

Suggested citation for this article: Pitzer VE, Olsen SJ, Bergstrom CT, Dowell SF, Lipsitch M. Little evidence for genetic susceptibility to influenza A (H5N1) from family clustering data. *Emerg Infect Dis.* 2007 Jul; [Epub ahead of print]

Little Evidence for Genetic Susceptibility to Influenza A (H5N1) from Family Clustering Data

Virginia E. Pitzer,* Sonja J. Olsen,† Carl T. Bergstrom,‡ Scott F. Dowell,† and Marc Lipsitch*

*Harvard School of Public Health, Boston, Massachusetts, USA; †Centers for Disease Control and Prevention, Atlanta, Georgia, USA; and ‡University of Washington, Seattle, Washington, USA

The apparent clustering of human cases of influenza A (H5N1) among blood relatives has been considered as evidence of genetic variation in susceptibility. We show that, by chance alone, a high proportion of clusters are expected to be limited to blood relatives when infection is a rare event.

Since December, 2003, 36 family clusters among 261 confirmed human cases of influenza A (H5N1) have been documented (1,2). These clusters range in size from 2 to 8 infected persons; in only 4 clusters were 2 unrelated family members (e.g., husband and wife) infected. This pattern has been considered by the World Health Organization as evidence of genetic variation in susceptibility (3–5), but we show this observation provides little grounds for this inference. We describe a null model in which nuclear families experience a common exposure to an avian influenza virus. The observed degree of clustering in blood relatives is consistent with that expected by chance alone in the absence of genetic variation in susceptibility; other features of the data are also consistent with the null model.

Our model assumes all persons are equally susceptible, such that they have the same probability of infection, τ , and ignores possible human-to-human transmission (see Technical Appendix). The number of infected family members follows a binomial distribution with mean $n\tau$, where n is the number of exposed persons in each family. A cluster is defined as a

family in which >1 person is infected; clusters are limited to blood relatives unless both parents are infected.

We compare our model to the observation that 32 of 36 clusters that occurred from December 2003 to December 2006 consisted only of blood relatives ($p_B = 0.89$, 95% confidence interval 0.74–0.97; Table in Technical Appendix). When the probability of infection is low, most clusters consist of 2 infected family members, and by simple combinatorics, these 2 are usually blood relatives (Figure 1).

For a given a nuclear family size, the null model also predicts the proportion of all cases that are part of a cluster and the average number of cases per cluster. Neither of these measures follows a simple distribution; we therefore use simulated data to determine what ranges of our parameters (τ and n) are consistent with the observed degree of clustering both in families and among blood relatives. We estimate the mean and 95% prediction intervals for the proportion of cases occurring in clusters when there are 261 cases, and for the average number of cases per cluster when there are 36 clusters. The expected proportion of cases occurring in clusters is similar to the observed data when the probability of infection is low ($\tau < 0.15$) (Figure 2). The observed average number of cases per cluster, however, is consistent with slightly higher probabilities of infection, larger family sizes, or both (Figure 2).

The discrepancy between the number of cases per cluster and the proportion of cases in clusters may be due to between-family variation in τ . If the probability of infection is low for members of most exposed families and higher for members of a few exposed families, then most cases may come from families in which τ is low, but most of the clusters will occur among families for which τ was higher. This will lead to a lower proportion of cases occurring in clusters and a higher average number of cases per cluster, as is observed. Although it is possible that such variation may be genetic, it could also result from between-household heterogeneity in intensity of exposure to infected birds (or intensity of shedding in birds to which different households are exposed), household hygiene, living conditions, and the like. Human-to-human transmission of the virus could also lead to larger than expected cluster sizes because having ≥ 1 case(s) within a family would increase the risk of subsequent cases occurring, and it could not be ruled out in several clusters (6,7).

Qualitatively, the data suggest the existence of nongenetic, between-household variation in risk. If such nongenetic variation were absent, then in any given village, nearly all pairs of cases occurring among unrelated persons in the same village would be in different

households. Roughly, the chance that a pair of cases in unrelated persons in a village would be from the same household as opposed to different households would be $1/H$, where H is the number of households in a village. With 4 pairs of cases in unrelated persons in the same household, $\approx 4H$ pairs of cases would be expected within a village, mostly in different households for large households. If the average village size of ≈ 138 households estimated for an area of Thailand (8) is typical, then if members of all households in a village were at equal risk, we would expect to see far more pairs of unrelated cases within a village than have actually been observed ($4H \approx 550$ pairs of cases in unrelated persons within a village $\gg 261$ total cases). Clearly, this argument is only heuristic, but when this argument is combined with the likelihood of biologic and behavioral differences between households, it seems likely that τ would vary considerably from 1 household to another.

Furthermore, the model does not account for additional individual variability in susceptibility possibly related to age, level of exposure, or other risk factors. If younger persons have a higher risk for infection or likelihood of exposure, clustering would be promoted, primarily within blood relatives, because siblings would be more likely than either parent to become infected. Approximately half of all cases have occurred in those <20 years of age (9). Similarly, if female persons (for example) were at higher risk for exposure, infection, or both, then clusters including non-blood relatives (e.g., spouses) would tend to include the low-risk sex and thus be less probable. Female persons of ages 10–29 years were slightly overrepresented among laboratory-confirmed case-patients, but the difference was not statistically significant (9).

The null model presented here is not designed to capture all of the heterogeneities in exposure and complexity of real families exposed to influenza subtype H5N1. Rather, it simply illustrates that a large proportion of family clusters limited to blood relatives may occur by chance in the absence of genetic variation in susceptibility, particularly when the probability of infection is low and family sizes are large. Although genetic heterogeneity may possibly contribute to the clustering of avian influenza cases within blood relatives, it is neither a necessary nor the most likely explanation for the data currently available.

This work was supported by US National Institutes of Health grants T32 AI07535 (V.E.P.) and cooperative agreement 5U01GM076497 (Models of Infectious Disease Agent Study [M.L.]).

Ms Pitzer is a Doctor of Science candidate in the Department of Epidemiology at Harvard School of Public Health. Her research interests include epidemiologic methods for emerging infectious diseases.

References

1. Olsen SJ, Ungchusak K, Sovann L, Uyeki TM, Dowell SF, Cox NJ, et al. Family clustering of avian influenza A (H5N1). *Emerg Infect Dis*. 2005;11:1799–801. [Medline](#)
2. World Health Organization. Cumulative number of confirmed human cases of avian influenza A/(H5N1) reported to WHO. 2006. [cited 2007 Apr 12]. Available from http://www.who.int/csr/disease/avian_influenza/country/cases_table_2006_12_27/en/index.html
3. World Health Organization. Influenza research at the human and animal interface. 2006. [cited 2006 Nov 3]. Available from http://www.who.int/csr/resources/publications/influenza/who_cds_epr_gip_2006_3c.pdf
4. Perdue ML, Swayne DE. Public health risk from avian influenza viruses. *Avian Dis*. 2005;49:317–27. [Medline](#)
5. Beigel JH, Farrar J, Han AM, Hayden FG, Hyer R, de Jong MD, et al. Avian influenza A (H5N1) infection in humans. *N Engl J Med*. 2005;353:1374–85. [Medline](#)
6. Ungchusak K, Auewarakul P, Dowell SF, Kitphati R, Auwanit W, Puthavathana P, et al. Probable person-to-person transmission of avian influenza A (H5N1). *N Engl J Med*. 2005;352:333–40. [Medline](#)
7. Butler D. Family tragedy spotlights flu mutations. *Nature*. 2006;442:114–5. [Medline](#)
8. Longini IM Jr, Nizam A, Xu S, Ungchusak K, Hanshaoworakul W, Cummings DA, et al. Containing pandemic influenza at the source. *Science*. 2005;309:1083–7. [Medline](#)
9. Epidemiology of WHO-confirmed human cases of avian influenza A (H5N1) infection. *Wkly Epidemiol Rec*. 2006;81:249–57. [Medline](#)

Address for correspondence: Virginia E. Pitzer, Department of Epidemiology, Harvard School of Public Health, 677 Huntington Ave, Boston, MA 02115, USA; email: vpitzer@hsph.harvard.edu

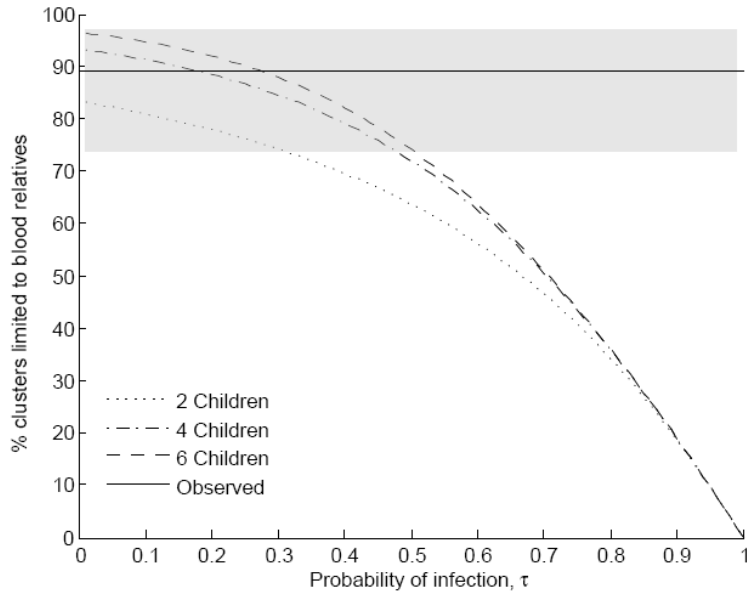


Figure 1. Proportion of clusters limited to blood relatives versus the probability of infection (τ) under the null hypothesis (no variation in susceptibility). Point estimates of the observed data are represented by the solid black line; the shaded region represents the 95% confidence interval.

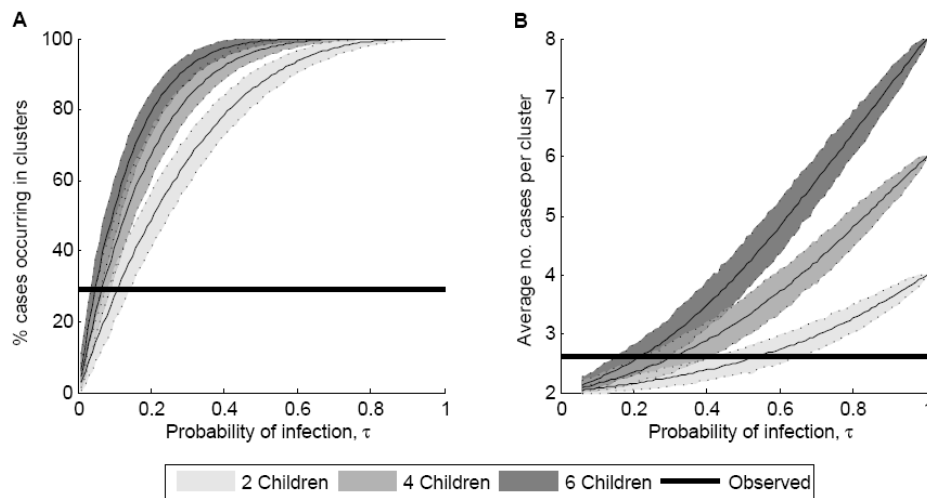


Figure 2. Relationship between data simulated under the null model and the observed pattern of family clustering for A) the proportion of cases occurring in clusters (given 261 total cases) and B) the average number of cases per cluster (given 36 clusters). Estimates of the mean are represented by solid lines; the shaded regions between the dotted lines show 95% prediction intervals for 1,000 simulations. The observed data are represented by the solid black lines.

Appendix: Supplementary Methods

Estimation of the Proportion of Clusters Limited to Blood Relatives

The expected proportion of clusters limited to blood relatives, p_B , for nuclear families (consisting of two parents and a defined number of children who are the offspring of both parents) who experience a common exposure to an avian influenza virus is given by

$$p_B = 1 - \frac{\tau^2}{1 - (1 - \tau)^n - n\tau(1 - \tau)^{n-1}}. \quad (1)$$

Here τ^2 is the probability that both parents are infected (hence the cluster is not restricted to blood relatives) and the denominator is the probability that any two or more persons in the household are infected, i.e., the probability of a household cluster. The model can easily be generalized to more complicated family structures.

Formally, let $p_B = \Pr(BC | C)$ where BC is the event that a family has a cluster of cases only in blood relatives, and C is the event that the family has a cluster of cases, i.e. more than one case. By the law of conditional probability,

$$p_B = \Pr(BC | C) = \frac{\Pr(BC \cap C)}{\Pr(C)} = \frac{\Pr(BC)}{\Pr(C)} = \frac{\Pr(C) - \Pr(NBC)}{\Pr(C)} = 1 - \frac{\Pr(NBC)}{\Pr(C)} \quad (2)$$

where NBC is the event that a family has a cluster including non-blood relatives: in this case, including both parents. Since a family cluster is defined as two or more cases in a family, $\Pr(C) = 1 - \Pr(0 \text{ cases}) - \Pr(1 \text{ case})$. The simple binomial probabilities of having 0 or 1 case in the family are given by: $\Pr(0 \text{ cases}) = (1 - \tau)^n$ and $\Pr(1 \text{ case}) = n\tau(1 - \tau)^{n-1}$. The probability that the two parents will be infected is just τ^2 . Substituting into equation (2) we obtain equation (1).

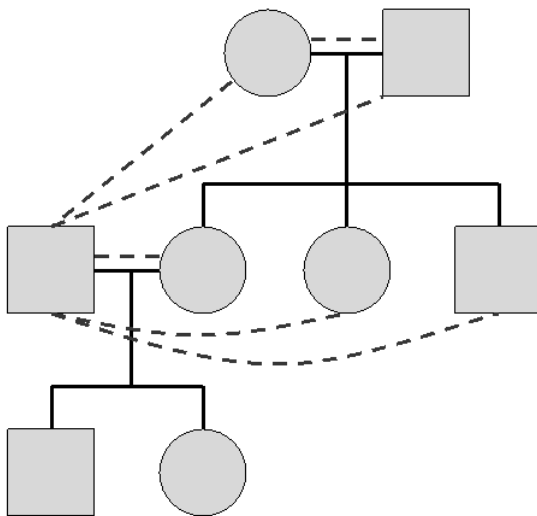
Accounting for Extended Family Structures

While our null model assumes a nuclear family structure, at least 11 out of the 36 clusters described here include members of the extended family (Appendix Table). We can account for such extended family structures by substituting $u\tau^2$, where u is the number of pairs of unrelated family members, for $\Pr(NBC)$ in equation (2) when τ is small, i.e. nearly all clusters are of size 2. Similarly, when τ is small, the denominator also includes only clusters of size 2 and can be approximated as

$$\binom{n}{2} \tau^2.$$

To take an arbitrary example, if we assume all families have an extended family structure such as the one depicted in Figure S1, there are $u = 6$ pairs of unrelated family members; therefore, the proportion of clusters limited to blood relatives occurring among such families would be:

$$p_B \approx 1 - \frac{6\tau^2}{\binom{8}{2}\tau^2} = 1 - \frac{6}{28} = \frac{11}{14}.$$



Appendix Figure. Example of an extended family pedigree. There are 8 family members (2 parents, their 3 children, son-in-law, and 2 grandchildren) and 6 pairs of non-blood relatives (depicted by the dashed lines).

Appendix Table. Family clusters of influenza A (H5N1), Dec 2003–Dec 2006*

Cluster	Onset of index case	Country	Age, y/Sex	Relation to index case	H5N1	Onset	Outcome
1	Dec/03	Viet. (N)	12/F	Self	+	Dec 25	D
			30/F	Mother	+	Jan 1	D
2	Dec/03	Viet. (N)	5/M	Self	+	Dec 29†	D
			7/F	Sister	NT	NN	D
3	Jan/04	Viet. (N)	31/M	Self	NT	Jan 7†	D
			30/F	Sister	+	Jan 10	D
			28/F	Wife	+	Jan 10	R
			23/F	Sister	+	Jan 11	D
4	Jan/04	Thailand	6/M	Self	+	Jan 8	D
			33/F	Mother	NT	Jan 8	D
5	Jul/04	Viet. (S)	19/M	Self	NT	Jul 23	D
			22/F	Cousin	NT	NN	D
			25/F	Sister	+	Jul 31	D
6	Sep/04	Thailand	11/F	Self	NT	Sep 2	D
			26/F	Mother	+	Sep 11	D
			32/F	Aunt	+	Sep 16	R
7	Dec/04	Viet. (N)	46/M	Self	+	Dec 26	D
			42/M	Brother	+	Jan 10†	R
			36/M	Brother	+	Not ill	Not ill
8	Jan/05	Viet. (S)	17/M	Self	+	Jan 10†	D
			22/F	Sister	NN	NN	Unk#
9	Jan/05	Viet (S)	35/F	Self	+	Jan 14	D
			13/F	Daughter	+	Jan 20	D
10	Jan/05	Camb.	14/M	Self	NT	NN	D
			25/F	Sister	+	Jan 21	D
			21/M	Self	+	Feb 14	Unk#
11	Feb/05	Viet. (N)	14/F	Sister	+	Feb 23	Unk#
			80/M	Grandfather	+	Not ill	Not ill
			69/M	Self	+	Feb 19	D
12	Feb/05	Viet. (N)	61/F	Wife	+	Not ill	Not ill
			13	Mar/05	Viet.	13/F	Self
5/M	Brother	+	Mar 12†			R	
14	Mar/05	Viet. (N)	Adult/F	Aunt	P	NN	Unk#
			39/M	Self	+	Mar 22†	Unk#
			Adult/F	Wife	+	Mar 22†	Unk#
			4 mo/NN	Child	+	Mar 22†	Unk#
			3/NN	Child	+	Mar 22†	Unk#
15	Jul/05	Indonesia	10/NN	Child	+	Mar 22†	Unk#
			8/F	Self	+	Jun 24	D
			1/F	Sister	NT	Jun 29	D
16	Sep/05	Indonesia	38/M	Father	+	Jul 2	D
			21/M	Self	+	Sep 20	Unk#
17	Oct/05	Thailand	4/M	Nephew	+	Oct 4	R
			48/M	Self	+	Oct 13	D
18	Oct/05	Indonesia	7/M	Son	+	Oct 16	Unk#
			19/F	Self	+	Oct 19	D
19	Oct/05	China	8/M	Brother	+	Oct 25	Unk#
			12/F	Self	NT	16 Oct	D
20	Nov/05	Indonesia	9/M	Brother	+	17 Oct	R
			7/M	Self	NT	3 Nov	D
21	Dec/05	Turkey	20/M	Brother	NT	3 Nov	D
			16/M	Brother	+	6 Nov	Unk#
			14/M	Self	+	26 Dec	D
22	Dec/05	Turkey	15/F	Sister	+	28 Dec	D
			12/F	Sister	+	28 Dec	D
			9/F	Self	+	29 Dec	R
23	Jan/06	Turkey	3/M	Brother	+	30 Dec	R
			5/M	Self	+	8 Jan	R
24	Jan/06	Indonesia	14/F	Sister	+	8 Jan	D
			13/F	Self	+	6 Jan	D
			4/M	Brother	+	8 Jan	D

25	Jan/06	Iraq	15/F	Self	+	17 Jan‡	D
			39/M	Uncle	+	18 Jan	D
26	Feb/06	Azer.	17/F	Self	+	23 Feb‡	D
			20/F	Cousin	+	3 Mar‡	D
			16/M	Cousin	+	10 Mar‡	D
			17/F	Friend	+	8 Mar‡	D
27	Mar/06	Azer.	15/F	Self	+	11 Mar	R
			17/F	Cousin	+	11 Mar	R
28	Mar/06	Egypt	6/F	Self	+	NN	Unk#
			18 mo/F	Sister	+	NN	Unk#
29	Apr/06	Indonesia	37/F	Self	NT	27 Apr	D
			15/M	Son	+	9 May‡	D
			17/M	Son	+	12 May‡	D
			28/F	Sister	+	10 May‡	D
			18 mo/F	Niece	+	14 May‡	D
			25/M	Brother	+	NN	Unk#
			10/M	Nephew	+	13 May‡	D
			32/M	Brother	+	15 May	D
30	May/06	Indonesia	10/F	Self	+	16 May	D
			18/M	Brother	+	16 May	D
31	May/06	Indonesia	15 F	Self	+	17 May	Unk#
			27 M	Brother	+	28 May	R
32	May/06	Indonesia	7/F	Self	+	26 May	D
			10/M	Brother	NT	29 May‡	D
33	Jul/06	Indonesia	17M	Self	+	26 Jul	R
			20M	Cousin	NT	26 Jul	D
34	Sep/06	Indonesia	11M	Self	+	16 Sep	D
			21F	Sister	+	19 Sep	Unk#
35	Sep/06	Indonesia	24 M	Self	NT	16 Sep	D
			20M	Brother	+	17 Sep	D
36	Dec/06	Egypt	30F	Self	+	17 Dec†	D
			16F	Niece	+	19 Dec†	D
			26M	Brother	+	17 Dec†	D

*Camb., Cambodia; D, respiratory death; N, north; NT, not tested; NN, not noted; P, pending; R, recovered; S, south; Unk, unknown; Viet., Vietnam; Azer., Azerbaijan.

†Date of hospitalization.

‡Date of death.

#Had respiratory symptoms, was hospitalized (unknown for #13 and #29), and outcome was unknown.