PLOS ONE

Sample Size Considerations for One-to-One Animal Transmission Studies of the Influenza A Viruses

Hiroshi Nishiura[1,2]*, **Hui-Ling Yen[3]**, **Benjamin J. Cowling[1]**

1 School of Public Health, The University of Hong Kong, Hong Kong SAR, China, 2 PRESTO, Japan Science and Technology Agency, Saitama, Japan, 3 Centre of Influenza Research, School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong SAR, China

Abstract

Background: Animal transmission studies can provide important insights into host, viral and environmental factors affecting transmission of viruses including influenza A. The basic unit of analysis in typical animal transmission experiments is the presence or absence of transmission from an infectious animal to a susceptible animal. In studies comparing two groups (e.g. two host genetic variants, two virus strains, or two arrangements of animal cages), differences between groups are evaluated by comparing the proportion of pairs with successful transmission in each group. The present study aimed to discuss the significance and power to estimate transmissibility and identify differences in the transmissibility based on one-to-one trials. The analyses are illustrated on transmission studies of influenza A viruses in the ferret model.

Methodology/Principal Findings: Employing the stochastic general epidemic model, the basic reproduction number, R_0, is derived from the final state of an epidemic and is related to the probability of successful transmission during each one-to-one trial. In studies to estimate transmissibility, we show that 3 pairs of infectious/susceptible animals cannot demonstrate a significantly higher transmissibility than $R_0 = 1$, even if infection occurs in all three pairs. In comparisons between two groups, at least 4 pairs of infectious/susceptible animals are required in each group to ensure high power to identify significant differences in transmissibility between the groups.

Conclusions: These results inform the appropriate sample sizes for animal transmission experiments, while relating the observed proportion of infected pairs to R_0, an interpretable epidemiological measure of transmissibility. In addition to the hypothesis testing results, the wide confidence intervals of R_0 with small sample sizes also imply that the objective demonstration of difference or similarity should rest on firmly calculated sample size.

Competing Interests: The authors declare that co-authors B.J. Cowling and H. Nishiura are PLOS ONE Editorial Board members. Also, Hui-Ling Yen will also become academic editor from January 2013. This does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials.

* E-mail: nishiura@hku.hk

Introduction

The transmission potential of a respiratory virus is commonly measured by the basic reproduction number, R_0, i.e. the average number of secondary cases produced by a typical primary case throughout the entire course of host infection, which has been regarded as one of the most important quantities in infectious disease epidemiology [1]. The value of R_0 not only informs how transmissible an infected individual is, but also gives three epidemiological insights into the transmission dynamics, i.e., (i) the risk of observing an epidemic given a certain number of infected individuals, (ii) the risk of infection in an individual throughout the course of an epidemic (given the epidemic), and (iii) the minimum control effort that is required to prevent or curb an epidemic based on a threshold theorem [1]. Among various methods for estimating R_0, animal transmission experiments have been used as a useful tool for measuring the transmissibility in controlled conditions [2], although it should be noted that R_0 is not only the property of a pathogen but also reflects the specific entire host-pathogen-environmental system, and thus, R_0 in the experimental setting is not directly applicable to other (e.g. natural) settings. However, through transmission experiments, one can identify the importance of various factors on transmissibility, including host (e.g. genetic variants or immune status), virus (e.g. different strains), and environmental (e.g. ambient temperature/humidity) factors and their interactions, thereby providing important insights into mechanisms of infection and transmission.

Transmission experiments of influenza A viruses have helped to determine the molecular mechanisms of adaptation in human host due to the multi-host nature of the virus [3–5]. It is possible to study transmission in humans in controlled experimental settings [6] and natural settings [7], but such studies are resource intensive. Moreover, prior influenza exposure history in humans varies between individuals and is difficult to control. Although animal studies cannot replace human studies, they can provide complementary information on factors affecting transmission. The ferret model tends to be preferred over the mouse and guinea pig models

for transmission experiments, because ferrets display clinical signs and symptoms which include fever, nasal discharge, and sneeze reflex [8–10]. In some studies, the potential of respiratory droplet transmission (inclusive of both droplet and aerosol transmissions) has been examined by placing a susceptible ferret in a cage next to another cage with a ferret inoculated with an influenza virus and allowing an exchange of air between the cages. Typically, such one-to-one transmission has been examined for two to four pairs for each virus, and the proportion of pairs with successful transmission has been compared between two (or more) different influenza viruses.

Despite a number of published ferret transmission experiment studies of influenza viruses, there has been no explicit estimation of R_0 from experiments that typically involve very small sample sizes, and moreover, the sample size rationale of such an experiment has not been extensively discussed. Although a common approach is now to use three pairs of ferrets in each virus group, and to compare the proportions of pairs with successful transmission between each group, the appropriate sample size and proper analysis of results have not been investigated. For instance, suppose that researchers did not observe any transmission for one virus (i.e. $k/n = 0/3$ where n and k represent the numbers of pairs and infected pairs, respectively), while all three pairs resulted in transmission for the other virus (i.e. $k/n = 3/3$). Given this result, we would like to know (i) if the difference in the transmissibility between the two groups is statistically significant and (ii) the degree of difference in transmissibility given the small sample size. The present study aimed to discuss the significance and power of one-to-one animal transmission experiments with particular reference to ferret transmission studies of influenza A viruses as a case study.

Materials and Methods

One-to-one transmission experiment data

We start by presenting summary results of the published transmission experiment studies of influenza A viruses using the ferret model. Table 1 summarizes a total of 12 transmission experiment studies that were conducted under a one-to-one transmission experiment design [3,5,8,11–19]. In the present study, we restrict our interest to the ferret model, especially its use in examining respiratory transmission, for consistency and clarity both in theory and biology. Among the total of 12 studies, nine investigated H1N1 viruses including two on 1918–19 pandemic viruses and seven on 2009 pandemic viruses. Three studies investigated the transmissibility of H5N1 viruses, two on H3N2 viruses and one on H2N2 viruses. In principle, those studies share the experimental design (i.e. inoculation of one ferret in a cage and exposure of the other ferret in an adjacent cage), but the details have been variable. The air flow (e.g. direction and air exchange rate) has not been strictly regulated by common rules, and viral dose for inoculation have not been identical among these studies, and thus, various differences in experimental designs prohibit pooling of data from differently designed studies.

With regard to the sample size (i.e. the number of pairs for each virus), all studies used small numbers of ferrets, with most choosing 3 pairs of infectious/susceptible ferrets in each group. There were two studies that used 4 pairs. Two studies used different pair numbers between two groups, i.e., one study used 2 pairs for the control and 4 pairs for the comparison group, and the other study used 3 pairs for the control and 6 pairs for the comparison group. The results (i.e. the number of pairs with successful transmission) shown in Table 1 represent the highest and lowest reported numbers among the all combinations of two viruses, and the judgment of difference (or similarity) was drawn in the original

publications based on the corresponding results. The primary objective of the original studies was either to identify molecular mechanisms (e.g. specific viral gene, amino acid or protein) governing the transmissibility ($n = 5$) or to quantify or compare the capacity of aerosol transmissibility ($n = 7$).

Stochastic general epidemic model

To allow comparison of the transmissibility, here we express the result from one-to-one transmission experiment (i.e. the proportion of pairs with transmission) as a function of R_0. First of all, we adopt an assumption that each pair is independent of other pairs, including no air-exchange between pairs. Let $p_{s,i}(t)$ be the conditional probability of observing s susceptible and i infected ferrets at time t given the initial condition of susceptible and infected ferrets (s_0, i_0) at time 0, i.e.,

$$p_{s,i}(t) = \Pr\{(S(t), I(t)) = (s, i) | (S(0), I(0)) = (s_0, i_0)\}, \qquad (1)$$

then the so-called "stochastic general epidemic" model is described by the following differential-difference equation:

$$\frac{d}{dt} p_{s,i}(t) = (i+1) p_{s,i+1}(t) + \frac{R_0}{N-1}(s+1)(i-1) p_{s+1,i-1}(t)$$
$$- \frac{R_0}{N-1} s i p_{s,i}(t) - i p_{s,i}(t), \qquad (2)$$

where N is the total population size (N = 2 in the case of one-to-one experiment) and t represents the multiple of the mean infectious period (i.e. the time unit is normalized by the mean infectious period). Here it should be noted that the mass action part has been scaled by (N−1), and not by N, because of small population size that requires us to precisely consider the impact of N on the incidence term. That is, in the case of small N, the transient risk of infection should be proportional to I(t)/(N−1) which can be exemplarily understood for N = 2 (i.e. if we use I(t)/N for calculating the incidence, it would indicate erroneously that the half of infectious individuals I(t)/2 would contribute to the transmission). Since the initial condition gives $p_{s0,i0}(0) = 1$, the probability of successful transmission by infinite time, q, is computed by $p_{0,0}(\infty)$ [20–22] and the solution is

$$q = \frac{R_0}{R_0 + 1}. \qquad (3)$$

Note that the analytical solution of q for small $\mathcal{N}$ is $q = R_0/(R_0 + \mathcal{N} - 1)$ which is different from what has been previously discussed [22]. Since the one-to-one transmission experiment handles the binary outcome (i.e. success or failure of transmission), the probability of transmission is computed by employing a binomial distribution. That is, for n independent pairs of one-to-one transmission experiments, the probability of observing k pairs with successful transmission is

$$\Pr(K = k; n, R_0) = \binom{n}{k} \left(\frac{R_0}{R_0 + 1}\right)^k \left(\frac{1}{R_0 + 1}\right)^{n-k}. \qquad (4)$$

The maximum likelihood estimator of R_0 based on the observed average frequency of successful transmission, k/n, is given by equating $R_0/(R_0 + 1) = k/n$ which yields.

Table 1. Study characteristics of the published one-to-one animal transmission experiment of influenza A viruses.

	Year of publication	Subtype	Infected animals in control group (k/n)[†]	Infected animals in comparison group (k/n)[†]	Judgment[‡]	Study objective(s)
Maines et al. [11]	2006	H3N2 & H5N1	3/3	0/3	Different	To identify the importance of virus internal protein genes in regulating transmissibility
Tumpey et al. [12]	2007	H1N1	3/3	0/3	Different	To show that only a modest change in the 1918 influenza hemagglutinin receptor binding site alters the transmissibility
van Hoeven et al. [8]	2009	H1N1	3/3	0/3	Different	To identify genetic determinants that govern airborne transmission among 1918-avian H1N1 ifnluenza reassortant viruses
Munster et al. [13]	2009	H1N1	4/4	4/4	Similar	To study aerosol transmission of the pandemic 2009 A(H1N1) influenza virus as compared with a seasonal 2007 A(H1N1) virus
Maines et al. [14]	2009	H1N1	3/3	2/3	Different	To compare the transmissibility of H1N1-2009 against a seasonal H1N1 virus through respiratory droplets
Pappas et al. [15]	2010	H2N2	3/3	0/3	Different	To examine the transmissibility of human H2N2 viruses isolated during the 1957/58 pandemic
Kiso et al. [16]	2010	H1N1	3/3	3/3	Similar	To compare the transmissibility between oseltamivir-resistant and sensitive H1N1-2009.
Van Doremalen et al. [17]	2011	H1N1	3/3	3/3	Similar	To investigate the effect of residue 227 in hemagglutinin on cell tropism and transmission of pH1N1 2009.
Koster et al. [18]	2012	H1N1	4/4	4/4	Similar	To develop a method to detect exhaled viral aerosol transmission between unanesthetized infected and susceptible ferrets.
Pearce et al. [19]	2012	H3N2	3/3	3/3	Similar	To analyze the transmissibility of four A(H3N2)v influenza viruses isolated from humans in 2009, 2010 and 2011.
Herfst et al. [5]	2012	H5N1 & H1N1	0/2	3/4	Different	To address the concern that the virus could acquire the ability of airborne transmission under natural conditions
Imai et al. [3]	2012	H5N1 & H1N1	0/3	4/6	Different	To assess the molecular changes in HA that would allow a virus possessing subtype H5 HA to be transmissible

[†]n represents the sample size, i.e. the number of pairs of a single infected animal and a susceptible animal. k represents the number of pairs with successful transmission. The numbers among control group are the highest reported frequency of infected animals in control group, while the numbers among comparison group show the lowest reported frequency of infection. [‡]Judgment corresponds to the interpretation of difference in transmissibility.
doi:10.1371/journal.pone.0055358.t001

$$\hat{R}_0 = \frac{k}{n-k}. \tag{5}$$

The 100(1-α)% confidence interval of R_0 is calculated from the solution of $R_{0,\mathrm{CI}} = x/(n-x)$ in which x satisfies

$$\sum_{i=x}^{n} \binom{n}{i} \left(\frac{\hat{R}_0}{\hat{R}_0+1}\right)^i \left(\frac{1}{\hat{R}_0+1}\right)^{n-i} = \frac{\alpha}{2}, \tag{6}$$

for the upper bound and

$$\sum_{i=0}^{x} \binom{n}{i} \left(\frac{\hat{R}_0}{\hat{R}_0+1}\right)^i \left(\frac{1}{\hat{R}_0+1}\right)^{n-i} = \frac{\alpha}{2}, \tag{7}$$

for the lower bound, except that the lower bound is 0 when $x = 0$ and the upper bound is infinity when $x = n$. In these exceptional circumstances, the upper bound for $x = 0$ is calculated as $R_{0,\mathrm{upper}} = ((\alpha/2)^{(1/n)}-1)/(\alpha/2)^{(1/n)}$ and the lower bound for $x = n$ is calculated as $R_{0,\mathrm{lower}} = (\alpha/2)^{(1/n)}/(1-(\alpha/2)^{(1/n)})$ as can be derived from the binomial distribution [23]. In addition to the final size discussed above, statistical consideration of transient state has been given elsewhere [24].

Hypothesis testing

We subsequently consider the difference in the transmissibility in published experimental studies in two different ways, because the null hypothesis has not necessarily been mentioned in the original articles in Table 1. Let $R_{0,\mathrm{ref}}$ be a specified reference value of the basic reproduction number. The first possible way to compare the transmissibility is to regard the result from each virus as one-sample comparison, which may be the case when $R_{0,\mathrm{ref}}$ of control virus can be assumed known (e.g. pre-determined) from published studies and so on. In this scenario, we compare R_0 against $R_{0,\mathrm{ref}}$, i.e.

$$H_0 : R_0 \le R_{0,ref} \quad \text{vs.} \quad H_a : R_0 > R_{0,ref}, \tag{8}$$

which may sometimes be intended to support the notion that some key molecular structure helped to acquire substantial transmissibility for a specific virus (e.g. by setting $R_{0,\mathrm{ref}} = 1$ or $R_{0,\mathrm{ref}} = 0$). It should be noted that R_0 depends on experimental design (air change rate per hour, air flow direction, etc) and is not comparable between differently designed experiments. Using the relationship (5), the issue of comparing transmissibility is replaced by one-sample comparison of a binomial proportion. The p-value for testing (8) given that k or more pairs resulted in infection is computed by

$$\Pr(X \ge k; n, R_{0,ref}) = \sum_{i=k}^{n} \binom{n}{i} \left(\frac{R_{0,ref}}{R_{0,ref}+1}\right)^i \left(\frac{1}{R_{0,ref}+1}\right)^{n-i}, \tag{9}$$

which rejects H_0 if less than or equal to $\alpha = 0.05$. Demonstrating $R_0 > R_{0,\mathrm{ref}} = 0$ would help researchers to demonstrate the capacity of a ferret infected with a virus to "cause" (any number of) secondary transmission. Demonstrating $R_0 > R_{0,\mathrm{ref}} = 1$ is particularly useful in practice, because satisfying this condition indicates that, at least under experimental conditions, a single primary case can on average generate one or more secondary cases through the respiratory route (and thus, is regarded as possessing a substantial

transmission potential to cause an epidemic only through that particular mode of transmission). Of course, rejecting H_0: $R_0 \ge 1$ can be achieved similarly by calculating $\Pr(X < k; n, R_{0,\mathrm{ref}} = 1) \le 0.05$. The power for (8) is given by

$$\sum_{k=0}^{n} \binom{n}{k} \left(\frac{R_0}{R_0+1}\right)^k \left(\frac{1}{R_0+1}\right)^{n-k} I(\Pr(X \ge k; n, R_{0,ref}) \le \alpha), \tag{10}$$

where $I(.)$ is the indicator function.

The second way to test the transmissibility is to consider the two virus groups within the same study in Table 1 as the comparison of two binomial proportions, q_1 and q_2 (or equivalently the comparison of two basic reproduction numbers estimated for respective viruses, $R_{0,1}$ and $R_{0,2}$) under the hypotheses

$$H_0 : q_1 = q_2 \quad \text{vs.} \quad H_a : q_1 \ne q_2, \tag{11}$$

(or H_0: $R_{0,1} = R_{0,2}$) and the implementation is exactly the same as two-tailed exact test for two samples that has been already discussed elsewhere [25]. It should be noted that differing number of pairs between two virus groups can be easily addressed by varying sample sizes in the exact test. In both one-sample and two-sample cases, the sample size estimation would have to be made directly from the binomial distribution (e.g. from (10) with a desired power). However, as an alternative, the power calculation could rest on a modified Wald test statistic, i.e., the well-known score confidence interval proposed by Agresti and Coull [26,27,28].

For numerical illustrations, we consider the p-value and power for all possible patterns of final size for the number of pairs, $n = 3$, 4 and 5 for both one-sample and two-sample comparisons. These numbers of pairs are specifically considered, because 3 pairs have been conventionally adopted, and we anticipate that 6 or more pairs may not be logistically very feasible for testing many types of influenza virus at present. A one-sample comparison is made by a one-tailed Fisher's exact test, while a two-sample comparison rests on a two-tailed test. While restricting our consideration to $n = 3$, 4 and 5, we also examine the p-value for one-sample test with varying reference values of the basic reproduction number and the number of pairs (from 1 to 10), especially in the case we have $k = n$ (i.e. all pairs resulted in infection) or $k = n-1$ which are frequently the case in published experimental studies.

Results

One-sample comparison

Table 2 shows the p-value and power for one-sample comparisons given that the number of pairs was 3, 4 or 5. Even when all pairs result in infection during 3 or 4 pair study, the experiment cannot indicate that the R_0 is significantly greater than 1. Only when we have a result of 5/5, the difference can be stated as significantly greater than R_0 of 1. Moreover, in the cases all pairs result in infection, one can quantitatively examine only the lower bound of R_0, and the expected value and the upper bound of R_0 are calculated as infinite. Figure 1 shows the p-value with different numbers of pairs in the case that all pairs are infected (i.e. $k = n$) or all pairs minus 1 resulted in infection (i.e. $k = n-1$). When the reference value of R_0 is as large as 2, three pairs are not large enough to demonstrate $R_0 > 2$ at a significant level $\alpha = 0.05$, and one may need at least 8 pairs and all the eight pairs need to be infected. At a stricter threshold (e.g. $\alpha = 0.01$), seven or more pairs would be required to reject $R_{0,\mathrm{null}} = 1$. In the case of $k = n-1$, even

Table 2. One-tailed test results of the basic reproduction number based on one-to-one transmission experiment.

Number of pairs	Final size	R_0[†]	(95% CI[‡])	p-value to demonstrate $R_0>1$	power to demonstrate $R_0>1$	power to demonstrate $R_0>0$	p-value to demonstrate $R_0<1$
3	0	0.0	(0, 4.8)	1.00	0.00	1.00	1.00
	1	0.5	(0, ∞)	0.88	0.00	1.00	0.74
	2	2.0	(0, ∞)	0.50	0.00	1.00	0.70
	3	∞	(0.8, ∞)	0.13	0.00	1.00	1.00
4	0	0.0	(0, 3.0)	1.00	0.00	0.00	1.00
	1	0.3	(0, 6.0)	0.94	0.00	0.68	0.74
	2	1.0	(0, ∞)	0.69	0.00	0.94	0.69
	3	3.0	(0.7, ∞)	0.31	0.00	0.99	0.68
	4	∞	(1.3, ∞)	0.06	0.00	1.00	1.00
5	0	0.0	(0, 2.2)	1.00	0.00	0.00	1.00
	1	0.3	(0, 3.0)	0.97	0.00	0.67	0.74
	2	0.7	(0, 8.0)	0.81	0.00	0.92	0.68
	3	1.5	(0.5, ∞)	0.50	0.08	0.99	0.66
	4	4.0	(1.3, ∞)	0.19	0.33	1.00	0.67
	5	∞	(1.8, ∞)	0.03[¶]	1.00	1.00	1.00

[†]The basic reproduction number, estimated from the one-to-one transmission experiment. [‡]CI, confidence intervals. [¶]R_0 is significantly greater than 1 by one-sample Fisher's exact test.
doi:10.1371/journal.pone.0055358.t002

ten pairs would not be enough to demonstrate that $R_0>2$ at $\alpha=0.05$ given $k=n-1$.

Also, when one pair escapes infection (i.e. $k=n-1$), Figure 1 and Table 2 consistently suggest that a five-pair or smaller study
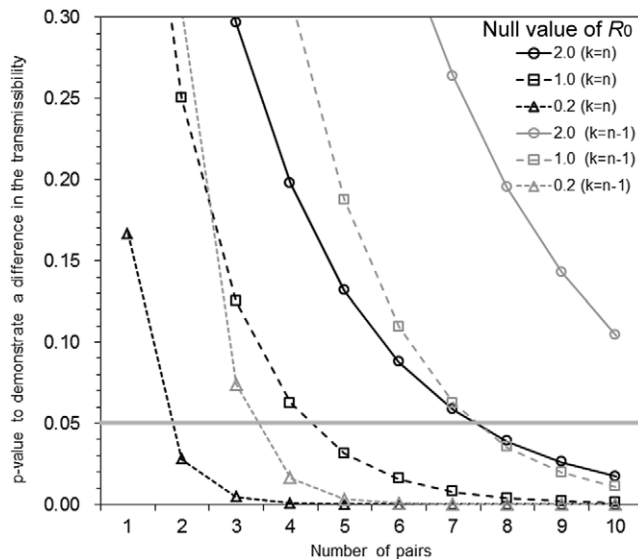

Figure 1. The p-value to demonstrate the significant difference in the transmissibility based on fully or nearly fully successful one-to-one transmission experiment. The p-values are shown to indicate the significance level at which the estimated basic reproduction number is significantly greater than the null value ($R_{0,null}$ are set to be 0.2, 1.0 and 2.0 with the null hypothesis, $H_0: R_0 \leq R_{0,null}$) given that all pairs resulted in infection (black lines; $n=k$ where n and k are the numbers of pairs and infected pairs, respectively) or all pairs minus 1 resulted in infection (grey lines; $n=k+1$). The hypothesis testing is based on one-sample comparison using the one-tailed Fisher's exact test. The horizontal grey bold line represents the significance level at 0.05.
doi:10.1371/journal.pone.0055358.g001

cannot help judge if R_0 is significantly greater than 1. The results 2/3, 3/4 or 4/5 does not indicate significant difference from $R_0=1$. All other combinations in Table 2 cannot determine if R_0 is significantly greater than 1, and more importantly, either lower or upper 95% confidence interval of estimated R_0 for these combinations always takes an extreme value (i.e. either lower bound being 0 or upper bound being infinite).

Provided that a transmission study intends to demonstrate $R_0>0$, the presence of at least one successful transmission (i.e. any k/n except for $k=0$) can yield a significant result with $p<0.01$. However, it should be remembered that power is not substantial for $k/n=1/4$ and $1/5$ (Table 2). When one intends to demonstrate that the transmission potential is less than 1, the examined total sample sizes are not enough to argue significant differences (Table 2). That is, given the number of pairs is 3, 4 or 5, it is more feasible to show that $R_0>1$ than demonstrating $R_0<1$.

Two-sample comparison

Table 3 summarizes the p-value and power for two-sample comparisons given that the number of pairs was 3, 4 or 5. Given three pairs for each sample, it is impossible to demonstrate any significant difference between two sample groups. In the case of four pairs, only a combination of 0/4 and 4/4 can indicate a significant difference. Given five pairs, three combinations (i.e. 0/5 vs 5/5, 1/5 vs 5/5 or 0/5 vs 4/5) could suggest significant difference in the transmission potential.

Discussion

The present study discussed the sample size considerations for one-to-one experimental studies of the transmission of influenza A viruses. Employing the stochastic general epidemic model, R_0 was derived from the final state of an epidemic [29,30,31], and its relevance to the probability of successful transmission during the one-to-one trial was explained. Three findings are particularly notable. First, $k/n=3/3$ and $4/4$ are not indicative of significant excess of R_0 from 1 in one-sample comparison. At least, five pairs

Table 3. Two-tailed comparison of the basic reproduction numbers based on one-to-one transmission experiment (H_0: $R_0 = R_{0,null}$[†]).

Number of pairs	Final size of comparison group	0		1		2		3		4		5	
		p-value	power	p-value	power	p-value	power	p-value	power	p-value	power	p-value	power
3	0	1.00	0.00	1.00	0.04	0.40	0.30	0.10	1.00				
	1	1.00	0.04	1.00	0.00	1.00	0.09	0.40	0.30				
	2	0.40	0.30	1.00	0.09	1.00	0.00	1.00	0.04				
	3	0.10	1.00	0.40	0.30	1.00	0.04	1.00	0.00				
4	0	1.00	0.00	1.00	0.05	0.43	0.31	0.14	0.74	0.03¶	1.00		
	1	1.00	0.05	1.00	0.04	1.00	0.13	0.49	0.37	0.14	0.74		
	2	0.43	0.31	1.00	0.13	1.00	0.07	1.00	0.13	0.43	0.31		
	3	0.14	0.74	0.49	0.37	1.00	0.13	1.00	0.04	1.00	0.05		
	4	0.03¶	1.00	0.14	0.74	0.43	0.31	1.00	0.05	1.00	0.00		
5	0	1.00	0.00	1.00	0.01	0.44	0.09	0.17	0.34	0.05¶	0.74	0.01¶	1.00
	1	1.00	0.01	1.00	0.01	1.00	0.03	0.52	0.14	0.29	0.38	0.05¶	0.74
	2	0.44	0.09	1.00	0.03	1.00	0.02	1.00	0.05	0.52	0.14	0.17	0.34
	3	0.17	0.34	0.52	0.14	1.00	0.05	1.00	0.02	1.00	0.03	0.44	0.09
	4	0.05¶	0.74	0.21	0.38	0.52	0.14	1.00	0.03	1.00	0.01	1.00	0.01
	5	0.01¶	1.00	0.05¶	0.74	0.17	0.34	0.44	0.09	1.00	0.01	1.00	0.00

[†]$R_{0,null}$, the basic reproduction number to be used in null hypothesis. ¶Estimated transmissibility from one-to-one transmission experiment is significantly different from $R_{0,null}$ by two-sample Fisher's exact test.
doi:10.1371/journal.pone.0055358.t003

would be required to demonstrate significant difference in the one-sample comparison. Second, $n = 3$ is not enough to show any significant difference in two-sample comparisons. Third, $k = n$ can yield the significant difference when $n = 5$ or greater in one-sample comparison, but one has to remember that the expected value and the upper confidence interval of R_0 would be calculated as infinite for small n. That is, while the experiment may be able to show significant difference from reference value, $k = n$ can inform only the lower bound of R_0. With the very limited sample sizes such as $n = 3$, 4 or 5, it is always the case that either lower or upper 95% confidence interval takes an extreme value (i.e. lower $= 0$ or upper $= \infty$). Keeping these points in mind, one can plan the one-to-one transmission study with reference to our computed results in Tables 2 and 3, while relating the observed proportion of infected pairs to R_0, an interpretable epidemiological measure of transmissibility.

The most important caveat in the present study in relation to the common practice is that $n = 3$ is not enough to show a significant difference as well as $R_0 > 1$ for one-sample comparison, while it can demonstrate $R_0 > 0$. Moreover, comparing a group with $n = 3$ against a reference group with the same sample size does not allow researchers to demonstrate any significant difference in the transmissibility between two sample groups. If two samples have to be compared, $n = 4$ would be regarded as minimum, and moreover, $k = n$ for $n = 4$ and $k = n$ or $k = n - 1$ for $n = 5$, respectively, would have to be required along with the absence of infected pairs in the control group. To interpret some results of the published studies in Table 1 which concluded difference in the transmissibility between two viruses, one sample interpretation may better be adopted for each virus against $R_{0,null} = 0$ (rather than comparing two sample groups). Over-interpretation of results without significant difference should be avoided.

It should be noted that demonstrating similar transmissibility between two groups is even more difficult in this context for two reasons. First, the sample size (i.e. the number of pairs) is very limited, and thus, it is too frequent that we do not observe any significant difference between two sample groups (Table 3). Second, demonstrating similarity must be distinguished from showing the absence of a significant difference, especially when the sample sizes are very limited. Demonstrating similarity in transmissibility would likely require much larger sample sizes, as seen in noninferiority and equivalence randomized trials [32]. Similarly, demonstrating the absence of substantial transmissibility in a single sample group is also a difficult task. As was shown in Table 2, showing $R_0 < 1$ cannot be achieved by $n = 3$, 4 or 5.

Two limitations should be noted. First, every transmission experiment study examines not only the frequency of successful transmissions but also other factors including mortality, weight loss, patterns in virus shedding, clinical signs and symptoms, behavioral changes and so on. Thus, although the quantification of transmissibility should strictly adhere to the frequency of transmission events, the interpretation of adaptation, pathogenicity and infectiousness (e.g. in the sense of virus replications in infected hosts) should be judged from multiple results. Moreover, study objectives of transmission experiment may not necessarily be to demonstrate differential transmissibility (e.g. may only be to prove that pre-symptomatic transmission can occur [33]). In this regard, the present study has focused only on a single aspect of experimental findings. Second, we adopted an independence assumption between pairs which may not be strictly the case in all published studies. If the transmission studies handled infectious virus with substantially high aerosol transmission potential, this assumption is violated. However, explicitly addressing this point cannot be achieved by employing a simple general stochastic epidemic model, and rather, would require much more complex

modeling analysis. Resolving second problem (of dependence between pairs) by mathematical modeling are our forthcoming future studies.

It is important that the researchers can calculate the most appropriate sample size (with or without a reference sample group) depending on study objectives. Sample size rationale should be formulated in the future including other types of transmission experiment design, such as exposing multiple animals to multiple infectious animals, or exposing two animals in two different manners (e.g. with and without direct contact). As the translation of the proportion of infected pairs into R_0 can permit object-oriented experimental design, the objectives, design and findings of

transmission experiment studies have to be reviewed and the corresponding sample size rationale should be discussed. The present study could be regarded as a starting point to extensively consider hypothesis testing of animal transmission experiment results using epidemiologically well-defined parameters.

Author Contributions

Conceived and designed the experiments: HN. Performed the experiments: HN. Analyzed the data: HN. Contributed reagents/materials/analysis tools: HN HLY BJC. Wrote the paper: HN HLY BJC.

References

1. Diekmann O, Heesterbeek JAP (2000) Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation. New York: Wiley.
2. De Jong MC, Kimman TG (1994) Experimental quantification of vaccine-induced reduction in virus transmission. Vaccine 128: 761–766.
3. Imai M, Watanabe T, Hatta M, Das SC, Ozawa M, et al. (2012) Experimental adaptation of an influenza H5 HA confers respiratory droplet transmission to a reassortant H5 HA/H1N1 virus in ferrets. Nature. 486: 420–428.
4. Russell CA, Fonville JM, Brown AE, Burke DF, Smith DL, et al. (2012) The potential for respiratory droplet-transmissible A/H5N1 influenza virus to evolve in a mammalian host. Science. 336: 1541–1547.
5. Herfst S, Schrauwen EJ, Linster M, Chutinimitkul S, de Wit E, et al. (2012) Airborne transmission of influenza A/H5N1 virus between ferrets. Science. 336: 1534–1541.
6. Killingley B, Enstone J, Booy R, Hayward A, Oxford J, et al. (2011) Potential role of human challenge studies for investigation of influenza transmission. Lancet Infect Dis. 11: 879–886.
7. Lee VJ, Yap J, Cook AR, Chen MI, Tay JK, et al. (2010) Effectiveness of public health measures in mitigating pandemic influenza spread: a prospective sero-epidemiological cohort study. J Infect Dis. 202: 1319–1326.
8. Van Hoeven N, Pappas C, Belser JA, Maines TR, Zeng H, et al. (2009) Human HA and polymerase subunit PB2 proteins confer transmission of an avian influenza virus through the air. Proc Natl Acad Sci U S A. 106: 3366–3371.
9. Tellier R (2009)Aerosol transmission of influenza A virus: a review of new studies. J R Soc Interface. 6: S783–S790.
10. Belser JA, Katz JM, Tumpey TM (2011) The ferret as a model organism to study influenza A virus infection. Dis Model Mech. 4: 575–579.
11. Maines TR, Chen LM, Matsuoka Y, Chen H, Rowe T, et al. (2006) Lack of transmission of H5N1 avian-human reassortant influenza viruses in a ferret model. Proc Natl Acad Sci U S A. 103: 12121–12126.
12. Tumpey TM, Maines TR, Van Hoeven N, Glaser L, Solórzano A, et al. (2007) A two-amino acid change in the hemagglutinin of the 1918 influenza virus abolishes transmission. Science. 315: 655–659.
13. Munster VJ, de Wit E, van den Brand JM, Herfst S, Schrauwen EJ, et al. (2009) Pathogenesis and transmission of swine-origin 2009 A(H1N1) influenza virus in ferrets. Science. 325: 481–483.
14. Maines TR, Jayaraman A, Belser JA, Wadford DA, Pappas C, et al. (2009) Transmission and pathogenesis of swine-origin 2009 A(H1N1) influenza viruses in ferrets and mice. Science. 325: 484–487.
15. Pappas C, Viswanathan K, Chandrasekaran A, Raman R, Katz JM, et al. (2010) Receptor specificity and transmission of H2N2 subtype viruses isolated from the pandemic of 1957. PLoS One. 5: e11158.
16. Kiso M, Shinya K, Shimojima M, Takano R, Takahashi K, et al. (2010) Characterization of oseltamivir-resistant 2009 H1N1 pandemic influenza A viruses. PloS Pathog. 6: e1001079.
17. van Doremalen N, Shelton H, Roberts KL, Jones IM, Pickles RJ, et al. (2011) A single amino acid in the HA of pH1N1 2009 influenza virus affects cell tropism

18. in human airway epithelium, but not transmission in ferrets. PLoS One. 6: e25755.
19. Koster F, Gouveia K, Zhou Y, Lowery K, Russell R, et al. (2012) Exhaled aerosol transmission of pandemic and seasonal H1N1 influenza viruses in the ferret. PLoS One. 7: e33118.
20. Pearce MB, Jayaraman A, Pappas C, Belser JA, Zeng H, et al. (2012) Pathogenesis and transmission of swine origin A(H3N2)v influenza viruses in ferrets. Proc Natl Acad Sci U S A. 109: 3944–3949.
21. Billard L, Zhao Z. (1993) The stochastic general epidemic revisited and a generalization. IMA J Math Appl Med Biol 10: 67–75.
22. Bartlett MS (1949) Some evolutionary stochastic processes. J R Stat Soc Ser B 11: 211–229.
23. Velthuis AG, Bouma A, Katsma WE, Nodelijk G, De Jong MC (2007) Design and analysis of small-scale transmission experiments with animals. Epidemiol Infect. 135: 202–217.
24. Nishiura H, Klinkenberg D, Roberts M, Heesterbeek JAP (2009) Early epidemiological assessment of the virulence of emerging infectious diseases: A case study of an influenza pandemic. PLoS ONE 4: e6852.
25. Velthuis AG, de Jong MC, de Bree J, Nodelijk G, van Boven M (2002) Quantification of transmission in one-to-one experiments. Epidemiol Infect 128: 193–204.
26. Thomas RG, Conlon M (1992) Sample size determination based on Fisher's Exact Test for use in 2×2 comparative trials with low event rates. Control Clin Trials. 13: 134–147.
27. Agresti A, Coull BA (1998) Approximate is better than 'exact' for interval estimation of binomial proportions. Am Stat 52: 119–126.
28. Krishnamoorthy K, Peng J (2007) Some properties of the exact and score methods for binomial proportion and sample size calculation. Communications in Statistics –Simulation and Computation 36: 1171–1186.
29. Nishiura H (2012) Early detection of nosocomial outbreaks caused by rare pathogens: A case study employing score prediction interval. Osong Public Health and Research Perspectives 3: 121–127.
30. Nishiura H, Yan P, Sleeman CK, Mode CJ (2012) Estimating the transmission potential of supercritical processes based on the final size distribution of minor outbreaks. J Theor Biol 294: 48–55.
31. Klick B, Nishiura H, Ng S, Fang VJ, Leung GM, et al. (2011) Transmissibility of seasonal and pandemic influenza in a cohort of households in Hong Kong in 2009. Epidemiology 22: 793–796.
32. Nishiura H, Chowell G, Castillo-Chavez C (2011) Did modeling overestimate the transmission potential of pandemic (H1N1-2009)? Sample size estimation for post-epidemic seroepidemiological studies. PLoS ONE 6: e17908.
33. Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJ, et al. (2006) Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. JAMA. 295: 1152–1160.
34. Roberts KL, Shelton H, Stilwell P, Barclay WS (2012) Transmission of a 2009 H1N1 pandemic influenza virus occurs before fever is detected, in the ferret model. PLoS One. 7: e43303.