

RESEARCH ARTICLE

The transmissibility of noroviruses: Statistical modeling of outbreak events with known route of transmission in Japan

Ryota Matsuyama^{1,2}, Fuminari Miura^{1,3}, Hiroshi Nishiura^{1,2*}

1 Graduate School of Medicine, Hokkaido University, Kita 15 Jo Nishi 7 Chome, Kita-ku, Sapporo, Japan, **2** CREST, Japan Science and Technology Agency, 4-1-8, Honcho, Kawaguchi-shi, Saitama, Japan, **3** Department of Urban Engineering, Graduate School of Engineering, the University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan

* nishiurah@med.hokudai.ac.jp



OPEN ACCESS

Citation: Matsuyama R, Miura F, Nishiura H (2017) The transmissibility of noroviruses: Statistical modeling of outbreak events with known route of transmission in Japan. PLoS ONE 12(3): e0173996. <https://doi.org/10.1371/journal.pone.0173996>

Editor: Christiane E. Wobus, University of Michigan, USA, UNITED STATES

Received: December 20, 2016

Accepted: March 1, 2017

Published: March 15, 2017

Copyright: © 2017 Matsuyama et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All datasets used in this study are within the paper and its Supporting Information files.

Funding: HN received funding support from the Japan Agency for Medical Research and Development, the Japan Society for the Promotion of Science (JSPS) KAKENHI (Grant numbers 16KT0130, 16K15356 and 26700028), the Program for Advancing Strategic International Networks to Accelerate the Circulation of Talented Researchers, the Japan Science and Technology

Abstract

In Japan, the fraction of norovirus outbreaks attributable to human-to-human transmission has increased with time, and the timing of the increased fraction has coincided with the increase in the observed fraction of genogroup II genotype 4 (GII.4). The present study aimed to estimate the time-dependent changes in the transmissibility of noroviruses. The effective reproduction number (R_y), for year y , was estimated by analyzing the time series surveillance data for outbreak events from 2000 to 2016. R_y was estimated by using the fraction of outbreak events that were attributable to human-to-human transmission and by employing three different statistical models that are considered to mechanistically capture the possible data-generating process in different ways. The R_y estimates ranged from 0.14 to 4.15 in value, revealing an overall increasing trend ($p < 0.05$ for all three models). The proportion of outbreaks caused by GII and GII.4 viruses among the total events also increased with time, and positive correlations were identified between transmissibility and these proportions. Parametric modeling of R_y indicated that the time-dependent patterns of R_y were better described by a step function plus linear trend rather than the step function alone that reflects the widespread use of reverse transcriptase PCR (RT-PCR) in and after 2007 for laboratory diagnosis. Accordingly, we conclude that norovirus transmissibility has increased over the past 16 years in Japan. The change is at least partially explained by the time-dependent domination of the contagious GII genogroup (e.g., GII.4), indicating that noroviruses better fitted to humans have selectively caused the human-to-human transmissions, thereby altering the epidemiology of this pathogen.

Introduction

Norovirus, a member of the *Caliciviridae* family, is a positive-sense single-stranded RNA virus. The virus is known to be a major causal agent of acute gastroenteritis in humans, especially during the winter season. Outbreaks of norovirus infection (NVI) have huge societal and health care costs globally because the virus has a strong infectious nature and perceived acute

Agency (JST) CREST program and RISTEX program for Science of Science, Technology and Innovation Policy. RM also received funding support from JSPS KAKENHI (Grant number 16H06581). The funders played no roles in the study design, data collection and analysis, decision to publish, or manuscript preparation.

Competing interests: The authors have declared that no competing interests exist.

symptoms shortly after illness onset [1,2]. The major mode of NVI transmission includes food-borne as well as human-to-human transmission via fecal-oral and vomit-oral routes. The median infective dose is as small as 18 to 10^3 virus particles [3], and viruses remain stable in the environment [4]. Since norovirus was first recognized, possible interventions to prevent secondary transmissions with it have been sought, but controlling it is difficult because of its speed of infection, the mild, nonspecific symptoms it causes, and its multiple transmission routes.

As part of the evolutionary process, viruses change their pathogenesis, virulence and transmissibility characteristics. As seen in the evolution of influenza virus [5], the multistrain dynamics of norovirus is considered to involve antigenic drift (i.e., changes in epitopes within the capsid P2 domain) and shift (e.g., overlapping recombination in open reading frame, ORF 1, and ORF 2), which allows the virus to escape the pressure of human immunity [6–9]. For instance, genogroup II (GII) and genotype 4 (GII.4) strains have caused at least six global epidemics [10–15] during the past 20 years, and have experienced antigenic evolution [6,7]. The newly emerged genotype GII.17 is no exception in this context: it has been the predominant strain in a number of recent disease outbreaks [16–18].

Antigenic evolution most likely involves evolutionary changes in virus transmissibility. To assess such transmissibility, the reproduction number (i.e., the average number of secondary infections generated by one primary case) hereafter called R_y , is an indispensable and representative epidemiological measurement. R_y does not dynamically change in a finer time scale than year, and in every single year the interpretations of R_y are analogous to those of R_0 , except the fact that R_y is assumed to be able to vary by years. When it comes to the transmissibility of NVIs, R_y has been estimated in a few published studies, where the estimated basic reproduction number (R_0) broadly ranged from 2–18 in closed or semi-closed environments [19–23]. However, the transmissibility has yet to be estimated for an entire human population and, notably, we have yet to explore any possible signature of time-dependent changes in the R_y of NVI to decipher the transmissibility.

Therefore, in the present study, we have explored whether a time-dependent signature exists for norovirus transmissibility by investigating the epidemiological dataset for Japan [24]. Our purpose was to estimate the R_y for NVI as a function of time using the national surveillance data and employing a few parsimonious mathematical models. We also explored the possible impact of the predominance of genogroup GII and genotype GII.4 on the time-dependent change in NVI transmissibility. These study objectives were accomplished using the following methods and data.

Materials and methods

Data source

In Japan, NVI outbreak events with confirmed virus detection are notified via the prefectural institutes of public health to the National Institute of Infectious Diseases as part of the National Epidemiological Surveillance of Infectious Diseases (NESID; <http://www.nih.go.jp/niid/ja/iasr-noro.html>) requirements [24]. The present study used the time series of the notified outbreak events (i.e., the count data of outbreak events) from September 2000 to August 2016 across Japan. When an outbreak is notified via the prefectural institute of public health, the notification document includes whether the outbreak was foodborne or mainly caused by human-to-human transmission. The notification system trusts prefectural institutes to judge whether the outbreak involves human-to-human transmission or is foodborne based on the epidemiological investigations, and if the cause remains unclear even after an epidemiological investigation, the outbreak events are reported as cause “unknown”. In many instances,

foodborne outbreaks (attributed to “environmental” exposure) are notified to this surveillance due to the need to virologically verify that a foodborne outbreak was caused by noroviruses, because Food Hygiene Law enforces the suspension of business at any restaurant or other eating facilities with norovirus outbreaks. Because the type of virus responsible for the outbreak is typed using the detected virus, the notification data also contains genogroup and genotype information. Also, because the notification relies on virological confirmation, the absolute number of outbreak events will be under stated, but here we assume that the proportion of observed foodborne outbreaks among all notified outbreaks reflects the unbiased proportion. Potential ascertainment biases are examined in sensitivity analysis.

Mathematical model

Fig 1 illustrates our proposed model building. Once a person is environmentally exposed to norovirus (i.e., via shellfish or other contaminated foods), index case(s) are classified as foodborne NVIs. Some of the foodborne NVIs decline to extinction, but the remainder involve chains of human-to-human secondary transmissions with R_y in year y (hereafter, we employ a piecewise constant reproduction number by 1 year). If the value of R_y is very large, a certain number of secondary, tertiary and further cases may occur. In such a situation, the majority of cases are caused by human-to-human transmission. In contrast, foodborne cases will be in the majority if a large number of individuals acquire NVIs directly from contaminated foods and if very few human-to-human transmissions follow. The judgment at the prefectural institute of public health regarding the most plausible route of transmission (i.e., whether the outbreak event was foodborne or human-to-human) must take place by investigating the following: (i)

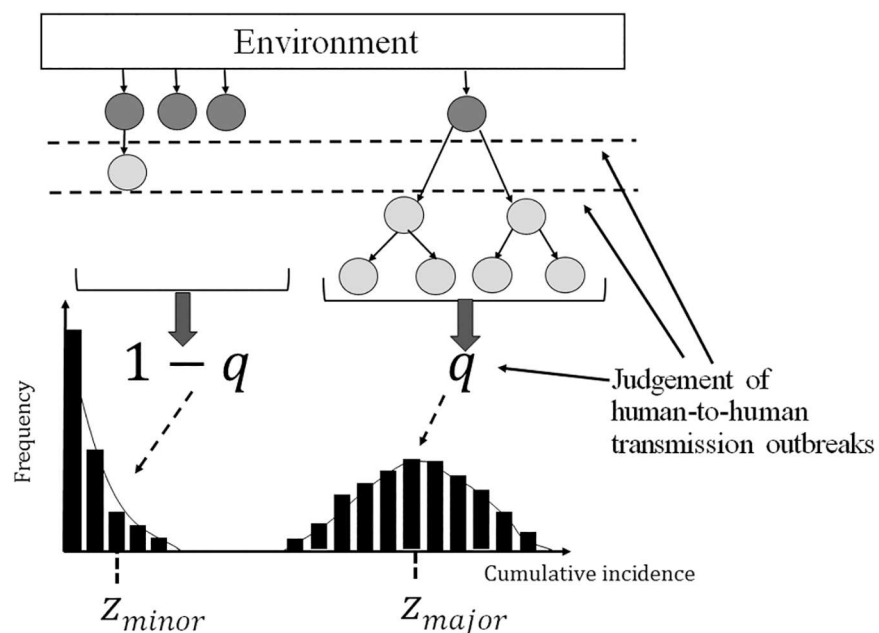


Fig 1. The data-generating process used for investigating norovirus outbreak events. Foodborne cases (dark gray) generate secondary cases via human-to-human transmissions (light gray). Given the final size distribution, let q be the proportion of major epidemics among the total environmental exposures. The expected final sizes of the major and minor outbreaks are expressed as Z_{major} and Z_{minor} , respectively. Judgment of a human-to-human transmission outbreak is considered to have taken place based on the presence of chains of secondary transmissions (i.e. major epidemic), the presence of at least one secondary case, or simply the random sampling result from the final size.

<https://doi.org/10.1371/journal.pone.0173996.g001>

the presence of a common source of infection (e.g., common food items) among the diagnosed cases and, (ii) the contact history among the cases during the outbreak investigation (Fig 1). It is not surprising that the R_y mirrors the fraction of human-to-human transmission events among the total.

Here, we have built a parsimonious statistical model to describe the data-generating process and estimate the R_y from the dataset. Let q be the proportion of the human-to-human transmission events among the total of outbreak events. Because q is interpreted in several different ways, we built three different models to capture the data-generating process of each interpretation scenario. In each model, the following assumptions were made. (1) We disregard a partly immunized heterogeneous population due to exposures in earlier years and any herd immunity effects are reflected in the estimate of R_y (in a homogeneous manner). (2) Heterogeneous susceptibility and contact are ignored. (3) An individual experiences one exposure event per year at most (i.e. our model does not account for multiple infection events with noroviruses in a year). (4) Ascertainment bias of cases does not affect the fraction of foodborne NVI outbreak events among the total notified events (i.e., the ascertainment process is independent of the reporting process of the major route of transmission). (5) Index cases were environmentally transmitted (e.g. can be attributed to a responsible food item).

Interpretation scenario 1: Reported outbreaks are the result of random sampling of the outbreak events. Let q be the proportion of major outbreaks among the total environmental exposure events. In this scenario, we assume that the observed fraction of human-to-human transmission outbreak events mirrors the fraction of major epidemics in the final size distribution. Assuming that the generation time of human-to-human transmission of norovirus is exponentially distributed, we have $q = 1 - 1/R_y$ given one index case [25], because the probability of extinction is given by $1/R_y$. Suppose that n_y outbreak events were foodborne among a total of $(n_y + m_y)$ outbreaks in year y , assuming that observation of the fraction of foodborne outbreaks is approximated by a binomial sampling process, the likelihood function to estimate R_y is

$$L(R_y; n_y, m_y) = \binom{n_y + m_y}{n_y} \left(\frac{1}{R_y}\right)^{n_y} \left(1 - \frac{1}{R_y}\right)^{m_y} \tag{1}$$

Note that this scenario requires an assumption of $R_y > 1$.

Interpretation scenario 2: Reported outbreaks are the result of random sampling of the patients. In this scenario, we assume that the observed fraction of human-to-human transmission outbreak events mirrors the fraction of cases that were involved in major epidemics. Hereafter, a “major epidemic” refers to a large epidemic that does not stochastically decline to extinction without depletion of susceptibles or successful control, while a “minor outbreak” represents a small scale occurrence of cases that is stochastically terminated. The average outbreak size differs between the foodborne and the human-to-human transmission outbreak events. Assuming that the observed fraction of human-to-human transmission outbreak events among the total events reflects the fraction of cases attributable to human-to-human transmissions, the final sizes of the minor and major outbreaks appear to be useful. Assuming that the generation time is exponentially distributed, and thus, the offspring distribution follows a geometric distribution [26], the final size of a minor outbreak (z_{minor}) is

$$z_{minor} = \left(\frac{R_y}{R_y - 1}\right) \frac{1}{N}, \tag{2}$$

where N is the population size. The derivation of (2) is discussed in [S1 Methods](#). The final size

of a major outbreak (z_{major}) satisfies

$$1 - z_{major} = \exp(-z_{major}R_y) \tag{3}$$

Employing a second-order Taylor series approximation, we get a closed form solution of z_{major} :

$$z_{major} = \frac{2(R_y - 1)}{R_y^2}. \tag{4}$$

Using (2) and (4), the likelihood function to estimate R_y is

$$L(R_y; n_y, m_y) = \binom{n_y + m_y}{n_y} \left(\frac{\frac{1}{R_y} \left(\frac{R_y}{R_y - 1} \right) \frac{1}{N}}{\frac{1}{R_y} \left(\frac{R_y}{R_y - 1} \right) \frac{1}{N} + \left(1 - \frac{1}{R_y} \right) \left(\frac{2(R_y - 1)}{R_y^2} \right)} \right)^{n_y} \left(\frac{\left(1 - \frac{1}{R_y} \right) \left(\frac{2(R_y - 1)}{R_y^2} \right)}{\frac{1}{R_y} \left(\frac{R_y}{R_y - 1} \right) \frac{1}{N} + \left(1 - \frac{1}{R_y} \right) \left(\frac{2(R_y - 1)}{R_y^2} \right)} \right)^{m_y} \tag{5}$$

Interpretation scenario 3: Reported fractions of human-to-human outbreak events reflect the presence of at least one secondary transmission event given an environmental exposure. In this scenario, we consider the possibility that any secondary transmission event from an index case may result in the outbreak notification recorded as arising mainly from human-to-human transmission. In such an instance, the frequency of foodborne infections is the probability of extinction within one generation (i.e., the absence of secondary transmission). Assuming again that the offspring distribution is geometrically distributed, the conditional probability of extinction within one generation is $1/(1 + R_y)$ (see [S1 Methods](#)). Accordingly, the likelihood function to estimate R_y in this scenario is

$$L(R_y; n_y, m_y) = \binom{n_y + m_y}{n_y} \left(\frac{1}{1 + R_y} \right)^{n_y} \left(\frac{R_y}{1 + R_y} \right)^{m_y} \tag{6}$$

Note that this scenario does not impose any assumption of criticality, that is, R_y can be either smaller or greater than the value of 1.

Maximum likelihood estimates of R_y were obtained by minimizing the negative logarithm of the above mentioned likelihood functions. The 95% confidence intervals (CI) were derived from the profile likelihood. To compare goodness-of-fit among all scenarios, we calculated Akaike’s information criterion (AIC). To assess the presence of the time-dependent trend in the R_y estimates, a test of trend for a continuous variable (i.e., Jonckheere-Terpstra test) was employed. We also explored possible correlations between R_y and the proportions of GII and GII.4 among the total of reported virus detections. Pearson’s moment correlation was employed for determining the correlation.

Moreover, to interpret estimates of R_y in relation to an associated time event, we have also decomposed the time-dependent patterns of R_y into one or more interpretable components. That is, the use of real-time polymerase chain reaction (RT-PCR) was initiated in 2001 in Japan and there has been a revision to the 2003 guidelines. For this reason, the use of RT-PCR has become widespread in and after the year 2007, potentially improving laboratory detection around that year. If the proportion of a certain genotype (e.g. GII.4) is caused by that change, the estimated R_y that rests on the observed proportion should be approximated by

$$R_y = \begin{cases} k, & \text{for } y \leq 2006/07 \\ ak, & \text{for } y > 2006/07 \end{cases} \tag{7}$$

where k and a are constant parameters. Alternatively, if there was a growing trend of R_y in addition to the time-dependent introduction of RT-PCR, the estimated R_y should better be described by

$$R_y = \begin{cases} k + bt, & \text{for } y \leq 2006/07 \\ a(k + bt), & \text{for } y > 2006/07 \end{cases} \quad (8)$$

where b measures the liner time-dependent trend. Parameterizing R_y in these simple ways, we compared the AIC between models (7) and (8), identifying better fitted model to the empirically observed data.

Sensitivity analysis

When the reported outbreak events were classified as foodborne or human-to-human transmissions, a substantial number of outbreak events remained cause “unknown”. Such events accounted for 17.1% to 47.5% of all the reported outbreak events from 2000 to 2016. We conducted the following two types of analysis: (i) we discarded those cause “unknown” events by assuming that the missing data occurred completely at random, and (ii) we added a certain fraction of the cause “unknown” data to the foodborne transmissions and the remaining fraction to the human-to-human transmission events. When we allocated the unknown causes to the two known groups, we varied the proportion to be added to the human-to-human transmission events from 0%–100%.

Another sensitivity analysis was carried out to examine the impact of ascertainment bias on our estimates of R_y , because outbreaks with human-to-human transmissions may have been more likely detected and reported, and it is valuable to quantify the impact of such bias. To do so, we varied the count data of outbreaks, m_y human-to-human and n_y foodborne outbreaks, by introducing a factor of ascertainment, c . That is, without ascertainment, c is assumed to be equal to 1. Nevertheless, if human-to-human outbreaks are c times more likely reported, the observed proportion of human-to-human outbreaks should change to $cm_y/\{(n_y+m_y)(1-m_y/(n_y+m_y))+cm_y/(n_y+m_y)\}$ where m_y and n_y are unbiased (actual) numbers of human-to-human and foodborne outbreaks in that adjusting formula. R_y was then estimated using the biased counts of human-to-human and foodborne outbreaks by varying c from 0.5 to 5.0.

Ethical considerations

The present study reanalyzed data that is publicly available in Japan. As such, the datasets used in our study were de-identified and fully anonymized in advance, and the analysis of publicly available data without identity information does not require ethical approval.

Supporting data availability

The present study used publicly available data, and the essential components of the epidemiological data were extractable from a website [24] and also [S1 Table](#) and [S2 Table](#).

Results

[Fig 2](#) shows the time-dependent dynamics of the reported norovirus outbreaks from 2000 to 2016 in Japan. A total of 9,264 outbreaks were reported from 2000/01 to 2015/16 seasons, among which 4,516 were human-to-human, 2,465 were foodborne and the remaining 2,283 were of unknown etiology. Mean and standard deviation (SD) of the reported outbreaks per year were 579 and 280 outbreaks, respectively, and the maximum was 1,388 in 2006/07 and minimum was 220 outbreaks in 2002/03. The minimum after 2006/07 was 373 outbreaks in

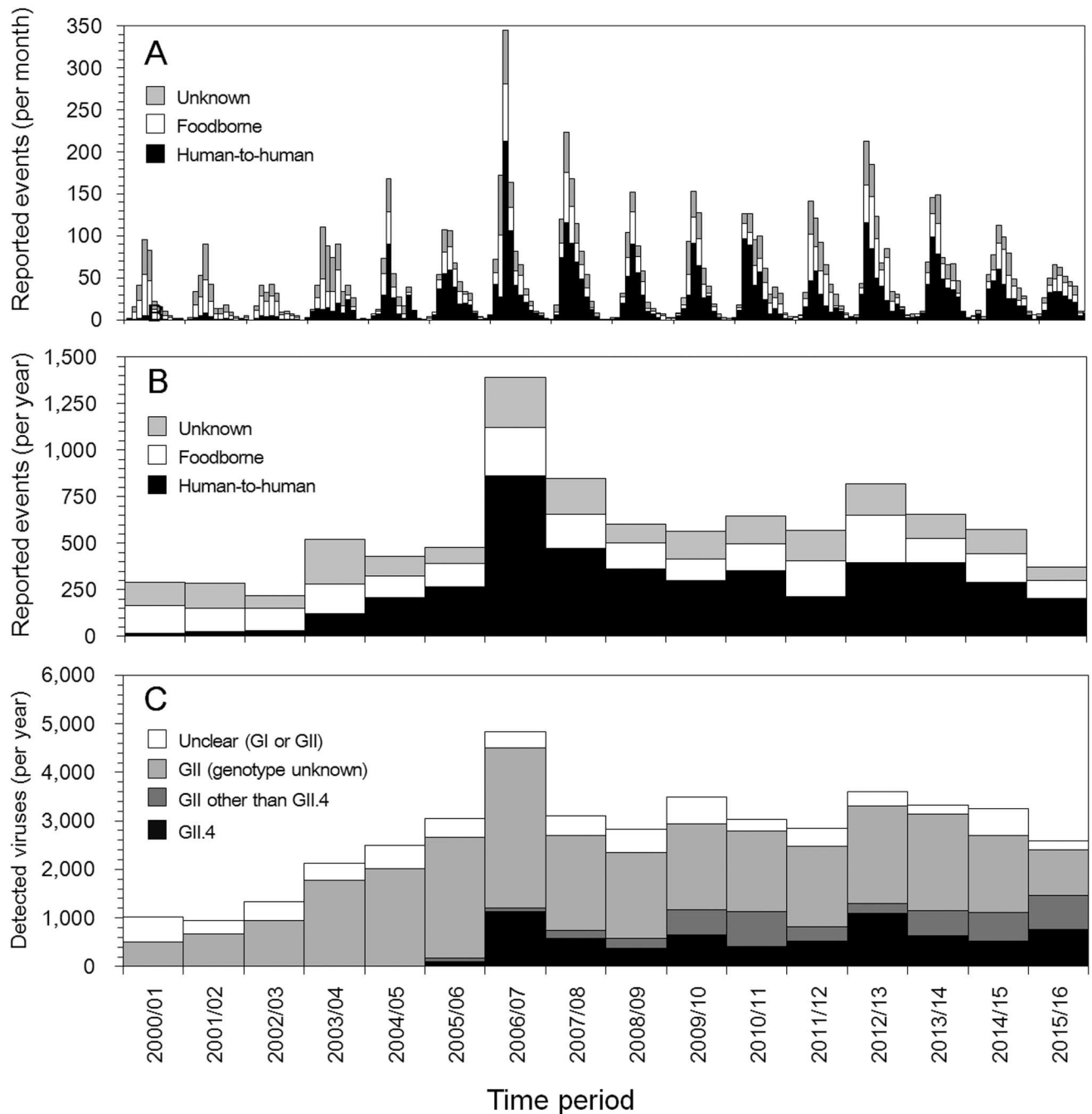


Fig 2. Outbreak events for noroviruses from 2000/01 to 2015/16 in Japan. (A) The reported number of outbreak events from 2000/01 to 2015/16 as monthly data. Every tick mark represents the beginning of September. (B) Outbreak event counts by year. Both (A) and (B) classify the transmission routes as outbreaks caused by human-to-human transmission (black), foodborne infection (white) and unknown causes (gray). (C) The reported number of each detected norovirus genogroup or genotype (detected viruses). White, light gray, dark gray and black colors correspond to unclear (GI or GII), GII (genotype unknown), and GII (except for GII.4 and GII.4), respectively.

<https://doi.org/10.1371/journal.pone.0173996.g002>

2015/16. Every year the peak month for infections was December (Fig 2A). An abrupt increase in the proportion of human-to-human transmission outbreaks is apparent after the 2003/2004 season, the proportion of which peaked in 2006/2007 (Fig 2B). The timing of the increase in

human-to-human transmission outbreaks coincided with the time at which GII, and especially GII.4, increased, and this is especially notable in 2006/07 (Fig 2C).

Fig 3 shows the R_y estimates based on each interpretation scenario. Under scenario 1, the R_y ranged from 1.11 (95% CI; 1.09–1.14) to 4.15 (95% CI; 3.91–4.38) and peaked in 2006/07. Adopting scenario 2, R_y was estimated to range from 1.06 (95% CI; 1.06–1.07) to 1.16 (95% CI; 1.16–1.17). Small increases in the estimate occurred in 2005/06, 2013/14 and 2015/16. Under scenario 3, R_y ranged from 0.11 (95% CI; 0.09–0.14) to 3.29 (95% CI; 3.05–3.52). The estimate exceeded the value of 1 in 2004/05. Employing the trend test, time-dependent increases in estimates of R_y were identified in all three scenarios ($p < 0.05$). Using AIC, scenarios 1 and 3 both appeared to yield the best fit (AIC = 100.3), followed by scenario 2 (AIC = 310.1).

Possible correlations were explored between the estimated R_y value and the GII genogroup and the GII.4 genotype proportions among the total (Fig 4). As we can identify the time-dependent increasing trend in the proportions of the GII genogroup and the GII.4 genotype among the total ($p < 0.05$), a positive correlation was observed between R_y in each scenario and the proportion of GII. A positive correlation was also the case between R_y and the proportion of GII.4, although the correlation in scenario 2 was only marginally significant.

Using parametric models of R_y as described by Eqs (7) and (8), we examined if observed proportions were biased by time-dependent artefact (e.g. caused by the time-dependent change in the laboratory diagnostic method). Considering that R_y could be described only by the step function dividing the overall time-dependent pattern into those before and during 2006/07 season and afterwards, AIC values of scenario 1, 2 and 3 were 542.5, 3774.4 and 8793.5, respectively (Fig 5). If we additionally account for the linearly increasing time trend in

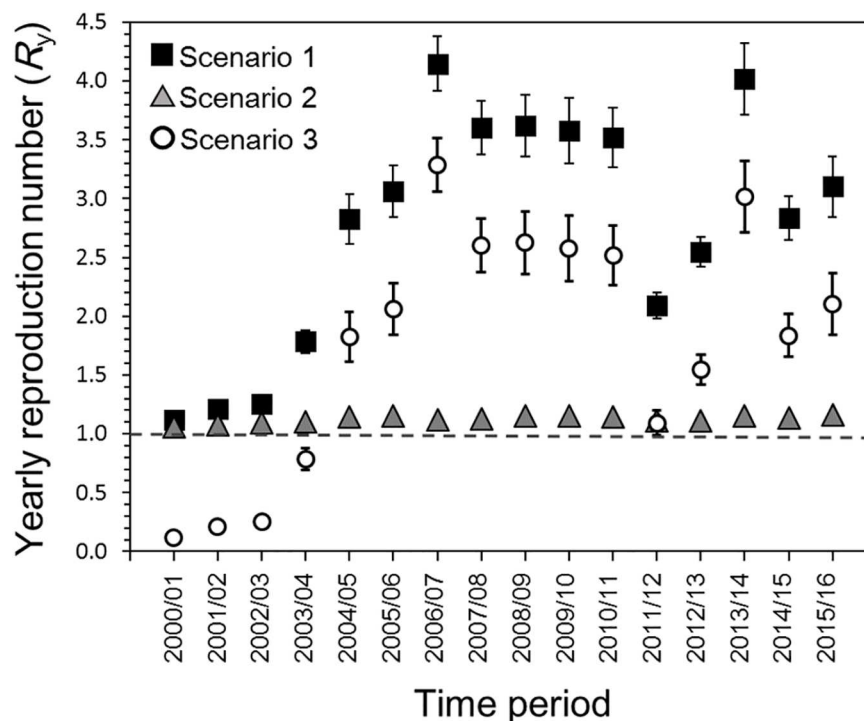


Fig 3. Reproduction number (R_y) from 2000/01 to 2015/16 in Japan. The R_y s of three different interpretation scenarios are presented. The whiskers for each R_y extend to the upper and lower 95% confidence intervals. The horizontal dotted line corresponds to the R_y at unity (i.e., the value of 1), below which a major epidemic event cannot occur.

<https://doi.org/10.1371/journal.pone.0173996.g003>

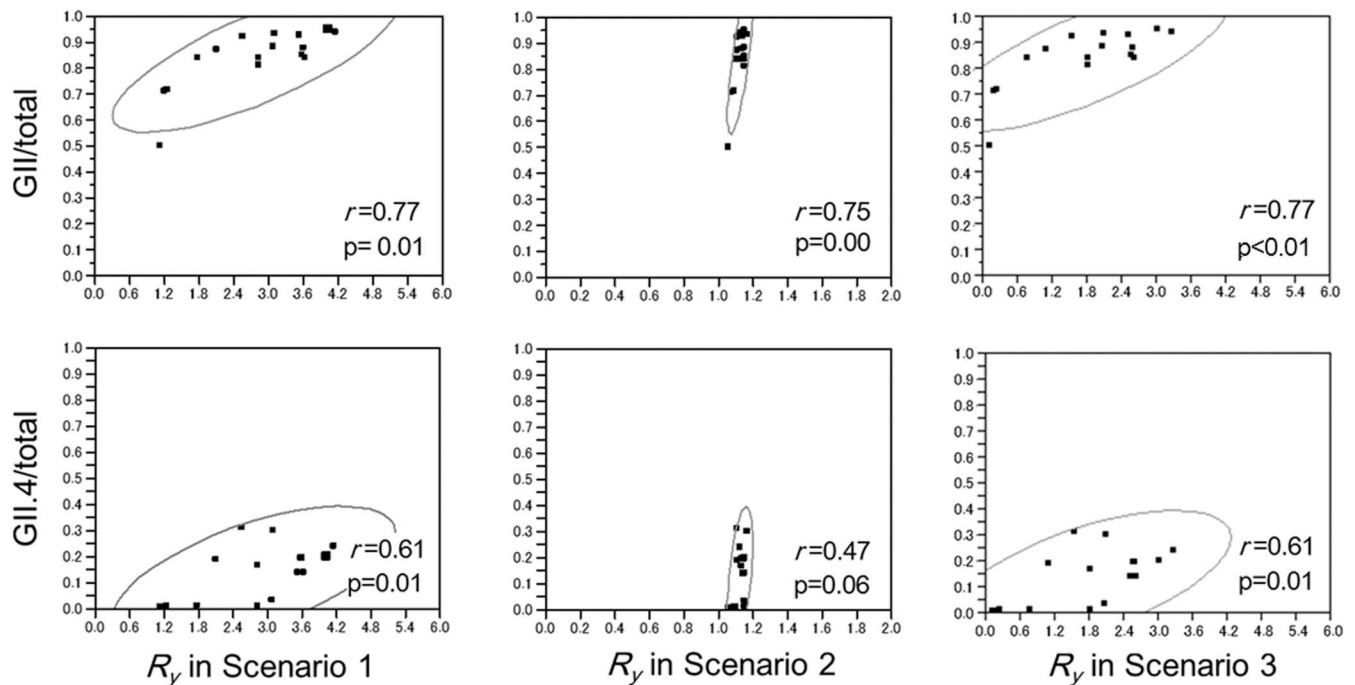


Fig 4. Correlation between the reproduction number (R_y) and the proportion of GII or GII.4 among all the detected noroviruses. The top three panels show the correlations between R_y and the proportion of GII (GII.4) among all the detected noroviruses, as based on scenarios 1, 2 and 3, respectively. Similarly, the bottom three panels examine the correlation between R_y and GII.4.

<https://doi.org/10.1371/journal.pone.0173996.g004>

R_y , the AIC was calculated at 480.7, 3728.9 and 8782.8, respectively, for scenarios 1, 2 and 3. Namely, the model fit was always better for R_y that combines step function (separating before and after 2007) with linear trend than R_y with step function only.

As part of the sensitivity analysis, Fig 6 explores the impact of the cause “unknown” category data on our estimate of R_y . As one can imagine, the estimate of R_y would be greater than R_y when we discarded unknown data. The scenario-related estimates order (Fig 6) and the time-dependent patterns of R_y remained unchanged, even when we assigned a certain fraction of the “unknown” causes to the human-to-human transmission events.

Using the empirical data in 2015/16 season, Fig 7 explores the impact of ascertainment bias of human-to-human outbreaks on the estimates of R_y . Overall, R_y appears to be overestimated if the ascertainment bias is large. Among three scenarios, scenarios 1 and 3 yielded R_y that are more sensitive to the increase in ascertainment factor c than scenario 2. In scenarios 1 and 3, increase of c by 1.0 led to an increase in R_y by 1.05.

Discussion

The present study estimated the R_y for NVIs from 2000/01 to 2015/16 by assessing their time-dependent patterns and exploring the possible association between R_y and the norovirus genotype. Since the data generating process cannot fully ascertain the mode of transmission, the corresponding uncertainties were addressed by employing three possible scenarios to interpret the data. An apparent increase in R_y (with small fluctuations) was observed under all the scenarios, along with positive correlations between R_y and the proportions of GII and GII.4 among the total number of viruses identified. The increasing pattern of R_y was better captured by a simple model with a linear trend plus the step function effect of introducing RT-PCR to laboratory diagnosis than a model with step function alone, implying that the increase in R_y

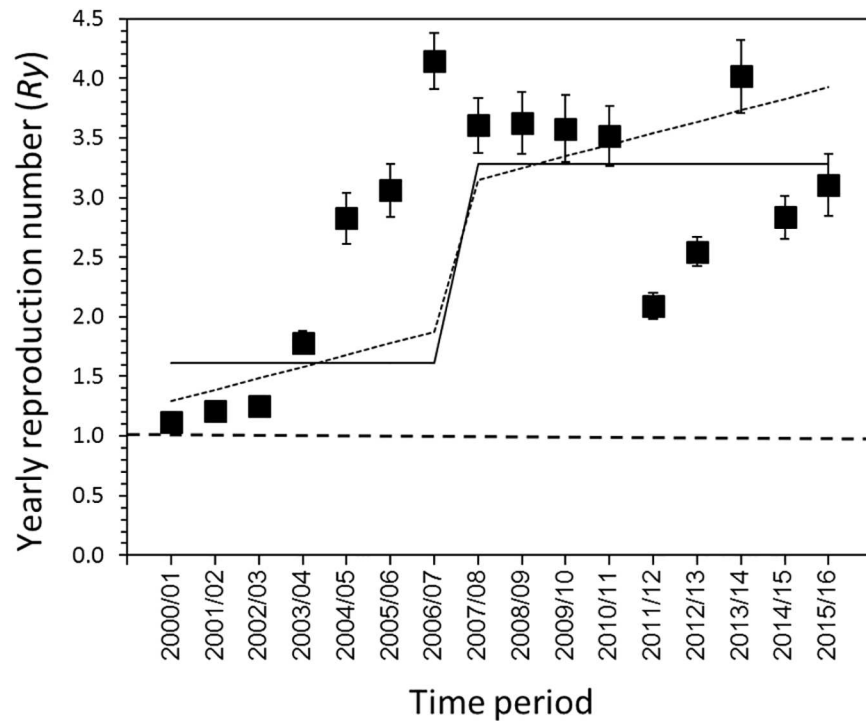


Fig 5. Parametric trend of the yearly reproduction number of noroviruses. Black squares represent the yearly reproduction number from 2000/01 to 2015/16 under scenario 1. The best fit models for the estimated yearly reproduction number with step function alone (solid line) and with step function plus a linear trend (dotted line) are shown. Akaike Information Criterion (AIC) values for solid and dotted lines were 542.5 and 480.7, indicating that the model with linear trend fitted to the data better.

<https://doi.org/10.1371/journal.pone.0173996.g005>

was not only caused by the widespread introduction of RT-PCR at one time. To the best of our knowledge, the present study is the first to successfully describe the time-dependent change in NVI transmissibility using empirically observed epidemiological data.

Growth in norovirus transmissibility over time implies that the infection has become a more human-to-human transmissible disease. This finding was achieved by exploring the long-time series data from 2000 to 2016, and the time-dependent change indicates that NVI in the early period may differ somewhat from the time-dependent changes seen in recent infections. For instance, foodborne cases of NVI arising from bivalve molluscan shellfish (raw oysters) have decreased relatively as a fraction among the total number of infected individuals in Japan [27], while noroviruses better fitted to humans have selectively caused the human-to-human transmissions, thereby changing the epidemiology of this pathogen. Our order of 3 to 4 estimates for R_y in recent years is consistent with the assumed values of the basic reproduction number R_0 in a previously published study [20]. The comparability of R_y to R_0 and its time dependency endorse its utility to monitor strain specific transmissibility over time in the future.

Moreover, the increased time-dependent pattern of the R_y appears to have coincided with an increase in the proportions of GII and GII.4 genotypes (especially in 2006 when the GII.4 2006b strain was identified [28–30]), implying that the viral genogroup or genotype is responsible for characterizing the increased transmissibility. The positive association between transmissibility and GII is consistent with the findings of a published systematic review [31]. While GII.4 was not linked to transmissibility in the same report [31], another study showed that GII.4 is likely to be involved in human-to-human transmissions [8, 32, 33]. GII.4 is known to

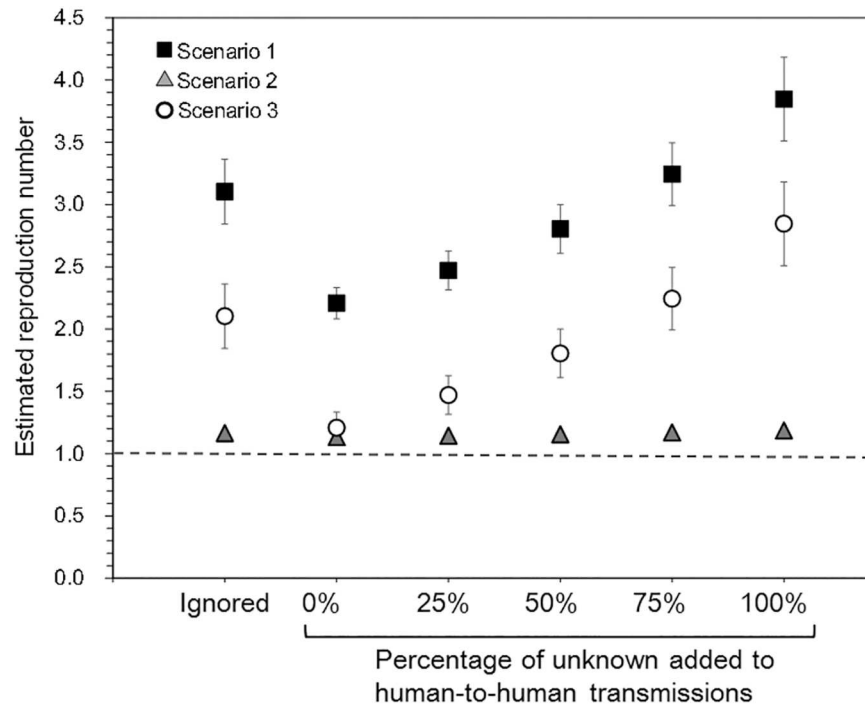


Fig 6. Estimates of the reproduction number of noroviruses in relation to missing data. Black squares, gray triangles, and white circles correspond to the estimated reproduction number in 2015/2016 as based on interpretation scenarios 1, 2 and 3, respectively. “Ignored” results represent the data obtained after discarding the “unknown” cause category datasets. By allocating a certain fraction of “unknown” causes to the human-to-human transmission outbreaks, the relative changes in the estimated reproduction number could be explored. Whiskers extend to the upper and lower 95% confidence intervals.

<https://doi.org/10.1371/journal.pone.0173996.g006>

bind to histo-blood group antigens more than other genotypes, and the diversity of its P2 domain implies that this genotype may experience more rapid molecular evolution than other genotypes [6,7,34]. Two studies have indicated that while transmission of GII.4 2006b has decreased [35, 36], it has been replaced by later genotypes such as GII.4 New Orleans 2009 and Sydney 2012, and new GII.4 variants may likely replace the earlier GII.4 genotype because substantial herd immunity to earlier genotypes probably exists [37–39]. The present study has contributed to improved understanding of the evolutionary dynamics of NVI by demonstrating that the time-dependent infection dynamics can be partly explored by estimating the R_y and examining the epidemiological and virological data concurrently. Continued monitoring of the major routes of transmission for NVIs is of the utmost importance, while better understanding of the mechanisms underlying norovirus evolution is eagerly awaited.

Scenarios 1 and 3, which rely on the random sampling of outbreaks or the random distinction of sporadic cases without any offspring from other series of cases, appeared to fit to observed data equally well in statistical sense. While our estimates of R_y in recent years were consistent with those published in the USA [20], it must be noted that the use of proportional data for human-to-human transmission among the total number of reported cases would inherently face ascertainment and reporting biases. That is, compared with sporadic cases, cases that are attributable to human-to-human transmission are more likely to be diagnosed and are also more likely to be notified, as has been noted in modeling studies that have focused on the human-animal interface [40,41]. In fact, one prefecture arbitrarily but descriptively defined a human-to-human transmission outbreak as an outbreak in various facilities where

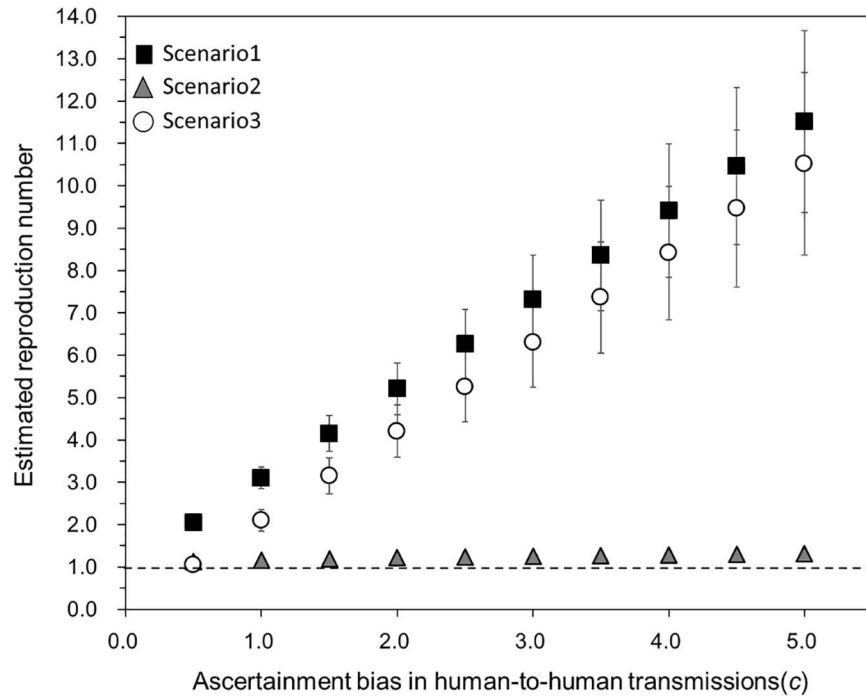


Fig 7. Estimates of the yearly reproduction number of noroviruses in relation to the ascertainment of human-to-human transmission outbreaks. Black squares, gray triangles, and white circles represent the yearly reproduction numbers in 2015/16 season in our interpretation scenarios 1, 2, and 3, respectively. Whiskers extend to the upper and lower 95% confidence intervals. We set the factor of ascertainment c , i.e. the relative frequency of human-to-human outbreaks to be reported compared with foodborne outbreaks, as a coefficient that represents the ascertainment bias. Observed frequency data in 2015/16 season were assumed as unbiased and $c = 1$.

<https://doi.org/10.1371/journal.pone.0173996.g007>

more than 10 cases a day or more than half of all residents developed severe symptoms [37]. We have thus conducted a sensitivity analysis, quantifying the impact of ascertainment bias on R_y estimates (Fig 7). Our R_y may have been potentially overestimated, influencing any policy relevant decisions derived from the magnitude of R_y (i.e. absolute value of R_y), but the relative change in R_y over time is likely to be unaffected by ascertainment and reporting biases.

As mentioned in Methods, foodborne outbreaks have been frequently differentiated from human-to-human outbreaks due to the need to link the outbreak to a source of infection (e.g. a shared food item) among cases. Cases in the foodborne outbreaks are likely tested, because public health needs to verify that a particular foodborne outbreak was caused by noroviruses, and in that sense, the environmental exposure event can be regarded as well identified in the empirical observation. Rather than the validity of environmental exposures, we have not been sure as to the data generating process of “unknown” category data. We have thus conducted a sensitivity analysis in Fig 6 and the time-dependent change in R_y remained similar to Fig 3.

It should be noted that we employed three possible scenarios to address the uncertainty surrounding the interpretation of the mode of transmission. Scenario 1 assumed random sampling of all outbreaks, while it is likely that large epidemics may be easily reported compared with small outbreaks. Scenario 2 assumed random sampling of cases, while in reality more severe cases are likely reported than mild ones. Scenario 3 assumed that the presence of one secondary transmission as human-to-human outbreak, but it does not capture the tendency that the presence of two or more secondary transmissions lead to greater ascertainment than

only one secondary transmission. While it is evident that each did not fully capture the observed patterns of the data, we were not able to exclude any scenario as long as each could theoretically lead to the observation of empirical data in Japan, and thus, we believe that employing three models and comparing estimates were the best practice to address the uncertainty surrounding the mode of transmission.

Four other limitations should be noted. First, technical changes in laboratory methods, reporting rules for NVIs and any other time events were not explicitly taken into account for the 16 year study period. As mentioned in Methods, the widespread use of RT-PCR in and after 2007 was one important factor that could have contributed to increased proportion of reported GII.4 cases from around 2006. We have thus modeled R_y using a step function with and without linear trend, and it was shown that the step function with linear trend was favored. Nevertheless, it should be remembered that the step function qualitatively captured the increase in R_y in any case, and we cannot exclude the possibility that RT-PCR had an impact on R_y . At least, we have shown that there was a linearly increasing trend that has not been explained by the introduction of single laboratory technique. Other reasons, e.g. disproportionately improved surveillance of human-to-human transmission especially at healthcare facility, cannot be excluded, and indeed, we have shown that an improved ascertainment would lead to an increase in R_y . Nevertheless, we do not have any explicit indication of such disproportionate improvement as compared to published studies on the relationship between the transmissibility and GII or GII.4 [8,31–33], and clarification on such alternative explanations would be the subject for future studies.

As the second limitation, by grouping all noroviruses in a single epidemic season into one category our analysis rests on an estimation of the R_y . While the present study explored possible correlations between R_y and the proportion of certain genogroups and genotypes, an aggregated analysis had to be adopted because of the unlinked nature of the empirical data between each single outbreak event and the specific genogrouping and genotyping. Third, our estimation did not explicitly account for herd immunity. Only by assuming that R_y reflects both the transmissibility and susceptible fraction of the population could a homogeneously defined estimate be derived from the data. The ignorance did not allow us to decompose R_y into the actual transmissibility of a particular year y and a possible partially immune fraction in the same year. Only by collecting more structured data (e.g. an age-structured long-time series combining the transmission route and virus genogrouping), would we be able to explicitly address the issue of herd immunity, and R_y estimates presented here might have been potentially overestimated. Fourth, we did not account for any index cases arising from (other) unrecognized chains of transmission. Given that such index cases exist, the estimate of R_y could be greater than what have been estimated in this article.

In summary, a time-dependent increase in NVI transmissibility was identified in Japan, indicating that noroviruses better fitted to humans have selectively caused the human-to-human transmissions. This increase in norovirus transmissibility has coincided with an increase in the proportions of GII and GII.4 among all the viruses detected, implying a possible association between transmissibility and this genotype and genogroup. Evolutionary dynamics are perhaps responsible for the changing epidemiology of noroviruses, thereby elevating the relative importance of human-to-human transmission as compared with classical foodborne infections. We believe that our findings will greatly enhance current understanding of the ecoepidemiology of noroviruses. However, we cannot help but stress the importance of continued monitoring of the fraction of human-to-human transmission outbreaks among the total outbreak events to better capture the evolutionary nature of norovirus and investigate effective control strategies against it.

Supporting information

S1 Methods.

(DOCX)

S1 Table.

(XLSX)

S2 Table.

(XLSX)

Acknowledgments

We acknowledge Kenji Mizumoto, Ryo Kinoshita, Kyeongah Nah and Yusuke Asai for their valuable technical advice and discussion over the design of the statistical models presented herein.

Author Contributions

Conceptualization: HN.

Data curation: HN RM.

Formal analysis: HN RM.

Funding acquisition: HN RM.

Investigation: HN RM.

Methodology: HN.

Project administration: HN.

Resources: HN RM.

Software: HN RM FM.

Supervision: HN.

Validation: HN FM.

Visualization: HN RM.

Writing – original draft: HN RM.

Writing – review & editing: HN RM FM.

References

1. Lopman BA, Steele D, Kirkwood CD. The vast and varied global burden of norovirus: Prospects for prevention and control. *PLoS Med.* 2016; 13(4): e1001999. <https://doi.org/10.1371/journal.pmed.1001999> PMID: 27115709
2. Hall AJ, Lopman BA, Payne DC, Patel MM, Gastañaduy PA, Vinjé J, et al. Norovirus disease in the United States. *Emerg Infect Dis.* 2013; 19: 1198–1205. <https://doi.org/10.3201/eid1908.130465> PMID: 23876403
3. Teunis PF, Moe CL, Liu P, Miller SE, Lindesmith L, Baric RS, et al. Norwalk virus: how infectious is it? *J Med Virol.* 2008; 80: 1468–1476. <https://doi.org/10.1002/jmv.21237> PMID: 18551613
4. Lopman B, Gastañaduy P, Park GW, Hall AJ, Parashar UD, Vinjé J. Environmental transmission of norovirus gastroenteritis. *Curr Opin Virol.* 2012; 2: 96–102. <https://doi.org/10.1016/j.coviro.2011.11.005> PMID: 22440972

5. Hay AJ, Gregory V, Douglas AR, Lin YP. The evolution of human influenza viruses. *Philos Trans R Soc Lond B Biol Sci.* 2001; 356(1416): 1861–1870. <https://doi.org/10.1098/rstb.2001.0999> PMID: 11779385
6. Bull RA, White PA. Mechanisms of GII.4 norovirus evolution. *Trends Microbiol.* 2011; 19: 233–240. <https://doi.org/10.1016/j.tim.2011.01.002> PMID: 21310617
7. White PA. Evolution of norovirus. *Clin Microbiol Infect.* 2014; 20: 741–745. <https://doi.org/10.1111/1469-0691.12746> PMID: 24980204
8. Lindesmith LC, Beltramello M, Donaldson EF, Corti D, Swanstrom J, Debbink K, et al. Immunogenetic mechanisms driving norovirus GII.4 antigenic variation. *PLoS Pathog.* 2012; 8(5): e1002705. <https://doi.org/10.1371/journal.ppat.1002705> PMID: 22615565
9. Siebenga JJ, Vennema H, Renckens B, de Bruin E, van der Veer B, Siezen RJ, et al. Epochal evolution of GII.4 norovirus capsid proteins from 1995 to 2006. *J Virol.* 2007; 81: 9932–9941. <https://doi.org/10.1128/JVI.00674-07> PMID: 17609280
10. White PA, Hansman GS, Li A, Dable J, Isaacs M, Ferson M, et al. Norwalk-like virus 95/96-US strain is a major cause of gastroenteritis outbreaks in Australia. *J Med Virol.* 2002; 68: 113–118. <https://doi.org/10.1002/jmv.10177> PMID: 12210438
11. Lopman B, Vennema H, Kohil E, Pothier P, Sanchez A, Negredo A, et al. Increase in viral gastroenteritis outbreaks in Europe and epidemic spread of new norovirus variant. *Lancet.* 2004; 363: 682–688. [https://doi.org/10.1016/S0140-6736\(04\)15641-9](https://doi.org/10.1016/S0140-6736(04)15641-9) PMID: 15001325
12. Bull RA, Tu ET, McIver CJ, Rawlinson WD, White PA. Emergence of a new norovirus genotype GII.4 variant associated with global outbreaks of gastroenteritis. *J Clin Microbiol.* 2006; 44: 327–333. <https://doi.org/10.1128/JCM.44.2.327-333.2006> PMID: 16455879
13. Eden JS, Bull RA, Tu E, McIver CJ, Lyon MJ, Marshall JA, et al. Norovirus GII.4 variant 2006b caused epidemics of acute gastroenteritis in Australia during 2007 and 2008. *J Clin Virol.* 2010; 49: 265–271. <https://doi.org/10.1016/j.jcv.2010.09.001> PMID: 20888289
14. Yen C, Wikswo ME, Lopman BA, Vinje J, Parashar UD, Hall AJ. Impact of an emergent norovirus variant in 2009 on norovirus outbreak activity in the United States. *Clin Infect Dis.* 2011; 53: 568–571. <https://doi.org/10.1093/cid/cir478> PMID: 21832262
15. Eden JS, Tanaka MM, Boni MF, Rawlinson WD, White PA. Recombination within the pandemic norovirus GII.4 lineage. *J Virol.* 2013; 87: 6270–6282. <https://doi.org/10.1128/JVI.03464-12> PMID: 23536665
16. Matsushima Y, Ishikawa M, Shimizu T, Komane A, Kasuo S, Shinohara M, et al. Genetic analyses of GII.17 norovirus strains in diarrheal disease outbreaks from December 2014 to March 2015 in Japan reveal a novel polymerase sequence and amino acid substitutions in the capsid region. *Euro Surveill.* 2015; 20(26). pii: 2117
17. Gao Z, Liu B, Huo D, Yan H, Jia L, Du Y, et al. Increased norovirus activity was associated with a novel norovirus GII.17 variant in Beijing, China during winter 2014–2015. *BMC Infect Dis.* 2015; 15: 574. <https://doi.org/10.1186/s12879-015-1315-z> PMID: 26678989
18. Chan MC, Lee N, Hung TN, Kwok K, Cheung K, Tin EK, et al. Rapid emergence and predominance of a broadly recognizing and fast-evolving norovirus GII.17 variant in late 2014. *Nat Commun.* 2015; 6: 10061. <https://doi.org/10.1038/ncomms10061> PMID: 26625712
19. Simmons K, Gambhir M, Leon J, Lopman B. Duration of immunity to norovirus gastroenteritis. *Emerg Infect Dis.* 2013; 19(8): 1260–1267. <https://doi.org/10.3201/eid1908.130472> PMID: 23876612
20. Lopman B, Simmons K, Gambhir M, Vinje J, Parashar U. Epidemiologic implications of asymptomatic reinfection: a mathematical modeling study of norovirus. *Am J Epidemiol.* 2014; 179(4): 507–512. <https://doi.org/10.1093/aje/kwt287> PMID: 24305574
21. Vanderpas J, Louis J, Reynders M, Mascart G, Vandenberg O. Mathematical model for the control of nosocomial norovirus. *J Hosp Infect.* 2009; 71(3): 214–222. <https://doi.org/10.1016/j.jhin.2008.11.024> PMID: 19162373
22. Sukhrie FH, Teunis P, Vennema H, Copra C, Thijs Beersma MF, Bogerman J, et al. Nosocomial transmission of norovirus is mainly caused by symptomatic cases. *Clin Infect Dis.* 2012; 4: 931–937.
23. Heijne JC, Teunis P, Morroy G, Wijkman C, Oostveen S, Duizer E, et al. Enhanced hygiene measures and norovirus transmission during an outbreak. *Emerg Infect Dis.* 2009; 15: 24–30. <https://doi.org/10.3201/eid1501.080299> PMID: 19116045
24. Infectious Disease Surveillance Center, National Institute of Infectious Diseases. Epidemiology of Norovirus in Japan, 2010/11–2013/14 seasons. *Infectious Agents Surveillance Report.* 2014; 35: 161–163.
25. Harris TE. *The theory of branching processes.* Berlin: Springer; 1963.
26. Nishiura H, Yan P, Sleeman CK, Mode CJ. Estimating the transmission potential of supercritical processes based on the final size distribution of minor outbreaks. *J Theor Biol.* 2012; 294: 48–55. <https://doi.org/10.1016/j.jtbi.2011.10.039> PMID: 22079419

27. Takeda N. Virological information of norovirus outbreaks and transmission. Ministry of Health, Labour and Welfare, Tokyo, 2016 (in Japanese). Available from: <http://www.mhlw.go.jp/shingi/2007/08/dl/s0817-6e.pdf>
28. Kumazaki M, Usuku S. Genetic analysis of norovirus GII.4 variant strains detected in outbreaks of gastroenteritis in Yokohama, Japan, from the 2006–2007 to the 2013–2014 seasons. *PLoS One*. 2015;
29. Chan-It W, Thongprachum A, Khamrin P, Kobayashi M, Okitsu S, Mizuguchi M, et al. Emergence of a new norovirus GII.6 variant in Japan, 2008–2009. *J Med Virol*. 2012; 84: 1089–1096. <https://doi.org/10.1002/jmv.23309> PMID: 22585727
30. Dey SK, Phan TG, Mizuguchi M, Okitsu S, Ushijima H. Genetic diversity and emergence of norovirus GII.4-2006b in Japan during 2006–2007. *Clin Lab*. 2011; 57(3–4): 193–199. PMID: 21500726
31. Bitler EJ, Matthews JE, Dickey BW, Eisenberg JNS, Leon JS. Norovirus outbreaks: a systematic review of commonly implicated transmission routes and vehicles. *Epidemiol Infect*. 2013; 141: 1563–1571. <https://doi.org/10.1017/S095026881300006X> PMID: 23433247
32. Vega E, Barclay L, Gregoricus N, Shirley SH, Lee D, Vinje J. Genotypic and epidemiologic trends of norovirus outbreaks in the United States, 2009 to 2013. *J Clin Microbiol*. 2014; 52: 147–155. <https://doi.org/10.1128/JCM.02680-13> PMID: 24172151
33. Verhoef L, Hewitt J, Barclay L, Ahmed SM, Lake R, Hall AJ, et al. Norovirus genotype profiles associated with foodborne transmission. *Emerg Infect Dis*. 2015; 21: 592–599. <https://doi.org/10.3201/eid2104.141073> PMID: 25811368
34. Lindesmith LC, Donaldson EF, Lobue AD, Cannon JL, Zheng DP, Vinje J et al. Mechanisms of GII.4 norovirus persistence in human populations. *PLoS Med*. 2008; 5(2):e31. <https://doi.org/10.1371/journal.pmed.0050031> PMID: 18271619
35. Vega E BL, Gregoricus N, Williams K, Lee D, Vinje J. Novel surveillance network for norovirus gastroenteritis outbreaks, United States. *Emerg Infect Dis*. 2011; 17: 1389–1395. <https://doi.org/10.3201/eid1708.101837> PMID: 21801614
36. Lindesmith LC, Costantini V, Swanstrom J, Debbink K, Donaldson EF, Vinje J et al. Emergence of a norovirus GII.4 strain correlates with changes in evolving blockade epitopes. *J Virol*. 2013; 87: 2803–2813. <https://doi.org/10.1128/JVI.03106-12> PMID: 23269783
37. Sakon N, Yamazaki K, Nakata K, Kanbayashi D, Yoda T, Mantani M, et al. Impact of genotype-specific herd immunity on the circulatory dynamism of norovirus: a 10-year longitudinal study of viral acute gastroenteritis. *J Infect Dis*. 2015; 211: 879–888. <https://doi.org/10.1093/infdis/jiu496> PMID: 25210139
38. de Graaf M, van Beek J, Vennema H, Podkolzin AT, Hewitt J, Bucardo F, et al. Emergence of a novel GII.17 norovirus—End of the GII.4 era? *Euro Surveill*. 2015; 20: pii: 21178.
39. Debbink K, Lindesmith LC, Donaldson EF, Costantini V, Beltramello M, Corti D, et al. Emergence of new pandemic GII.4 Sydney norovirus strain correlates with escape from herd immunity. *J Infect Dis*. 2013; 208: 1877–1887. <https://doi.org/10.1093/infdis/jit370> PMID: 23908476
40. Cauchemez S, Epperson S, Biggerstaff M, Swerdlow D, Finelli L, Ferguson NM. Using routine surveillance data to estimate the epidemic potential of emerging zoonoses: application to the emergence of US swine origin influenza A H3N2v virus. *PLoS Med*. 2013; 10(3): e1001399. <https://doi.org/10.1371/journal.pmed.1001399> PMID: 23472057
41. Nishiura H, Mizumoto K, Ejima K. How to interpret the transmissibility of novel influenza A(H7N9): an analysis of initial epidemiological data of human cases from China. *Theor Biol Med Model*. 2013; 10: 30. <https://doi.org/10.1186/1742-4682-10-30> PMID: 23642092