Risk Factors for Cluster Outbreaks of Avian Influenza A H5N1 Infection, Indonesia

Tjandra Y. Aditama,1,a Gina Samaan,2,a Rita Kusriastuti,1 Wilfried H. Purba,1 Misriyah,1 Hari Santoso,1 Arie Bratasena,1 Anas Maruf,1 Elvieda Sariwati,1 Vivi Setiawaty,3 Alex R. Cook,4 Mark S. Clements,2,5 Kamalini Lokuge,2 Paul M. Kelly,2,6 and I. Nyoman Kandun1

1Directorate-General Disease Control and Environmental Health, Ministry of Health, Jakarta, Indonesia; 2National Centre for Epidemiology and Population Health, College of Medicine, Biology, and Environment, The Australian National University, Canberra; 3National Institute of Health Research and Development, Ministry of Health, Jakarta, Indonesia; 4Department of Statistics and Applied Probability, National University Singapore; 5Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden; and 6Population Health Division, Australian Capital Territory Government Health Directorate, Canberra

Background. By 30 July 2009, Indonesia had reported 139 outbreaks of avian influenza (AI) H5N1 infection in humans. Risk factors for case clustering remain largely unknown. This study assesses risk factors for cluster outbreaks and for secondary case infection.

Methods. The 113 sporadic and 26 cluster outbreaks were compared on household and individual level variables. Variables assessed include those never reported previously, including household size and genealogical relationships between cases and their contacts.

Results. Cluster outbreaks had larger households and more blood-related contacts, especially first-degree relatives, compared with sporadic case outbreaks. Risk factors for cluster outbreaks were the number of first-degree blood-relatives to the index case (adjusted odds ratio [aOR], 1.50; 95% confidence interval [CI]: 1.20–1.86) and index cases having direct exposure to sources of AI H5N1 virus (aOR, 3.20; 95% CI: 1.15–8.90). Risk factors for secondary case infection were being aged between 5 and 17 years (aOR, 8.32; 95% CI: 1.72–40.25), or 18 and 30 years (aOR, 6.04; 95% CI: 1.21–30.08), having direct exposure to sources of AI H5N1 virus (aOR, 3.48; 95% CI: 1.28–9.46), and being a first-degree relative to an index case (aOR, 11.0; 95% CI: 1.43–84.66). Siblings to index cases were 5 times more likely to become secondary cases (OR, 4.72; 95% CI: 1.67–13.35).

Conclusions. The type of exposure and the genealogical relationship between index cases and their contacts impacts the risk of clustering. The study adds evidence that AI H5N1 infection is influenced by, and may even depend on, host genetic susceptibility.
confirmed cases were identified from 139 households [5]. The Indonesian Ministry of Health investigates all human cases of AI H5N1 to determine the source of illness and disease exposure [4]. This is to our knowledge the first study globally to compare outbreaks involving a single case with those involving ≥1 case to identify risk factors for cluster outbreaks. There are 2 tiers of analyses reported in the study: (1) risk factors for cluster outbreaks and (2) risk factors for secondary case infection. Both tiers explore household and individual level variables including those never to our knowledge reported previously, such as household size and genealogical relationships between cases and their contacts.

METHODS

Definitions
The Ministry of Health AI H5N1 case database and detailed case investigation forms were reviewed and analyzed for outbreaks detected between July 2005 and July 2009. The study conformed with the World Health Organization (WHO) case definitions of probable and confirmed categories of human AI H5N1 infection [6], and definitions of cluster and sporadic outbreaks [7]. A cluster is a group composed of ≥1 confirmed cases of H5N1 virus infection and additional confirmed or probable cases associated with a specific setting, with the onset of cases occurring within 2 weeks of each other. A sporadic outbreak was defined as 1 confirmed case of H5N1 virus infection. For both sporadic and cluster outbreaks, a household contact was a person who had at least 4 hours contact with a case at home within the 7 days prior or 14 days after a case’s onset of illness.

Other data used include household setting, for which cities or towns were defined as urban, fringes of cities as semiurban, and villages as rural. Seasonal outbreaks were defined as outbreaks occurring in the wet season from December to March versus those occurring in the dry season (April until November). Three categories were used to describe disease exposure: direct, indirect and other: “Direct” exposure referred to cases who handled sick or dead poultry, handled poultry products such as fertilizers, or who had poultry deaths in the home; “indirect” exposure to cases where poultry deaths were reported in the case’s neighborhood, cases where healthy poultry were present in the neighborhood, and cases who visited live bird markets; and “other” to cases whose exposure was inconclusive despite investigation or who apparently were only exposed to a prior case but not to direct or indirect exposure types.

Household size was the number of people in the household including all cases. Household size was analyzed as both a continuous and categorical variable. Contacts were classified in their genealogical relationship to the index case. First-degree relatives comprised parents, offspring, and siblings; second-degree relatives were aunts, uncles, grandchildren, grandparents, nephews, and nieces; and third-degree relatives were cousins. Non–blood relations to the index case comprised spouses, family-in-law, and household help.

Data Collection
Field investigation teams were deployed to investigate and instigate disease control measures for every outbreak [4, 8–10]. District level teams were deployed on the same day of outbreak detection to initiate the investigation. A provincial/national team rapidly followed to systematically collect data and to cross-check the district level team findings and activities. Most provincial/national teams also comprised a WHO epidemiologist to support data collection.

Teams interviewed cases when possible (because many cases died before investigation teams arrived), family members (especially those who could report on the case), and key informants (including village leaders). Data collection forms, developed based on WHO guidelines [7, 11, 12], were used to obtain demographic, clinical, and epidemiological data. Household contacts were traced, and healthcare workers from the nearest government primary healthcare center were instructed to visit the household daily for 2 weeks to monitor and detect any additional cases in the household. The definition of household contact and their monitoring for 2 weeks was uniformly and systematically applied in both sporadic and cluster outbreaks.

Statistical Methods and Ethics
We used logistic regression to assess risk factors for cluster outbreaks, starting with univariate analyses and subsequently constructing multivariate models using variables significant at α = .1 in univariate analysis and sequentially discarding terms not significant at α = .05 starting with the one with the highest P value. We used the le Cessie-van Houwelingen-Copas-Hosmer unweighted sum of squares goodness-of-fit test to assess model validity, as advocated by Hosmer et al [13–15]. Independent predictors of cluster outbreaks were explored further using descriptive statistics (Table 1). We used Wilson score interval method to derive confidence intervals for proportions [16].

To assess risk factors for secondary case infection, logistic regression was used with adjustments for clustering (Tables 2 and 3) by computing a cluster robust standard error for the coefficient. Stata software, version 10-0 (StataCorp) and the R statistical environment were used for the descriptive and statistical analyses [17]. This study was part of an ongoing public-health investigation and is therefore exempt from review by an institutional review board.

RESULTS
In the 4-year study period, 139 outbreaks of human AI H5N1 infection were detected, of which 113 were sporadic case outbreaks and 26, clusters. There were 177 cases (159 laboratory-confirmed and 18 probable). For the 113 sporadic case outbreaks, only 1 case
was detected in each of those outbreaks despite investigation and monitoring of their household contacts. In the 26 cluster outbreaks, there were 64 confirmed and probable cases, where the average cluster size was 2.5 (median, 2; range, 2–8). Only 1 cluster had >4 cases—a cluster from North Sumatra province that had 7 confirmed and 1 probable case [18]. Case fatality rate was 85%. A map of cases and outbreak type by province can be seen in Figure 1.

### Table 1. Main Putative Exposure for Sporadic Cases (n = 113), Cluster Index Cases (n = 29), and Cluster Secondary Cases (n = 35)

<table>
<thead>
<tr>
<th>Exposure type</th>
<th>Exposure</th>
<th>Sporadic case (%)</th>
<th>Cluster index case (%)</th>
<th>Cluster secondary cases (%)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct</td>
<td>Bird deaths or H5N1 confirmed in birds in the home</td>
<td>13 (12)</td>
<td>9 (31)</td>
<td>9 (26)</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>Handled bird products</td>
<td>7 (6)</td>
<td>1 (3)</td>
<td>2 (6)</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>Handled sick/dead birds</td>
<td>25 (22)</td>
<td>9 (31)</td>
<td>6 (17)</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>Subtotal</td>
<td>45 (40)</td>
<td>19 (66)</td>
<td>17 (49)</td>
<td>81</td>
</tr>
<tr>
<td>Indirect</td>
<td>Healthy birds in neighborhood</td>
<td>18 (16)</td>
<td>1 (3)</td>
<td>2 (6)</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>Bird deaths in neighborhood</td>
<td>33 (29)</td>
<td>6 (21)</td>
<td>5 (14)</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>Visited live bird market</td>
<td>7 (6)</td>
<td>0*</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Sub-total</td>
<td>58 (51)</td>
<td>7 (24)</td>
<td>7 (20)</td>
<td>72</td>
</tr>
<tr>
<td>Other</td>
<td>Inconclusive but exposed to prior case</td>
<td>...</td>
<td>...</td>
<td>11 (31)</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Inconclusive despite investigation</td>
<td>10 (9)</td>
<td>3 (10)</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Subtotal</td>
<td>10 (9)</td>
<td>3 (10)</td>
<td>11 (31)</td>
<td>24</td>
</tr>
<tr>
<td>Totala</td>
<td></td>
<td>113</td>
<td>29</td>
<td>35</td>
<td>177</td>
</tr>
</tbody>
</table>

* Even though 1 cluster index case had exposure to live bird markets, the case’s main exposure was classified as handling sick/dead birds.

**Risk Factors for Cluster Outbreaks**

Thirteen variables at household and individual level were explored as potential risk factors for cluster outbreaks. Four variables had P values < .1 on univariate analyses: household size, number of blood contacts to the index case, the index case’s main exposure type, and the degree of blood relation to the index case (Figure 2). The risk factors and the multivariate model are presented below.

![Figure 1. Confirmed and probable cases of avian influenza H5N1 infection in Indonesia, by province and outbreak type, July 2005–July 2009.](cid:CID 2011:53 (15 December) 1239)
Household-Level Risk Factors

Information about household contacts was available for 80 of the 139 outbreaks (60 sporadic and 20 cluster outbreaks). For these 80 outbreaks, 607 individuals were investigated, of whom 111 developed illness (82 index cases and 29 secondary cases) and 496 remained healthy. Since 1 cluster outbreak only had coindex cases but no secondary cases, the 29 secondary cases came from 19 cluster outbreaks. The overall attack rate in the 80 outbreaks was 18%. Thirty-three households had 6–10 persons (41%), 32 had 1–5 persons (40%), 12 had 11–15 persons (15%), and 3 had >15 persons (4%). The mean size of households was 6 persons for sporadic outbreaks and 9 for cluster outbreaks. Each additional household member increased the odds for developing a cluster by 20% (odds ratio [OR], 1.20; 95% confidence interval [CI]: 1.09–1.33, \( P < .001 \)). The increased risk was marginally stronger for each additional blood relative contact (OR 1.25; 95% CI: 1.05–1.48, \( P = .012 \)). The majority of household contacts for both sporadic and cluster outbreak types were first-degree relatives (Figure 2), but this proportion was significantly higher in cluster outbreaks (\( P = .008 \)) and the risk of an infection developing into a cluster increased markedly with the number of first-degree relatives (OR, 1.51; 95% CI: 1.24–1.84, \( P < .001 \)). In contrast, these outbreaks provide no evidence that the risk increased with more second-degree (\( P = .27 \)), or third-degree (\( P = .08 \)) relatives, nor for unrelated cohabitants (\( P = .69 \)), suggesting that the increased risk attributable to large household sizes is likely due to the increased number of first-degree blood relatives such as siblings, parents, and offspring.

Among households, 28% were urban, 36% semiurban, and 36% rural. Both sporadic and cluster outbreaks occurred in all 3 settings (Figure 2). There was no evidence that other household level variables such as household location, time of year, or mean age of cohabitants were risk factors for cluster outbreaks (Figure 2).

### Table 3. Genealogical Relationships Associated With Secondary Cases, Comparing 35 Secondary Cases and 348 Healthy Contacts

<table>
<thead>
<tr>
<th>Relation</th>
<th>Secondary Cases (%)</th>
<th>Healthy Contacts (%)</th>
<th>OR (P value)</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sibling</td>
<td>22 (62)</td>
<td>106 (21)</td>
<td>4.72 (.003)</td>
<td>1.67–13.35</td>
</tr>
<tr>
<td>Other first degree</td>
<td>9 (26)</td>
<td>151 (30)</td>
<td>1.36 (.56)</td>
<td>.49–3.79</td>
</tr>
<tr>
<td>Father</td>
<td>1 (3)</td>
<td>47 (10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mother</td>
<td>3 (9)</td>
<td>55 (11)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Offspring</td>
<td>5 (14)</td>
<td>49 (10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Second and third degree</td>
<td>4 (11)</td>
<td>91 (18)</td>
<td>Reference group</td>
<td></td>
</tr>
<tr>
<td>Grandchild</td>
<td>0</td>
<td>2 (0.4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grandparent</td>
<td>0</td>
<td>15 (3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other (aunt/uncles, cousins, nephews/nieces)</td>
<td>4 (11)</td>
<td>74 (15)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: OR, odds ratio; CI, confidence interval.
Individual Level Risk Factors
There were 113 sporadic outbreak cases and 26 cluster outbreaks with 29 index cases (3 cluster outbreaks had 2 index cases with the same illness onset date). Similar proportions of sporadic cases and cluster index cases worked in bird-related occupations (12%) were likely to be infected in the home (51% sporadic and 68% cluster index) and had timely hospitalization (7% sporadic and 11% cluster index). These variables, along with age and sex, were not associated with sporadic or cluster outbreaks (Figure 2).

Index case exposure was significantly associated with outbreak type, where index cases with direct exposure to sources of AI H5N1 virus were more likely to lead to clusters (OR, 3.50; 95% CI: 1.35–9.10, P = .01, Figure 2). Table 1 presents the main

Figure 2. Risk factors for cluster outbreaks of avian influenza A H5N1 infection comparing 113 sporadic and 26 cluster outbreaks. Bars in the left column are empirical proportions of outbreaks that are clusters. White rectangles in the right column are data from sporadic outbreaks and black rectangles represent cluster outbreaks. The relationship between relatedness and cluster formation is presented by the proportion of all household contacts by degree of relatedness for cluster and sporadic outbreaks (right, bottom). Variables significant at the 5% level are indicated with P values next to the title. aFor relatedness to case, P value presented for the number of first-degree relatives. For second- or third-degree relatives or unrelated cohabitants, the P values were .27, .08, and .69, respectively. The 13 index cases whose exposure could not be determined despite investigation are omitted from the exposure of index panel. Analyses for index case level variables included 29 index cases for clusters (as 3 clusters had coindex cases).
putative exposures for cases in sporadic and cluster outbreaks. Compared to sporadic cases, a greater proportion of cluster index cases had bird deaths in the home (31% vs 12%) or handled sick/dead birds (31% vs 22%). A greater proportion of sporadic cases had indirect exposures as their main putative exposure, where a greater proportion was exposed to poultry deaths in their neighborhood (29% vs 21%), had healthy poultry in their neighborhood (16% vs 3%), or visited a live bird market (6% vs 0).

**Multivariate Model**
Variables significant at $z = .1$ on univariate analysis—exposure of index, household size, number of blood contacts, and number of first-degree relatives—were considered for a multivariate logistic regression model. Because risk associated with household size and relatedness were attributable to the number of first-degree relatives, and given the correlation between these 3 factors, we dropped household size and number of blood contacts. Both remaining variables were statistically significant as independent risk factors for cluster outbreaks: the number of first-degree relatives of the index case(s) (adjusted odds ratio [aOR], 1.50 per first-degree relative; 95% CI: 1.20–1.86, $P < .001$) and index case direct exposure to sources of AI H5N1 virus (aOR, 3.20; 95% CI: 1.15–8.90, $P = .026$). The final multivariate model passed the le Cessie-van Houwelingen-Copas-Hosmer unweighted sum of squares goodness-of-fit test ($P = .62$). No other household or individual level variables were associated with sporadic outbreaks (Figure 2 and Table 1).

**Risk Factors for Secondary Case Infection**
Three variables were found significant at $z = .1$ on univariate analyses as risk factors for secondary case infection (Table 2): age, exposure, and first-degree blood-relatives. There was no statistical difference in sex between secondary cases and healthy contacts.

Contacts under 30 years of age were at greater risk of becoming secondary cases (Table 2). Further analysis found that the age distribution of secondary cases (mean, 17.1; range, 1–39 years) was similar to cluster index cases (mean, 17.3; range, 3–38 years). Overall, the age of all cluster cases (mean, 17.5; range, 1–39 years) did not differ substantially from that of sporadic cases (mean, 20.6; range 2–67, $P = .054$).

Contacts who had direct exposure to sources of AI H5N1 virus were also at greater risk of infection than contacts who had indirect exposure ($P = .003$; Table 2). Secondary case exposures were similar to cluster index cases, where most had bird deaths in the home (26%) or handled sick/dead birds (17%) (Table 1). Being a first-degree relative to an index case was a risk factor for becoming a secondary case ($P = .001$). The majority of secondary cases (89%) were first-degree relatives (Table 2). No non-blood relatives of index cases were infected in any of the outbreaks.

**Multivariate Model** The 3 variables significant on univariate analyses were considered in the multivariate logistic regression model. The final model had 3 risk factors for secondary case infection (Table 2). These were age where individuals between 5 and 17 years of age (OR, 8.32; 95% CI: 1.72–40.25; $P = .008$) or 18–30 years of age (OR, 6.04, 95% CI: 1.21–30.08, $P = .028$) were more likely to be infected compared with contacts >30 years old, having direct exposure to sources of AI virus (OR, 348; 95% CI: 1.28–9.46, $P = .014$) and being a first-degree relative of an index case (OR, 11.0; 95% CI: 1.43–84.66, $P = .02$). The final multivariate model had good fit ($P = .21$).

The finding that first-degree relatives were at greater risk of secondary infection was explored further. Restricting analysis to only blood-relative contacts of index cases, we found that siblings were nearly five times more likely to become secondary cases compared with second- or third-degree blood-relatives (OR, 4.72; 95% CI: 1.67–13.35, $P = .003$; Table 3). Other first-degree relatives (parents or offspring) were statistically at no greater risk of infection than second- or third-degree relatives. The empirical infection rates for children and adults as a function of genealogical relation to index cases highlights that in addition to relatedness, age is an important determinant of infection (Figure 3).

**DISCUSSION**
Understanding risk factors for outbreaks, especially clustering, have important implications for disease control and prevention [2, 19, 20]. The major conclusion of this study is that an interplay of exposure type and genetic susceptibility predisposes the formation of AI H5N1 cluster outbreaks. Households with many blood-related contacts to the index case were more likely to develop secondary cases, and those who became secondary cases were more likely to be first-degree relatives of the index case. To minimize the risk of clustering, the public health implications are 2-fold: (1) household contacts, especially first-degree relatives, need to be traced and monitored for infection, and (2) household contacts should be educated about appropriate methods for handling birds, especially sick and dead birds, to minimize direct exposure to sources of virus.

Identifying the mechanisms most responsible for household clustering is difficult because genetic relationship and household membership are correlated. Even for diseases for which a genetic mechanism for infection has been identified through whole-genome research, such as for leprosy, the extent to which genetic versus household exposure factors explain clustering has been difficult to determine [21]. For AI H5N1, arguments for genetic susceptibility include the preponderance of familial clustering of cases, with 50 of the 54 clusters detected globally (as of March 2009) having cases that were all genetically linked [18]. Further arguments include the paucity of cases in highly exposed groups such as poultry workers and the occurrence of familial cases...
This may bias the estimates for the risk factors investigated. Second, outbreaks with more cases are more likely to be detected since healthcare workers are more likely to suspect and report these events to public health authorities. This inflates the secondary attack rate and may also bias the estimates for the risk factors investigated. These limitations could be addressed through prospective community-based studies.

A limitation of the study is that 10% of cases (18 of 177) met only the probable case definition [6]. Some of these cases may not have been AI H5N1 infection. However, we consider risk of misclassification to be low, considering the severity associated with illness, the proximate timeline of illness related to a confirmed case in the same household, and similar exposure patterns.

In conclusion, this is the first study to our knowledge to demonstrate that the interaction between household and individual level variables including genealogical relationships impacts the risk of clusters as well as secondary case infection. The study adds further evidence to the hypothesis that there is a genetic basis for susceptibility to infection.

### Notes

**Acknowledgements.** We would like to acknowledge the support of the provincial and district health offices, the Ministry of Agriculture, and laboratory staff in the outbreak investigation and data collection.

**Financial support.** No funding was required for this study as the data were public health surveillance data.

**Potential conflicts of interest.** All authors: No reported conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

### References


