**: 338–342.
12. **Lau EHY, et al.** Effect of interventions on Influenza A (H9N2) isolation in Hong Kong's live poultry markets, 1999–2005. *Emerging Infectious Diseases* 2007; **13**: 1340–1347.
13. **Phillips CV, Goodman KJ.** The missed lessons of Sir Austin Bradford Hill. *Epidemiological Perspectives and Innovations* 2004; **1**: 3.
14. **Koopman J.** Modeling infection transmission. *Annual Review of Public Health* 2004; **25**: 303–326.
15. **Snow SJ.** Commentary: Sutherland, Snow and water: the transmission of cholera in the nineteenth century. *International Journal of Epidemiology* 2002; **31**: 908–911.