



RESEARCH ARTICLE

REVISED Quantification and determinants of the amount of respiratory syncytial virus (RSV) shed using real time PCR data from a longitudinal household study [version 2; referees: 2 approved, 2 approved with reservations]

Miriam Wathuo¹, Graham F. Medley ², D. James Nokes^{1,3}, Patrick K. Munywoki^{1,4}

¹KEMRI - Wellcome Trust Research Programme, Centre for Geographic Medicine Research – Coast, Kilifi, Kenya

²Department of Global Health and Development, London School of Hygiene and Tropical Medicine, London, UK

³School of Life Sciences and SBIDER, University of Warwick, Coventry, UK

⁴Department of Nursing Sciences, Pwani University, Kilifi, Kenya

v2 First published: 14 Dec 2016, 1:27 (doi: [10.12688/wellcomeopenres.10284.1](https://doi.org/10.12688/wellcomeopenres.10284.1))
Latest published: 13 Mar 2017, 1:27 (doi: [10.12688/wellcomeopenres.10284.2](https://doi.org/10.12688/wellcomeopenres.10284.2))

Abstract





Background: A better understanding of respiratory syncytial virus (RSV) epidemiology requires realistic estimates of RSV shedding patterns, quantities shed, and identification of the related underlying factors.

Methods: RSV infection data arise from a cohort study of 47 households with 493 occupants, in coastal Kenya, during the 2009/2010 RSV season. Nasopharyngeal swabs were taken every 3 to 4 days and screened for RSV using a real time polymerase chain reaction (PCR) assay. The amount of virus shed was quantified by calculating the ‘area under the curve’ using the trapezoidal rule applied to rescaled PCR cycle threshold output. Multivariable linear regression was used to identify correlates of amount of virus shed.

Results: The median quantity of virus shed per infection episode was 29.4 (95% CI: 15.2, 54.2) log₁₀ ribonucleic acid (RNA) copies * days. Young age (<1 year), presence of upper respiratory symptoms, intra-household acquisition of infection, an individual’s first infection episode in the RSV season, and having a co-infection of RSV group A and B were associated with increased amount of virus shed.

Conclusions: The findings provide insight into which groups of individuals have higher potential for transmission, information which may be useful in designing RSV prevention strategies.

Open Peer Review

Referee Status:    

Invited Referees

1 2 3 4

REVISED

version 2

published
13 Mar 2017

version 1

published
14 Dec 2016

   
report report report report

- 1 **Benjamin J. Cowling**, The University of Hong Kong (HKU) China
- 2 **H. Rogier van Doorn**, National Hospital of Tropical Diseases Vietnam, **Lien Anh Ha Do**, Murdoch Children’s Research Institute Australia
- 3 **Peter F. Wright**, Dartmouth College USA
- 4 **John DeVincenzo**, University of Tennessee College of Medicine USA

Discuss this article

Comments (0)

Corresponding author: D. James Nokes (jnokes@kemri-wellcome.org)

How to cite this article: Wathuo M, Medley GF, Nokes DJ and Munywoki PK. **Quantification and determinants of the amount of respiratory syncytial virus (RSV) shed using real time PCR data from a longitudinal household study [version 2; referees: 2 approved, 2 approved with reservations]** Wellcome Open Research 2017, 1:27 (doi: [10.12688/wellcomeopenres.10284.2](https://doi.org/10.12688/wellcomeopenres.10284.2))

Copyright: © 2017 Wathuo M *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Grant information: This work was supported by the Wellcome Trust [102975], [090853].

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: No competing interests were disclosed.

First published: 14 Dec 2016, 1:27 (doi: [10.12688/wellcomeopenres.10284.1](https://doi.org/10.12688/wellcomeopenres.10284.1))

REVISED Amendments from Version 1

Differences between versions 1 and 2 of the article 'Quantification and determinants of the amount of respiratory syncytial virus (RSV) shed using real time PCR data from a longitudinal household study'.

Version two of this article;

1. Emphasizes that log transformed values are used for the viral densities and AUC values and why they are used. The reasons for this are given as being that the log transformed values have a smaller variance and they are less skewed. They are also more suitable for the linear regression analysis.
2. States that the reason sex is adjusted for in the final linear regression model despite not being statistically significant is because it was chosen *a priori* as an important factor for RSV infections.
3. Includes additional information in [Table 1](#) on the total number of individuals per category among all the individuals who took part in the study.
4. Adds a possible explanation as to why individuals who were living with smokers appeared to shed less virus than those who were not living with smokers. Some of the reasons stated are that there were very few individuals in the study who lived with smokers, and we do not have information on the contact patterns of individuals with the smokers.
5. Adds that we tried two different approaches in calculating the AUC. The first involved taking logs of the viral densities and calculating the AUC. The second involved calculating the AUC using untransformed viral densities, then taking logarithms of the AUC. These two approaches did not give substantially different results.

See referee reports

Introduction

Respiratory syncytial virus (RSV) is the most common viral cause of severe lower respiratory tract infection (LRTI) among infants and children under 5 years old worldwide, with the greatest burden occurring in developing countries¹. Most children experience an RSV infection episode by the age of two years, with peak rates of infection occurring in the first year of life². Re-infections occur throughout life³ as RSV infections provide incomplete or waning immunity. There are no licensed RSV vaccines but there is heightened activity in vaccine development⁴⁻⁸.

Vaccine delivery strategies should be informed by detailed understanding of RSV transmission dynamics. To reduce the circulation of RSV, identifying which groups are responsible for majority of the infections is critical. Transmission potential can be estimated by a combination of mixing patterns⁹ and virus shedding (viral density and duration). Human experimental infection with RSV reported that volunteers inoculated with a higher dose of virus ($4.7 \log_{10}$ tissue culture infectious dose (TCID)₅₀ of RSV A2) were more likely to be infected than those given low doses of the inoculum ($3.7 \log_{10}$ TCID₅₀ of RSV A2)¹⁰, suggesting that individuals who shed more virus are likely to be more infectious. Determining quantities shed and the related underlying factors, will help in predicting the spread of infection in a population and identify key groups to target for infection control. Previous studies have used

the duration of shedding to identify individuals with the greatest potential for infection spread¹¹, based on the assumption that they have a higher number of infectious contacts. Several studies report on the duration of RSV shedding with mean values ranging from 4.5 to 11.2 days^{12,13}. In addition to variation in assay type, determinants included history of RSV infection¹², age, infection severity, detection of other viruses before and during the RSV infection, and presence of concurrent RSV infections in the same household¹³. These studies and others^{14,15} do not account for the temporal changes in quantity of virus shed. Experimental RSV infection studies indicate that an individual infection episode begins with low viral shedding, which rises with time as the virus continues to replicate within epithelial cells and finally declines as the infection clears¹⁶.

The current analysis aims to include temporal changes in viral shedding in estimating the total amount of virus shed during an individual RSV infection episode. Relative to the duration of shedding, this may provide an improved correlate of infectiousness and help in identifying the key factors influencing RSV shedding patterns. Such data are informative in formulating vaccine product profiles and designing prevention strategies for RSV.

Methods

Data

The RSV infection data arise from an intensively followed cohort of 47 households with 493 occupants in rural coastal Kenya. The details of the study have been described elsewhere^{13,17,18}. In summary, throughout an RSV season spanning 26 weeks (December 2009–June 2010), nasopharyngeal swabs (NPS) were collected by trained field assistants every 3–4 days, irrespective of symptoms, from 47 RSV naïve infants and their household members. Households were selected through the Kilifi Health and Demographic Surveillance System (KHDSS) and local community health workers and were considered eligible if they had a child born after 1st April 2009. The infants were assumed to be RSV naïve because they were born after the 2008/2009 RSV season. A total of 16928 NPS collections were tested for RSV (groups A and B) and other prevalent respiratory viruses (adenoviruses, rhinoviruses and human coronaviruses (NL63, 229E and OC43)) using multiplex real time polymerase chain reaction (PCR) assay as previously described¹⁹. As PCR cycles proceed the quantity of product detected rises and at some point passes through a user assigned threshold for detection (set at the point of observable exponential rise in product), known as the cycle threshold (Ct). The earlier (lower) the Ct value the higher the starting concentration of target sequence (equated with viral density). As a cut-off, samples with Ct values of 35.0 and below were considered positive. An RSV infection episode was defined as the period within which an individual provided specimens which were PCR positive for the same infecting RSV group with no more than 14 days separating any two positive samples. A household outbreak was defined as a period in which a household experienced more than one individual infection episode, without there being more than 14 days between any two infection episodes. A primary/index case was a person first identified to have an RSV infection leading to a household outbreak. A first episode was the first episode in that season that an individual had while a subsequent episode was a second or third episode that a person

may have had during the follow up. Virus test results, recorded in Ct values, were converted to \log_{10} ribonucleic acid (RNA) copy numbers (direct measure of viral density) to enable plotting of the time – concentration curve. The equation $y = -3.308x + 42.9$ was used to convert Ct values (y), to their \log_{10} RNA equivalents (x), as a way of quantifying RNA²⁰. The converted values are hereafter referred to as viral densities. Amount of virus shed during an infection episode was estimated by calculating the area under the time-concentration curve (AUC) using the trapezoidal rule²¹, with units of copies * days. The peak viral density for each episode was identified as the highest measured viral density in an infection episode.

Calculation of area under curve

For every episode, calculation of the area under the curve involved plotting a time concentration curve of the different log viral densities over the duration of shedding of the episode (Figure 1). Three scenarios were explored which take account of uncertainty arising from the sampling intervals, estimating a minimum, midpoint and maximum AUC, for which calculations are described below, aided by illustrations in Figure 1A – C. For each scenario, two examples are described: the first with one positive observation, and the second with three positive observations (Figure 1). This analogy

can be extended to episodes with a different number of positive observations. The number of positive observations in an episode ranged from 1 to 11.

a) Minimum area under curve. The minimum AUC estimate assumed that an individual started shedding on the day of the first positive sample in an infection episode, and stopped shedding on the day of the last positive sample of the episode, in the form shown in Figure 1A, with viral density Y_1 (\log_{10} RNA copies), and hence calculated as,

$$AUC = 0.5 * Y_1 * 1 \text{ day.}$$

If an episode had two or more consecutive positive observations, AUC was calculated by including all days in which they were assumed to be shedding. Thus, for an example episode with three positive observations, with viral density Y_i , $i=1,3$, respectively, then:

$$AUC = (0.5 * ((D_2 - D_1) + 1) * (Y_1 + Y_2)) + (0.5 * (D_3 - D_2) * (Y_2 + Y_3)) \dots \dots \dots (i).$$

Note the addition of one day to the calculation such that shedding on day 1 is included. Days with zero Ct values in between samples

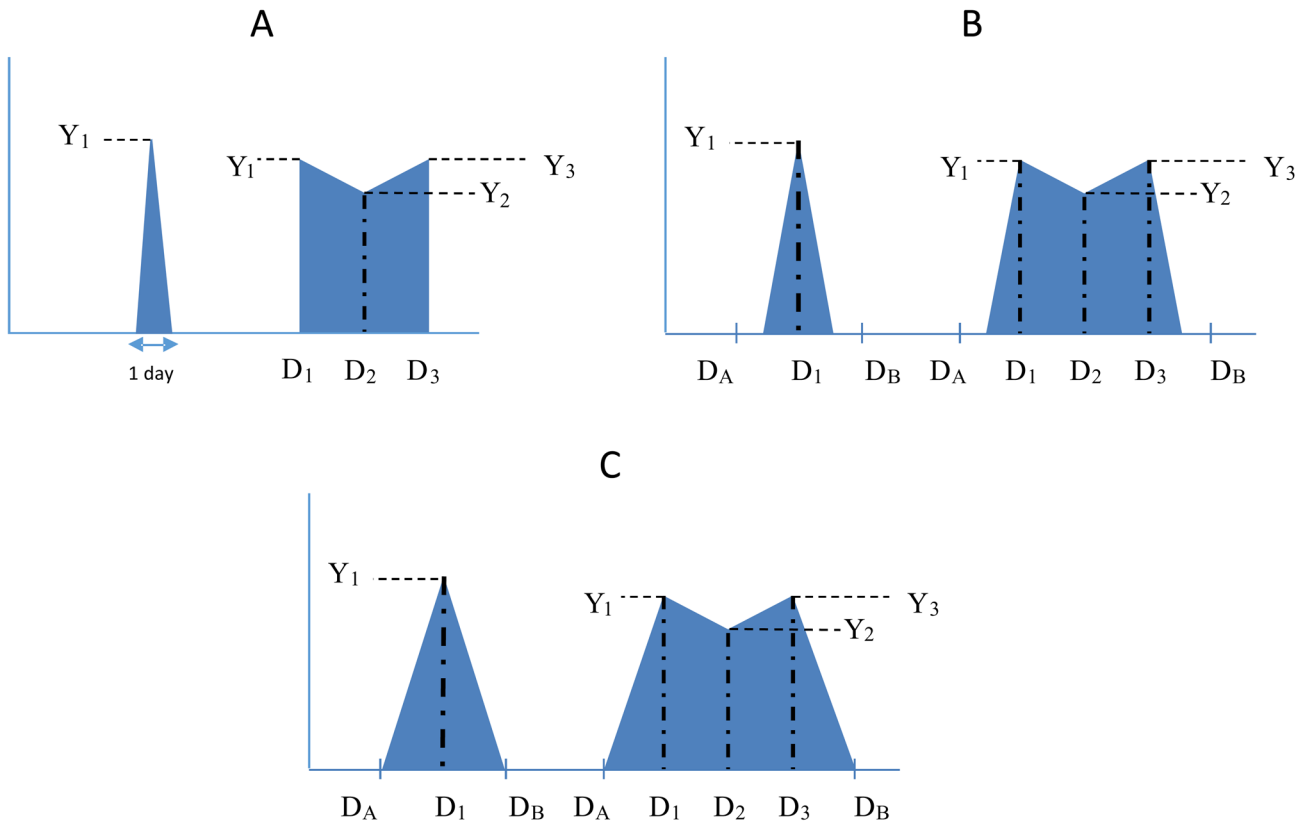


Figure 1. Schema depicting the patterns of virus shedding for three possible scenarios, **A**, **B** and **C**, graphically representing the minimum, midpoint and maximum AUC calculations, respectively, defined in the Methods text. For each scenario two example episodes are shown, the first with only one virus positive sample, and the second with three virus positive samples. Y_1 – Log viral density values; D_A – Day of last negative sample before start of episode; D_B – Day of first negative sample after end of episode; D_1 – Day of first positive sample; D_2 – Day of second positive sample; D_3 – Day of third positive sample.

with positive Ct values were taken to have zero viral density, after conversion of the Ct values to \log_{10} viral RNA copies. Therefore, AUC was calculated as in (i) above including both zero and non-zero observations. For left and right censored episodes, calculation of the amount of virus remained as in (i) above.

b) Midpoint area under curve. The midpoint AUC estimate assumed that an individual began shedding midway between the day of the first positive sample of an infection episode and the day of the last negative sample before the start of the episode. An individual was assumed to have stopped shedding midway between the day of the last positive sample of an episode and the day of the first negative sample after the episode (Figure 1B).

Hence to calculate the AUC for an episode with one positive observation:

Let the midpoint days be: $x = (D_A + D_1)/2$, and $y = (D_B + D_1)/2$, where D_A , D_B and D_1 are, respectively, the day of the last negative sample before the start of the episode; day of the first negative sample after the end of the episode; day of the first positive sample, hence

$$AUC = 0.5 * ((y - x) + 1) * Y_1$$

To calculate the AUC for an episode with three positive observations:

Let the midpoint days be: $x = (D_A + D_1)/2$, and $y = (D_B + D_3)/2$, where D_2 and D_3 are the days of the second and third positive samples, respectively, so

$$AUC = (0.5 * Y_1 * ((D_1 - x) + 1)) + (0.5 * (D_2 - D_1) * (Y_1 + Y_2)) + (0.5 * (D_3 - D_2) * (Y_2 + Y_3)) + (0.5 * Y_3 * (y - D_3)) \dots \dots \dots (ii).$$

Days with zero Ct values in between samples with positive Ct values were taken to have a Ct value of 40 (the Ct value translating to the lowest viral load), then converted to viral density along with other Ct values in the episode, and the AUC calculated as in (ii) above.

For left and right censored episodes, with unknown D_A and D_B , respectively, shedding was assumed to start 1.85 days (i.e. half the mean sampling interval) before the first observed day of shedding, D_1 , and to finish 1.85 days after the last positive sample (D_3 in the example of (ii) above). Hence, in (ii) above, $(D_1 - x)$ and $(y - D_3)$ were replaced with the value 1.85 days, for left and right censored data, respectively.

Maximum area under curve. To calculate the maximum AUC, individuals were assumed to have begun shedding immediately after the last negative sample before the start of an episode, and stopped shedding immediately before the first negative sample after the episode (Figure 1C). The viral densities for the days when individuals were considered to have begun and stopped shedding were assumed to be zero. Zero Ct values within an infection episode were assigned viral densities equal to the average recorded for the first positive sample before and first positive sample after the zero Ct samples.

Hence, to calculate the AUC for an episode with one positive observation:

$$AUC = 0.5 * (D_B - D_A) * Y_1,$$

and to calculate the AUC for an episode with three positive observations:

$$AUC = (0.5 * Y_1 * (D_1 - D_A)) + (0.5 * (D_2 - D_1) * (Y_1 + Y_2)) + (0.5 * (D_3 - D_2) * (Y_2 + Y_3)) + (0.5 * Y_3 * (D_B - D_3)) \dots \dots \dots (iii).$$

Left and right censored episodes were treated similarly to that for midpoint AUC, but assuming shedding began and ended 3.7 days (i.e. the mean sampling interval) before the first positive sample of an episode and after the last positive sample of an episode. Hence in (iii) above, $(D_1 - D_A)$ and $(D_B - D_3)$ were replaced with the value 3.7 days, for left and right censored data, respectively.

Statistical analysis

Data were analysed using Stata version 13.1 (StataCorp. College Station, TX: StataCorp LP. RRID:SCR_012763) and RStudio version 0.99.489 (RStudio Team. RStudio, Inc., Boston, MA. RRID:SCR_000432). The differences in the distribution of amount of virus by various characteristics was tested using the Wilcoxon rank-sum test and Kruskal Wallis test as appropriate. Spearman's rank correlation coefficient was used to find the association between the different measures of viral quantity calculated. Linear regression was used to identify the main factors associated with the amount of virus shed, i.e. AUC. Possible covariates were chosen from the dataset by selecting those that could plausibly be associated with AUC. A univariable analysis was carried out to identify factors associated with AUC. All variables with a p-value of <0.1 in the univariable analysis were included in the multivariable regression analysis. The final multivariable regression model was developed by a backward elimination procedure, removing variables with a p-value >0.05 in each step using likelihood ratio tests. Risk factors were removed in descending order of strength of association determined from the multivariable analysis. Two interactions were tested in the multivariable model. An interaction between age and symptom status was tested to determine if the effect of age on the amount of virus shed varied by symptom status. An interaction between age and primary/secondary case status was also tested because it was suspected that the effect of age on the amount of virus shed could vary by primary/secondary case status. We chose to adjust for sex *a priori* because it was considered an important risk factor to acute respiratory infections²².

Ethics

The household study was approved by the KEMRI Ethical Review Committee in Kenya, SSC No. 1651, and the Biomedical Research Ethics Committee at the University of Warwick in the United Kingdom. Data was anonymised by using unique identification numbers for each participant and household.

Consent

Written informed consent was obtained from all participants and/or their parents or guardians.

Results

Infection episodes

The mean (SD) sampling interval was 3.7 (2.3) days. The median number of swabs collected for an individual was 41, with the minimum being 1 swab and the maximum 48 swabs. There were 537 (3.2%) samples from 179 individuals that were positive for RSV. RSV group A only, group B only, and group A/B co-infections were detected in 231 (1.5%), 287 (1.8%), and 19 (0.1%)

NPS collections, respectively (Table 1). From the 179 infected individuals, a total of 208 infection episodes were observed during the six-month study period; 180 were fully observed episodes while 13 and 17 were left and right censored episodes respectively. Two episodes were both left and right censored. Eighty three (39.9%), 111 (53.4%), and 14 (6.7%) episodes were associated with RSV group A, group B and a co-infection respectively. One hundred and fifty two (84.9%) individuals had one episode,

Table 1. Distribution of RSV infection episodes among individuals during the 2009/2010 RSV season in a household cohort, coastal Kenya.

Characteristic	Categories	Individuals in study (n=493)		Individuals infected (n=179)		Infection episodes (n=208)		Total AUC ^a (n=7875.47)	
		n	%	n	%	n	%	n	%
Age at infection	<1y	55	11.2	31	17.3	35	16.8	2424.1	30.7
	1-<5y	82	16.6	41	22.9	51	24.5	2011.8	25.5
	5-<15y	165	33.5	64	35.8	73	35.1	2241.7	28.4
	15-<40y	147	29.6	35	19.6	40	19.2	854.0	10.8
	> = 40y	44	8.9	8	4.5	9	4.3	366.9	4.6
Sex	Female	272	55.2	96	53.6	118	56.7	4392.4	55.6
	Male	221	44.8	83	46.4	90	43.3	3506.2	44.4
Symptomatic	No	87	17.4	71	39.7	90	43.3	2145.2	27.2
	Yes	406	82.6	108	60.3	118	56.7	5753.3	72.8
Participant in school ^b	No	259	52.5	101	56.4	119	57.2	5184.2	65.6
	Yes	204	41.5	78	43.6	89	42.8	2714.4	34.4
Living with smoker in household	No	389	78.9	155	86.6	181	87	7081.2	89.7
	Yes	104	21.1	24	13.4	27	13	817.3	10.3
Household size	< = 5	37	7.5	22	12.3	25	12	992.6	12.6
	6 to 10	180	36.5	52	29.1	61	29.3	2605.3	33.0
	11 to 15	101	20.5	35	19.6	40	19.2	1360.5	17.2
	>15	175	35.5	70	39.1	82	39.4	2940.1	37.2
Infection with other viruses	No	43	8.7	83	46.4	110	52.9	5054.6	64.0
	Yes	450	91.3	96	53.6	98	47.1	2844.0	36.0
In household with outbreak	No	162	32.9	14	7.8	192	92.3	7460.6	94.5
	Yes	331	67.1	165	92.2	16	7.7	438.0	5.5
Infesting RSV group ^c	RSV A			66	36.9	83	39.9	2838.1	35.9
	RSV B			91	50.8	111	53.4	4170.0	52.8
	RSV A & B			22	12.3	14	6.7	890.5	11.3
Order of the episode in individual	First			152	84.9	179	86.1	7130.3	90.3
	Subsequent			27	15.1	29	13.9	768.3	9.7
Introducer of RSV infection in household	Primary case			66	36.9	70	33.7	2372.7	30.0
	Others			113	63.1	138	66.4	5525.9	70.0

^aTotal amount of virus shed (\log_{10} RNA copies * days). ^bIncludes some participants in the age groups 1-<5 years, 5-<15 years, and 15-<40 years only. ^cIndividuals could appear in one or more categories of this variable.

while 25 (14.0%) and 2 (1.1%) had two and three episodes respectively (Table 1). In addition, 24 (60.0%) households had one outbreak, while 5 (12.5%) had two outbreaks.

Overall amount of RSV shed

The amount of virus shed was presented in logarithmic form because the untransformed values were skewed to the right, and the residuals of the linear regression analysis using the untransformed outcome did not meet the normality assumption. Using log transformed values also helped to reduce the variance. The mean (variance) for the untransformed and log-transformed values was 3.73×10^7 (1.34×10^8) and 6.07 (1.51), respectively.

The median (interquartile range, IQR) of the amount of virus shed estimated by the minimum, midpoint and maximum AUC approaches was 31.1 (15.4, 57.9), 41.7 (24.3, 68.0), and 50.7 (31.0, 79.6) \log_{10} RNA copies per infection episode, respectively. The minimum and maximum AUC estimates were strongly correlated with the midpoint AUC estimates ($r=0.97$ $P < 0.001$ and $r=0.96$ $P < 0.001$, respectively) (Supplementary Material Figure S1). The alternative measures of RSV shedding, (shedding duration and peak viral density), were also strongly correlated to the midpoint AUC estimates; 0.94 ($P < 0.001$) and 0.74 ($P < 0.001$) (Figure 2). Note, however, that within a narrow range of AUC the range of peak virus can be wide. The distribution of the amount of RSV shed by the various characteristics explored was similar regardless of the estimation method (i.e. minimum, midpoint, or maximum AUC) (See Supplementary Material Table S1) and hereafter only the midpoint estimates are reported. RSV viral loads for both RSV A and RSV B seemed to be low at the beginning of the RSV season, but they later increased well into the season. However, the peak viral loads reached during the season for both RSV A and RSV B seemed to be similar (see Supplementary Material Figure S2).

Amount of RSV shed by various characteristics

The median (IQR) amount of virus shed was 71.0 (42.3, 96.7), 37.7 (24.6, 54.4), 25.0 (14.1, 36.7), 14.6 (9.4, 32.4), and 56.3 (5.7, 65.3) \log_{10} RNA copies for individuals aged $<1y$, $1-<5y$, $5-<15y$, $15-<40y$ and $\geq 40y$ respectively (Figure 3a) based on the midpoint approach, and there was strong evidence of a difference in distribution of AUC in different age groups ($P = 0.001$). Infection episodes associated with symptoms had a higher median (IQR) amount of virus shed than those without symptoms (42.5 (25.1, 66.0) vs 19.0

(9.8, 29.0) \log_{10} RNA copies; $P < 0.001$) (Figure 3b). Most symptomatic episodes were experienced by younger individuals (<20 years), while older individuals above 20 years mostly had asymptomatic episodes (Figure 4). Episodes associated with RSV A only, RSV B only, and RSV group A/B co-infection had a median amount of virus of 26.4 (15.1, 52.4), 28.8 (13.2, 48.9), and 66.1 (42.1, 75.5) \log_{10} RNA copies respectively ($P = 0.003$) (Figure 3c). Co-infection (associated with adenoviruses, rhinovirus, or coronaviruses) episodes had a median amount of virus of 38.7 (19.6, 65.3), while episodes not associated with any other infection had a median of 25.0 (11.1, 37.0) \log_{10} RNA copies, and there was strong evidence ($P = 0.003$) of a difference in the distribution of AUC in these two groups of episodes. Episodes that occurred during a household outbreak had a median (IQR) amount of virus of 31.6 (15.9, 54.8) \log_{10} RNA copies, compared to a median (IQR) of 20.3 (10.2, 35.4) \log_{10} RNA copies for episodes that were not associated with household outbreaks, though there was weak evidence of a difference in medians (see Supplementary Material Table S1). Infants had the highest peak viral density, and longest duration of shedding (Figure 5). Lowering the threshold below 35 Ct did not greatly change the results but reduced the total number of infections (see Supplementary Material Figure S3).

Factors associated with the amount of virus shed

The final multivariable regression model identified age, symptom status, RSV infecting group, order of the episode in an individual, and being a primary/secondary case in the household as the main factors associated with the amount of virus shed. The effect of age on the amount of virus shed seemed to vary with primary/secondary case status (Table 2). The results reported here are, therefore, adjusted for the above covariates. The difference in mean amount of virus in secondary case episodes for age groups $1-<5y$, $5-<15y$, $15-<40y$, and $\geq 40y$ when compared to infants $<1y$ was; -41.6 (95% CI: -53.7, -29.6), -43.1 (-55.5, -30.7), -50.6 (95% CI: -64.9, -36.3), and -29.4 (95% CI: -50.8, -8.1) \log_{10} RNA copies respectively ($P < 0.001$). There was no evidence ($p=0.594$) of a difference in amount of virus by age for primary cases. There was an increase in mean amount of virus for symptomatic episodes compared to asymptomatic episodes by 14.1 (95% CI: 6.3, 21.9) \log_{10} RNA copies, and strong evidence of a difference in means ($P < 0.001$). Mean amount of virus shed for subsequent episodes of individuals in the study RSV season compared to first infections was -10.4 (95% CI: -20.0, -0.8) \log_{10} RNA copies ($P = 0.03$). The difference

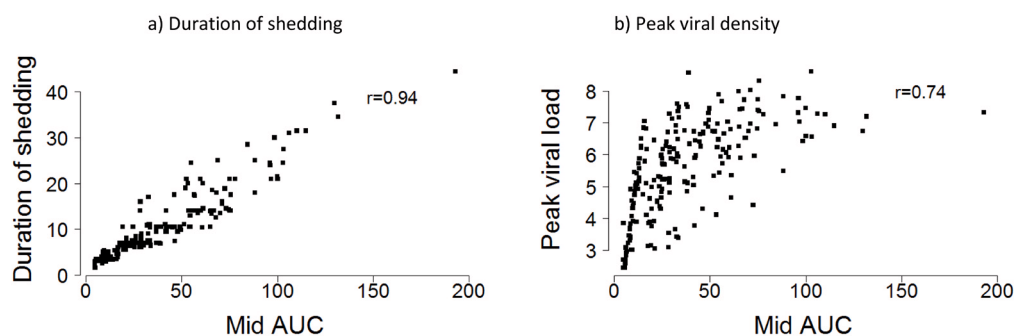


Figure 2. Scatter plots showing the correlation between (a) duration of shedding (days) and (b) peak viral density (\log_{10} RNA copies) and midpoint AUC (\log_{10} RNA copies*days) estimates.

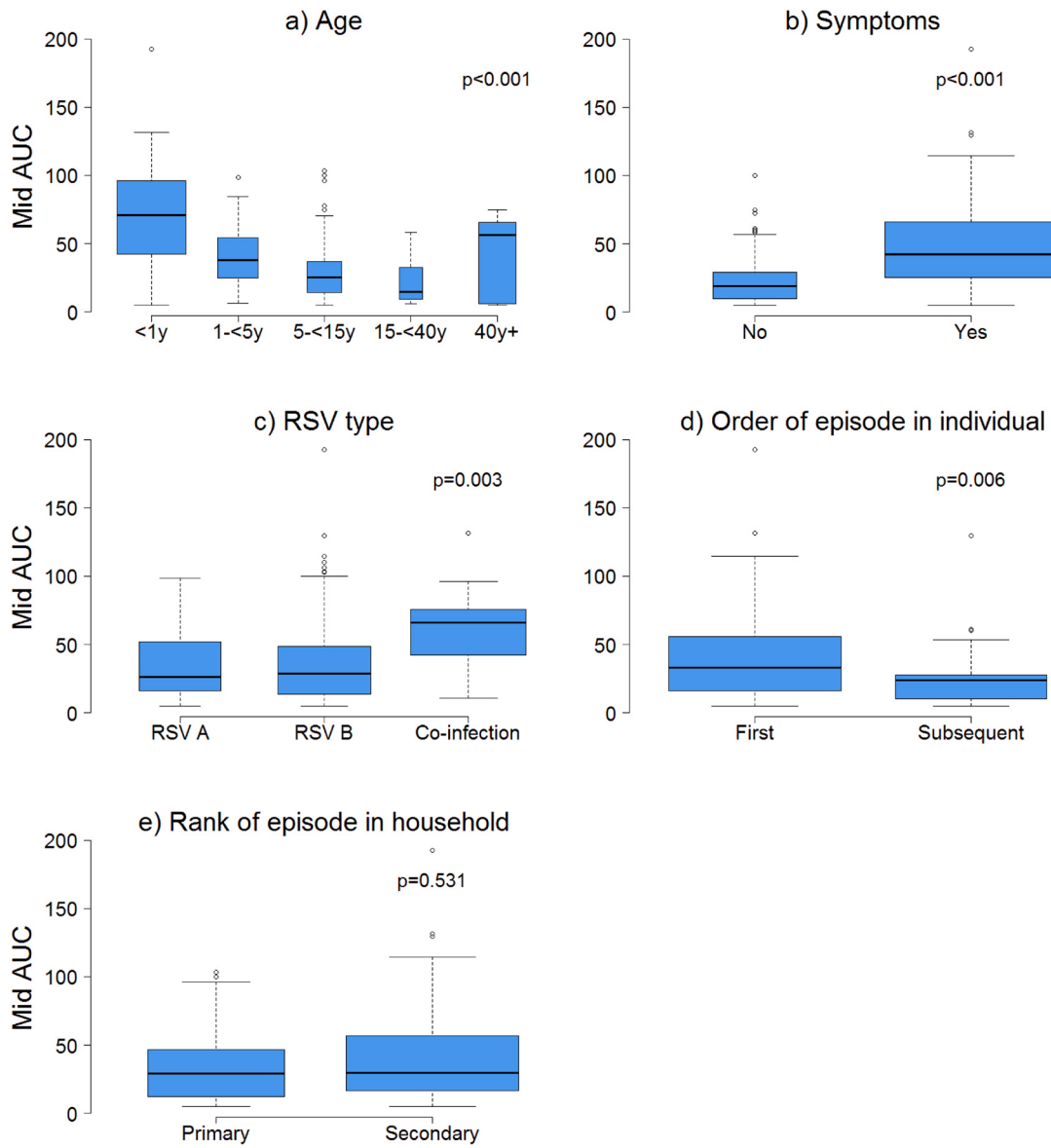


Figure 3. Box and whisker plots showing the distribution of amount of virus shed (AUC) using the midpoint estimates by **a)** age **b)** symptoms **c)** RSV type **d)** order of episode in individual, and **e)** primary or secondary episode in household. The width of a box relative to other boxes in the same plot is directly proportional to the mean duration of shedding in each category of that plot.

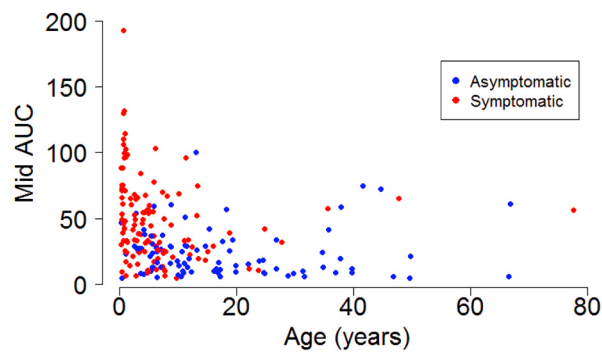


Figure 4. Scatter plot showing the relationship between midpoint AUC and age stratified by symptom status.

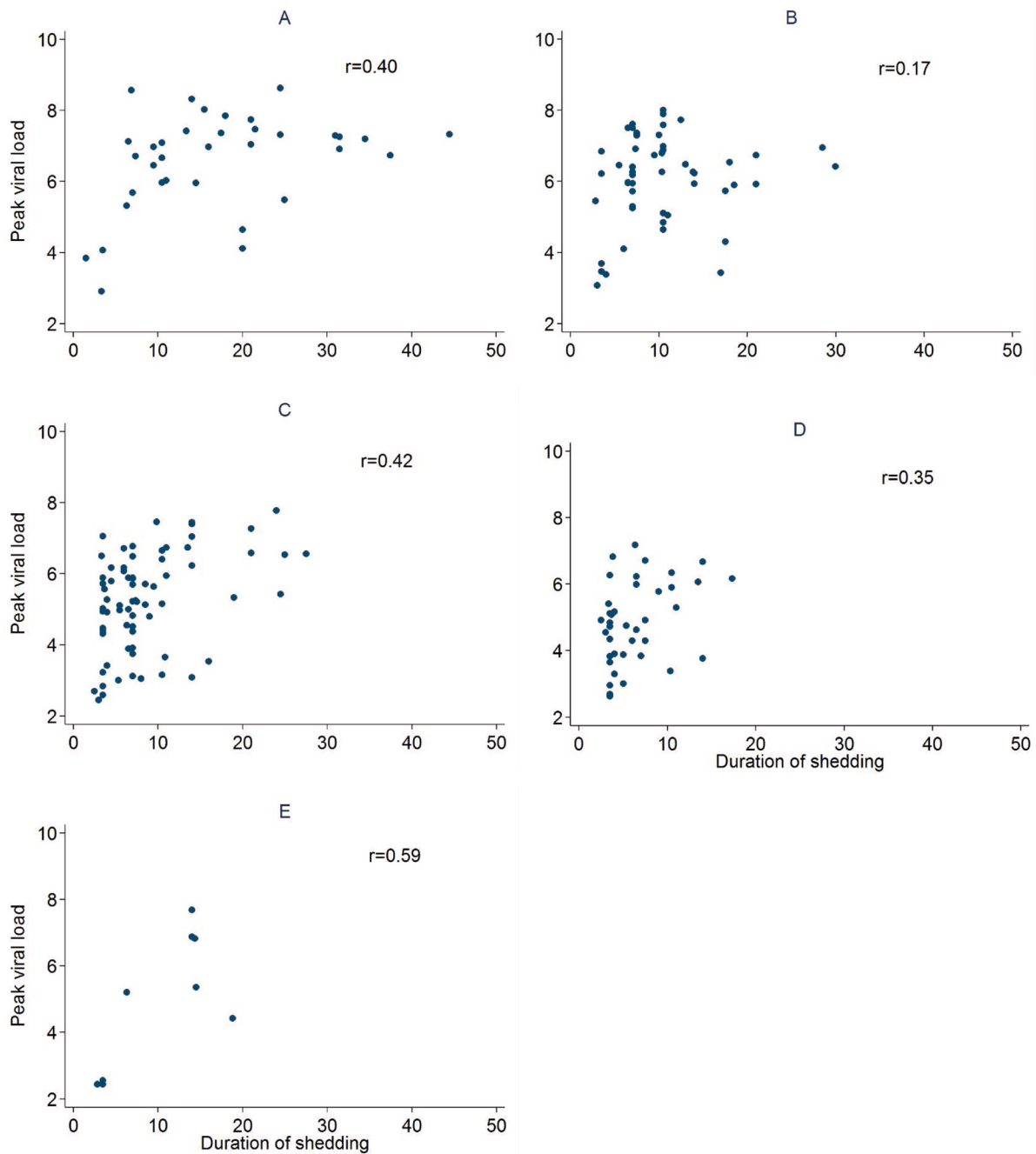


Figure 5. Scatter plots showing the relationship between duration of shedding (days) and peak viral density (\log_{10} RNA copies) for individuals aged **a)** less than 1 year old, **b)** 1 to 4 years, **c)** 5 to 14 years, **d)** 15 to 39 years, and **e)** 40 years and above.

in mean amount of virus for episodes of RSV B only and RSV A/B co-infections when compared to episodes of RSV group A only was 3.7 (95% CI: -3.1, 10.5) \log_{10} RNA copies, and 21.3 (95% CI: 7.3, 35.3) \log_{10} RNA copies ($P = 0.012$) respectively (Table 2). For all age groups, there was no evidence of a difference in amount of virus between primary and secondary episodes. There was no evidence that the effect of age on amount of virus varied by symptom status.

Discussion

We report a detailed analysis of RSV shedding patterns in which the amount of virus shed during the course of an infection was quantified, and factors associated with the quantity shed were identified. Young age at infection, presence of respiratory symptoms, intra-household acquisition of infection, being a first infection episode in the season for an individual, and having a co-infection were associated with increased amounts of RSV shed. The majority

Table 2. Factors influencing amount of virus shed in rural coastal Kenya during the 2009/2010 RSV season: Multiple linear regression analysis.

Factors	Categories	Beta ^a (log ₁₀ RNA copies)	95% CI	P-value
Primary cases: Age in years	<1y	Ref*		
	1-<5y	-0.21	-18.4, 17.9	0.594
	5-<15y	-7.3	-23.0, 8.4	
	15-<40y	-12.7	-30.7, 5.2	
	> = 40y	-9.4	-39.3, 20.5	
Secondary cases: Age in years	<1y	Ref*		
	1-<5y	-41.6	-53.7, -29.6	<0.001
	15-<15y	-43.1	-55.5, -30.7	
	15-<40y	-50.6	-64.9, -36.3	
	>=40y	-29.4	-50.8, -8.1	
Symptomatic	No	Ref*		
	Yes	14.1	6.3, 21.9	<0.001
Order of infection in individual	First	Ref*		
	Subsequent	-10.4	-20.0, -0.8	0.03
RSV group	RSV A	Ref*		0.012
	RSV B	3.7	-3.1, 10.5	
	Co-infection	21.3	7.3, 35.3	
Sex	Female	Ref*		0.287
	Male	-4.0	-10.2, 3.0	

^aDifference in mean AUC in comparison to reference category using midpoint estimates. *Reference category.

of virus available for transmission during an RSV epidemic appears to arise from individuals in their first year of life and therefore undergoing their first RSV infection episode (since all were born after the preceding RSV epidemic). Household size did not influence the per person amount of virus shed, and neither did the presence of a smoker in the household. In this rural community, indoor smoke from solid fuels (mainly firewood and charcoal) was ubiquitous and likely to conceal any specific effects of cigarette smoke. In addition, we only had 18 smokers out of the 493 participants. So, relatively few people lived with smokers. Also, we did not collect data on contact patterns of household members with the smokers or if the smoker did smoke in the house or outside which could affect the level of exposure to smoke within the households. The apparent effect of school going status on amount of virus shed was confounded by age, as school was attended by young individuals, who in turn shed more virus. RSV viral densities appeared to be low for most individuals in the beginning of the season but increased well into the season. This was a reflection of the occurrence of the RSV epidemic. Viral densities for RSV B were first to increase followed by viral densities for RSV A. RSV B infections spread to more individuals probably due to earlier onset by chance compared to RSV A infections.

The finding that 90.3% of virus shed was from individuals experiencing their first episode in the RSV season indicates that

first infections are most important in virus transmission compared to subsequent infections (Table 1). Modelling studies have been unable to determine to what extent reinfections contribute to transmission^{11,23}, which has a big influence on predictions of the effect of vaccination on transmission dynamics¹¹. Whilst these results do not include the opportunities to transmit, which will be generally higher for school-going children undergoing their second or third infection⁹, they suggest that people with their first episode in the season are the most likely to transmit infection when they contact susceptible individuals. Peak viral density and duration of infection episodes were highly correlated with the amount of virus shed. Infants had the highest peak viral density, and longest duration of shedding (Figure 5).

Infants shed the highest median amount of virus followed by individuals aged 40 and above. The adults (15 to 39 years) shed the lowest amounts (Figure 3a). Acquired immunity (following exposure to RSV) and physiological development of the airways with age have been linked with reduced risk of severe RSV disease^{12,13}. We speculate that elderly individuals are more prone to infection because of their deteriorating immune response which could explain the higher amount of virus in individuals over 40 years, compared to other adults. However, this study cannot discriminate between the two factors of young age and a first infection as the dominant factors associated with shedding (infants all had first infections,

and we do not know the history of previous infection in all other ages). Individuals with a co-infection of RSV group A and B shed twice the amount of virus compared to those infected with either RSV A or B only, suggesting that the physiological infection processes for two viruses might be independent (rather than competitive or synergistic). Similar patterns of shedding were reported based on shedding duration estimates using this dataset¹³ and others^{12,15}. These findings can provide a hint to pathogenesis of RSV disease⁶. A recent experimental study has reported disease severity to be closely related to viral load¹⁶ even though another shows contrasting observations on disease severity and viral load link¹⁰. Our study showed a strong positive association between virus shedding and presence of respiratory symptoms. Symptomatic individuals also shed higher amounts of virus than asymptomatic individuals irrespective of age. Overall those who were both young (<5 years) and symptomatic contributed to the highest amount of virus shed (51.8%). Using amount of virus as a correlate of infectiousness, these two factors are associated with virus spread within the community and in families and thus have direct implications in design and delivery of transmission-blocking interventions such as vaccines.

This analysis had some limitations. First, the nasal samples were collected with intervals of 3 to 4 days hence a complete profile of the viral density changes over time could not be captured. This has implications on the accuracy of our AUC estimates. However we minimize this by using a conservative approach providing three possible estimates i.e. minimum, midpoint and maximum estimates as detailed in the [Supplementary Material](#). Furthermore, the three measures were strongly correlated and their variations by various characteristics such as age and symptoms, were comparable. Second, Ct values were converted to RNA copies using published associations²⁰ hence providing only a relative measure of virus density in the sample taken rather than absolute values of the amount of virus shed. Using the inverse of the Ct values as a measure of viral load did not change the distribution of AUC by various characteristics or any conclusions made from the results. There are no published studies using our approach in a community setting to facilitate comparison. Our approach takes into account both changing viral load and the duration of shedding making it more robust. However, its accuracy also relies on user selected thresholds for measures of viral load, and this could result in some differences in findings across studies. A quantitative PCR would be preferable. Third, there is inherent variability in estimated viral load due to variation (both biological and methodological) in the amount of virus collected using a nasopharyngeal swab. Deep NPS collections might have a higher density than perinasal swabs but events prior to sampling, e.g. sneezing, might introduce variations in sample quality. However this is likely to be non-systematic hence only conservatively affects our final conclusions. Finally, the assumption is made that virus density as measured by PCR is a measure of infectious virus. The most likely limitation of this assumption is over-estimating the duration of and peak shedding of viable virus. Generally, the relationship between the measures of viral shedding and infectiousness to other people are unknown. In the current analysis we took logarithms of estimated viral densities, under the general assumption that the risk of transmission will saturate at higher viral shedding. We then integrated over this logarithmic measure to obtain a measure of total infectiousness. An alternative would be to have integrated over the untransformed

viral density, and then taken logarithms prior to statistical analysis. This approach did not give substantially different results. Estimation of the functional relationship between viral shedding and transmission is being considered in on-going analysis of the epidemics observed within households.

In conclusion, long shedders tend to shed more virus and are likely to be more infectious. However, it is also evident that individuals with similar durations of shedding may shed different amounts of virus. Individuals with a combination of high viral density and long shedding durations are likely to shed the most virus. This improves on existing literature that measures infectiousness using the duration of shedding only. The groups that are most likely to spread infection are individuals with a first infection episode, the symptomatic, and most predominantly, the young, which has implications on which groups to target for vaccination in RSV prevention strategies.

Data availability

The dataset used in this study, together with associated analysis code (.do), graphics code (.R), variable codebook (.pdf), and summary file (.txt) are available from Harvard Dataverse: doi:[10.7910/DVN/MOTEJH](https://doi.org/10.7910/DVN/MOTEJH)²⁴. The dataset has restricted access, retaining Personal Identifiable Information (specifically dates) for analysis duplication. Please request access as instructed in the Readme.txt file. Other files are open access.

Author contributions

Miriam Wathuo did the analysis and interpretation of the data, and drafted the article. Graham Medley was involved in conception and design of the household study, revising the article critically for intellectual content and final approval of the version to be submitted. James Nokes was involved in conception, design and implementation of the household study, revising the article critically for intellectual content and final approval of the version to be submitted. Patrick Munywoki designed the household study and directed its implementation. He was also involved in the analysis, interpretation of the data, revising the article critically for intellectual content and final approval of the version to be submitted.

Competing interests

No competing interests were disclosed.

Grant information

This work was supported by the Wellcome Trust [102975], [090853].

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgement

We would like to thank the Virus Epidemiology and Control Research Group at KEMRI-Wellcome Trust Research Programme in Kilifi, Kenya for their contribution towards the development of this manuscript, and the Data Management Group for their support during the analysis. Special thanks also goes to the field workers for their hard work in sample collection, and the laboratory team for their dedication in sample testing. This article was published with permission from the Director of KEMRI.

Supplementary material

This file provides (a) Scatter plots showing the association between minimum, midpoint and maximum AUC (**Figure S1**), (b) graphical presentation of the relationship between amount of virus and duration of shedding for different thresholds of positivity (**Figure S2**), and temporal patterns of virus load (**Figure S3**), and (c) descriptive analysis tables for virus shedding (**Table S1** and **Table S2**).

[Click here to access the data.](#)

References

- Nair H, Nokes DJ, Gessner BD, *et al.*: **Global burden of acute lower respiratory infections due to respiratory syncytial virus in young children: a systematic review and meta-analysis.** *Lancet.* 2010; **375**(9725): 1545–1555.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Glezen WP, Taber LH, Frank AL, *et al.*: **Risk of primary infection and reinfection with respiratory syncytial virus.** *Am J Dis Child.* 1986; **140**(6): 543–6.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Henderson FW, Collier AM, Clyde WA Jr, *et al.*: **Respiratory-syncytial-virus infections, reinfections and immunity. A prospective, longitudinal study in young children.** *N Engl J Med.* 1979; **300**(10): 530–4.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Modjarrad K, Giersing B, Kaslow DC, *et al.*: **WHO consultation on Respiratory Syncytial Virus Vaccine Development Report from a World Health Organization Meeting held on 23–24 March 2015.** *Vaccine.* 2016; **34**(2): 190–7.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Karron RA, Wright PF, Belshe RB, *et al.*: **Identification of a recombinant live attenuated respiratory syncytial virus vaccine candidate that is highly attenuated in infants.** *J Infect Dis.* 2005; **191**(7): 1093–104.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Wright PF, Karron RA, Belshe RB, *et al.*: **The absence of enhanced disease with wild type respiratory syncytial virus infection occurring after receipt of live, attenuated, respiratory syncytial virus vaccines.** *Vaccine.* 2007; **25**(42): 7372–7378.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Anderson LJ, Dormitzer PR, Nokes DJ, *et al.*: **Strategic priorities for respiratory syncytial virus (RSV) vaccine development.** *Vaccine.* 2013; **31**(Suppl 2): B209–15.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Güvenel AK, Chiu C, Openshaw PJ: **Current concepts and progress in RSV vaccine development.** *Expert Rev Vaccines.* 2014; **13**(3):333–44.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Kiti MC, Kinyanjui TM, Koech DC, *et al.*: **Quantifying age-related rates of social contact using diaries in a rural coastal population of Kenya.** *PLoS One.* 2014; **9**(8): e104786.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Lee FE, Walsh EE, Falsey AR, *et al.*: **Experimental infection of humans with A2 respiratory syncytial virus.** *Antiviral Res.* 2004; **63**(3): 191–196.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Kinyanjui TM, House TA, Kiti MC, *et al.*: **Vaccine Induced Herd Immunity for Control of Respiratory Syncytial Virus Disease in a Low-Income Country Setting.** *PLoS One.* 2015; **10**(9): e0138018.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Okiro EA, White LJ, Ngama M, *et al.*: **Duration of shedding of respiratory syncytial virus in a community study of Kenyan children.** *BMC Infect Dis.* 2010; **10**: 15.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Munywoki PK, Koech DC, Agoti CN, *et al.*: **Influence of age, severity of infection, and co-infection on the duration of respiratory syncytial virus (RSV) shedding.** *Epidemiol Infect.* 2015; **143**(4): 804–12.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hall CB, Geiman JM, Biggar R, *et al.*: **Respiratory syncytial virus infections within families.** *N Engl J Med.* [electronic article]. 1976; **294**(8): 414–9.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Frank AL, Taber LH, Wells CR, *et al.*: **Patterns of shedding of myxoviruses and paramyxoviruses in children.** *J Infect Dis.* [electronic article]. 1981; **144**(5): 433–41.
[PubMed Abstract](#) | [Publisher Full Text](#)
- DeVincenzo JP, Wilkinson T, Vaishnav A, *et al.*: **Viral load drives disease in humans experimentally infected with respiratory syncytial virus.** *Am J Respir Crit Care Med.* 2010; **182**(10): 1305–1314.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Munywoki PK, Koech DC, Agoti CN, *et al.*: **Frequent Asymptomatic Respiratory Syncytial Virus Infections During an Epidemic in a Rural Kenyan Household Cohort.** *J Infect Dis.* 2015; **212**(11): 1711–8.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Munywoki PK, Koech DC, Agoti CN, *et al.*: **The source of respiratory syncytial virus infection in infants: a household cohort study in rural Kenya.** *J Infect Dis.* 2014; **209**(11): 1685–92.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hammit LL, Kazungu S, Welch S, *et al.*: **Added value of an oropharyngeal swab in detection of viruses in children hospitalized with lower respiratory tract infection.** *J Clin Microbiol.* 2011; **49**(6): 2318–2320.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Nolan T, Hands RE, Bustin SA: **Quantification of mRNA using real-time RT-PCR.** *Nat Protoc.* 2006; **1**(3): 1559–1582.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Kreide D, Lahr D: **Trapezoid Rule and Simpson's Rule.** 2010.
[Reference Source](#)
- Simoes EA: **Environmental and demographic risk factors for respiratory syncytial virus lower respiratory tract disease.** *J Pediatr.* 2003; **143**(5 Suppl): S118–S126.
[PubMed Abstract](#) | [Publisher Full Text](#)
- White LJ, Waris M, Cane PA, *et al.*: **The transmission dynamics of groups A and B human respiratory syncytial virus (hRSV) in England & Wales and Finland: seasonality and cross-protection.** *Epidemiol Infect.* 2005; **133**(2): 279–89.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Munywoki PK, Nokes DJ: **Replication Data for: Quantification and determinants of the amount of respiratory syncytial virus (RSV) shed using real time PCR data from a longitudinal household study.** *Harvard Dataverse, V5.* 2016.
[Data Source](#)

Open Peer Review

Current Referee Status:    

Version 1

Referee Report 31 January 2017

doi:[10.21956/wellcomeopenres.11076.r19771](https://doi.org/10.21956/wellcomeopenres.11076.r19771)



John DeVincenzo

Department of Pediatrics, University of Tennessee College of Medicine, Memphis, TN, USA

This is a report that adds significantly to the medical literature on the subject. It is from a well-designed study from a longitudinal community household study in rural Kenya from whom collections of respiratory secretions were collected with and without bias with respect to symptomatology. As such, it is a very valuable database to assess many important clinical questions.

These authors attempt to leverage this impressive dataset to evaluate quantitative virology relationships to clinical and epidemiological endpoints. This is a very valuable goal and they have generated results and analyses that are important, however, there are numerous problems with this approach which deserve to be reviewed, and their conclusions need to be modified in light of these limitations.

- 1. Relatively non-quantitative collection technique:** Unfortunately, because of the nature of the study (outpatient from village homes), the collections themselves had to be performed using a swab technique. Swab collections (even if done using a deep swab approach) have been shown to produce specimens that are difficult to use for quantitative virology. Variations in depth of swab acquisition, variations in depth of the transition zone from squamous to respiratory epithelial cells within the nose, reduced sensitivity caused by the small volume collection and the large volume dilution required during the nucleic acid extraction technique, and sample to sample variations of the PCR technique (which can not be overcome by re-running the specimen because of lack of sufficient collection volume) all contribute to creating this relatively non-quantitative collection technique.
- 2. Relatively non-quantitative PCR techniques employed:** Molecular quantification of viruses within a sample requires a standard curve to assign a quantitative value based upon a Ct value. One cannot use the Ct value itself without it being read from a standard curve. These standard curves can be created externally from the PCR run itself (less desirable) or can be created internally during the actual run which includes the unknown samples from the patients themselves (more desirable). The creation and design of the standard curves is extremely important in determination of quantification and can be based on whole viral genome from infected cells (thereby controlling for the nucleic acid extraction and purification steps, and controlling for the reverse transcriptase step). Less optimal (but much easier to construct) standard curves can be created that utilize cloned cDNA (easily giving a copy number, but not properly controlling for either the nucleic acid extraction, purification, nor reverse transcription steps). Unfortunately, this study apparently has not constructed standard curves, and they rely solely on generated Ct values to imply quantity. It is, in some cases, acceptable to do so when using a single-plex assay (where relative viral concentrations can fairly be compared to the same technique on different

specimens). But this approach is never acceptable to use when multiplexed assays are employed (see below).

3. **Introduction of technique-driven systemic bias in molecular quantification:** This molecular quantification bias comes in two forms:
 - A) multiplex-PCRs detect viral amplicon differently when there are two molecular targets within the same sample based on competition for ions, primers, and probes, and enzymes within the reaction.
 - B) The relative amplification efficiency (and the probe-binding-induced fluorescent signal read out by the PCR machine, will be different depending on the probe and the amplicon size and the relative binding efficiencies of the forward and reverse primers. This makes it impossible to quantitatively assess the quantification of two different viruses (RSV-A and B) within a mixed sample and also makes it impossible to compare one target's quantity (RSV-A) with another target's quantity (RSV-B).
4. **Failure to evaluate infectious viral particles and using non-infectious viral particle counts to make conclusions about infectious viral load.** The authors find and conclude that: "The majority of virus available for transmission during an RSV epidemic appears to arise from individuals in their first year of life and therefore undergoing their first RSV infection episode." Although I agree with this statement, it is important to realize that from data within this manuscript itself, this statement assumes that infectious viral particles equate to quantitative PCR copy number. This assumption is certainly not true. Infectious virus particles are best quantified by a quantitative culture approach. Factors within the respiratory secretions themselves (such as IgA) have been shown to be associated with RSV culture negativity despite persistent high concentrations of RSV genome continuing to be detected within human respiratory secretions after a documented RSV infection^{1,2}. Infant's ability to generate RSV-specific IgA is significantly less than the ability of adults to do so during an RSV infection^{3,4}. The author's conclusion that infants likely contribute more than do adults to epidemic spread of RSV infection has been carefully modeled in a recent paper⁵ in which two key factors; infectious viral particle quantity over time, and quantitative measures of human contact activity during times of infection, were incorporated into the model. The implications of these factors promoting RSV community spread should influence the types of RSV vaccines to be developed and the age ranges targeted by these vaccines, and these different vaccine scenarios have been modeled⁵.
5. **Questions regarding viral load area under the curve calculations.** Any area is geometrically and mathematically defined by multiplying two linear dimensions together. The same is true for viral load area under the curve (in this paper it is called viral density area under the curve for some reason). Viral load AUC is Viral load (units) multiplied by Time (units). The resulting units of viral load are always containing the unit of time. (For example "log PFUe x days", or "Copies x hours"). The units of viral load AUC in this paper are incorrect and need to be redefined. When one realizes the mathematical definition of AUC, then it is no surprise that Fig 2a and 2b show a lot of correlation between duration of shedding and AUC (because time (duration) is actually in the denominator of the AUC value).
6. **Problems of determining co-infection with RSV-A and RSV-B.** Students of RSV need to be very careful in reporting and interpreting data regarding co-infections. This paper is no exception. This paper and others assume that when a multiplexed assay reports that the wavelength of the excited probe is orange and another wavelength of the excited probe is green (I am giving a hypothetical example) that this means that both RSV-A and B are present in the

sample. However this may or may not be the case. For example, if the virus is an RSV-A virus, the probe for RSV-A binds relatively tightly to the amplicon through strict base-pair matching, and then cleavage of the reporter molecule from the quencher molecule allows the detection of the amplicon (through inhibition of FRET). However, it is very possible that the RSV-B probe will also bind to the amplicon at a certain number of base pairs, and this can result in the RSV-B reporter being detected (erroneously causing the authors to read that RSV-B is also present in the sample). Because of this problem, this paper's results, analyses, and conclusions regarding "co-infections" of RSV-A and B need to be taken with caution.

7. **Problems with co-infections with RSV and other respiratory viruses** (Adeno, Rhinoviruses, and Coronaviruses). RSV infection causes massive shedding of respiratory epithelial cells into the airway lumen (which are then collected by the swab technique). Any persistent virus present in those shed cells will therefore be detected at higher frequency during an RSV infection. This is an alternative explanation of the correlation between viral load of RSV and detection of these other infections (Top of Pg. 8). This phenomenon carries much different implications than are assumed to be present by these authors.
8. **Associations with age and outcome (area under the curve viral load)**. The authors do an excellent job of assessing the associations between area under the curve and other factors. After observing the problems with the actual viral load area under the curve measurements themselves, (see points 1,2,3,5 above) it appears that their data shows that the primary host characteristic that affects area under the curve viral load is whether or not children have had a prior RSV infection. This fits with other data from the literature. However, it would be nice to know whether young age within the infant age category was predictive of increased area under the curve viral load. Unfortunately, the data set does not contain data on age in weeks, days, or months within the first year of life. I suspect that younger ages within the infant age range, i.e. the RSV-naïve population, will be related to the area under the curve viral load. Other databases will likely need to be leveraged in order to determine this important question. Although the number of infant infections in this database will likely be small, the authors should attempt to break down the ages within the infant age range in order to try to answer this important question. The infant antiviral functional immune response is likely not fully developed at birth, and RSV can be used to interrogate the ontogeny of this immune response⁶, which can have major implications on vaccine development.
9. **Associations with other factors and outcome (area under the curve viral load)**. The manuscript also indicates that presence of respiratory symptoms, intra-household acquisition of infection, being a first infection episode in the season for an individual, and having a co-infection were associated with increased amounts of RSV shed. We should think about the root causes of these associations so that we can understand principles of RSV infection. I do not know the intricacies of the subject acquisition and identification, but is it possible that intra-household transmission simply is a surrogate marker for earlier swabbing of the patient? We know that viral load peaks in the nose shortly after symptom onset, and that viral load declines thereafter. Therefore, having a first swab earlier would create a subject with a greater viral load AUC. An alternative explanation is that perhaps an inoculum effect is present and that children are inoculated within a household transmission with greater amounts of virus, thus causing greater AUC viral load. Past studies have tried to determine an inoculum effect for RSV and crowding does not seem to be a predictor of viral load^{7,8}. Likewise, in experimental RSV infection models of adults, there appears to be little effect of inoculum on AUC viral load². It is likely that RSV inoculum affects whether a person becomes infected, but once that person is infected that inoculum has little effect on viral load, AUC or disease severity. I have addressed the issue of co-infection (and

alternative explanations for the observed data) in points 6 and 7 above.

10. **Waning control of RSV infection with advanced age.** Perhaps the greatest contribution of this paper to the medical literature is the finding that area under the curve viral load begins to rise again in adulthood after the age of 40. This is a fascinating finding, and could result from either immune senescence, or reduced exposure to boosting infections of RSV, or a combination of both. Whether this adult age-related increase in RSV area under the curve viral load is similar in developed countries will need to be evaluated.

Overall this paper adds significantly to the medical literature into our understanding of RSV in communities. These authors should be lauded and encouraged in their efforts to extract as much information as possible from carefully constructed clinical trials.

References

1. Bagga B, Cehelsky JE, Vaishnav A, Wilkinson T, Meyers R, Harrison LM, Roddam PL, Walsh EE, DeVincenzo JP: Effect of Preexisting Serum and Mucosal Antibody on Experimental Respiratory Syncytial Virus (RSV) Challenge and Infection of Adults. *J Infect Dis.* 2015; **212** (11): 1719-25 [PubMed Abstract](#) | [Publisher Full Text](#)
2. DeVincenzo JP, Wilkinson T, Vaishnav A, Cehelsky J, Meyers R, Nochur S, Harrison L, Meeking P, Mann A, Moane E, Oxford J, Pareek R, Moore R, Walsh E, Studholme R, Dorsett P, Alvarez R, Lambkin-Williams R: Viral load drives disease in humans experimentally infected with respiratory syncytial virus. *Am J Respir Crit Care Med.* 2010; **182** (10): 1305-14 [PubMed Abstract](#) | [Publisher Full Text](#)
3. Mulroy N, DeVincenzo J, Bagga B, Meals E, Harrison L, Sealy B, Hurwitz J, Benhabiles L, Cormier S, Hijano DR: Infant mucosal RSV-specific IgA response dynamics during and after infection. *Southern Society for Pediatric Research.* 2016.
4. Hijano D et al.: Neonates fail to mount a mucosal immune response against RSV infection. *St. Jude/PIDS Symposium.* 2016.
5. Yamin D, Jones FK, DeVincenzo JP, Gertler S, Kobiler O, Townsend JP, Galvani AP: Vaccination strategies against respiratory syncytial virus. *Proc Natl Acad Sci U S A.* 2016; **113** (46): 13239-13244 [PubMed Abstract](#) | [Publisher Full Text](#)
6. Miller CR, Shah A, Harrison L, Meals E, Hyrne C, Bagga B, Kim-Hoehamer Y, Benhabiles L, Cormier S, DeVincenzo J: The dynamics of respiratory syncytial virus (RSV) replication reveals age-related functional deficiency and maturation of the infant antiviral immune response. *Southern Society for Pediatric Research.* 2016.
7. DeVincenzo JP, El Saleeby CM, Bush AJ: Respiratory syncytial virus load predicts disease severity in previously healthy infants. *J Infect Dis.* 2005; **191** (11): 1861-8 [PubMed Abstract](#) | [Publisher Full Text](#)
8. El Saleeby CM, Bush AJ, Harrison LM, Aitken JA, DeVincenzo JP: Respiratory syncytial virus load, viral dynamics, and disease severity in previously healthy naturally infected children. *J Infect Dis.* 2011; **204** (7): 996-1002 [PubMed Abstract](#) | [Publisher Full Text](#)

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Competing Interests: No competing interests were disclosed.

Author Response 14 Mar 2017

James Nokes,

Responses to John DeVincenzo

1. Relatively non-quantitative collection technique: Unfortunately, because of the nature of the study (outpatient from village homes), the collections themselves had to be performed using a swab technique. Swab collections (even if done using a deep swab approach) have been shown to produce specimens that are difficult to use for quantitative virology. Variations in depth of swab acquisition, variations in depth of the transition zone from squamous to respiratory epithelial cells within the nose, reduced sensitivity caused by the small volume collection and the large volume dilution required during the nucleic acid extraction technique, and sample to sample variations of the PCR technique (which can not be overcome by re-running the specimen because of lack of sufficient collection volume) all contribute to creating this relatively non-quantitative collection technique.

We agree that virus quantification from nasopharyngeal swab collections is imperfect and will result in variation arising from the collection technique. This variation is likely to be random and hence should not alter the patterns with respect to covariables, but make observations more noisy. We hope readers understand the limitations and the fact that there are no better collection alternatives.

2. Relatively non-quantitative PCR techniques employed: Molecular quantification of viruses within a sample requires a standard curve to assign a quantitative value based upon a Ct value. One cannot use the Ct value itself without it being read from a standard curve. These standard curves can be created externally from the PCR run itself (less desirable) or can be created internally during the actual run which includes the unknown samples from the patients themselves (more desirable). The creation and design of the standard curves is extremely important in determination of quantification and can be based on whole viral genome from infected cells (thereby controlling for the nucleic acid extraction and purification steps, and controlling for the reverse transcriptase step). Less optimal (but much easier to construct) standard curves can be created that utilize cloned cDNA (easily giving a copy number, but not properly controlling for either the nucleic acid extraction, purification, nor reverse transcription steps). Unfortunately, this study apparently has not constructed standard curves, and they rely solely on generated Ct values to imply quantity. It is, in some cases, acceptable to do so when using a single-plex assay (where relative viral concentrations can fairly be compared to the same technique on different specimens). But this approach is never acceptable to use when multiplexed assays are employed (see below).

While we do not disagree with the reviewer that more accurate quantification of viral RNA is possible, there is no disputing the negative consistent relationship between starting quantity and cycle threshold. Inter-assay variation was assessed through repeated positive controls which were monitored each run. There is, of course, the underlying high degree of underlying biological variation referred to earlier which diminishes the value of downstream precision in RNA load estimation. Hence we recognise we are using only crude estimates, but these were the data available and they produce results which look epidemiologically sensible.

3. Introduction of technique-driven systemic bias in molecular quantification: This molecular quantification bias comes in two forms:

A) multiplex-PCRs detect viral amplicon differently when there are two molecular targets within the same sample based on competition for ions, primers, and probes, and enzymes within the reaction.

B) The relative amplification efficiency (and the probe-binding-induced fluorescent signal read out by the PCR machine, will be different depending on the probe and the amplicon size and the relative binding efficiencies of the forward and reverse primers. This makes it impossible to quantitatively assess the quantification of two different viruses (RSV-A and B) within a mixed sample and also makes it impossible to compare one target's quantity (RSV-A) with another target's quantity (RSV-B).

The observed increase in AUC for co-infections of RSV A and B relative to single infections is perplexing, so we thank the reviewer for the insights regarding interpretation of multiplex PCR data.

4. Failure to evaluate infectious viral particles and using non-infectious viral particle counts to make conclusions about infectious viral load. The authors find and conclude that: "The majority of virus available for transmission during an RSV epidemic appears to arise from individuals in their first year of life and therefore undergoing their first RSV infection episode." Although I agree with this statement, it is important to realize that from data within this manuscript itself, this statement assumes that infectious viral particles equate to quantitative PCR copy number. This assumption is certainly not true. Infectious virus particles are best quantified by a quantitative culture approach. Factors within the respiratory secretions themselves (such as IgA) have been shown to be associated with RSV culture negativity despite persistent high concentrations of RSV genome continuing to be detected within human respiratory secretions after a documented RSV infection. Infant's ability to generate RSV-specific IgA is significantly less than the ability of adults to do so during an RSV infection. The author's conclusion that infants likely contribute more than do adults to epidemic spread of RSV infection has been carefully modeled in a recent paper in which two key factors; infectious viral particle quantity over time, and quantitative measures of human contact activity during times of infection, were incorporated into the model. The implications of these factors promoting RSV community spread should influence the types of RSV vaccines to be developed and the age ranges targeted by these vaccines, and these different vaccine scenarios have been modeled.

We accept that a limitation of the original study was the absence of virus culture. A caution on inferring virus infectiousness using the PCR quantities is explicit in the paper. While further work will be required to link quantitative PCR copy numbers and infectiousness it is worth noting in a carefully designed human challenge studies viral load dynamics by qPCR mirror that of quantitative culture with qPCR consistently overestimating viral load mainly in the recovery phase. This provides evidence of direct proportional relationship between quantitative culture (a measure infectious virus) and qPCR (measuring both infectious and non-infectious viruses) (DeVincenzo, Wilkinson et al. 2010) hence our interpretation albeit with caution.

5. Questions regarding viral load area under the curve calculations. Any area is geometrically and mathematically defined by multiplying two linear dimensions together. The same is true for viral load area under the curve (in this paper it is called viral density area under the curve for some reason). Viral load AUC is Viral load (units) multiplied by Time (units). The resulting units of viral load are always containing the unit of time. (For example "log PFUe x days", or "Copies x hours"). The units of viral load AUC in this paper are incorrect and need to be redefined. When one realizes the mathematical definition of AUC, then it is no surprise that Fig 2a and 2b show a lot of correlation between duration of shedding and AUC (because time (duration) is actually in the denominator of the AUC value).

Thanks for pointing this out. We have changed the units to log 10 copies x days.

6. Problems of determining co-infection with RSV-A and RSV-B. Students of RSV need to be very careful in reporting and interpreting data regarding co-infections. This paper is no exception. This paper and others assume that when a multiplexed assay reports that the wavelength of the excited probe is orange and another wavelength of the excited probe is green (I am giving a hypothetical example) that this means that both RSV-A and B are present in the sample. However this may or may not be the case. For example, if the virus is an RSV-A virus, the probe for RSV-A binds relatively tightly to the amplicon through strict base-pair matching, and then cleavage of the reporter molecule from the quencher molecule allows the detection of the amplicon (through inhibition of FRET). However, it is very possible that the RSV-B probe will also bind to the amplicon at a certain number of base pairs, and this can result in the RSV-B reporter being detected (erroneously causing the authors to read that RSV-B is also present in the sample). Because of this problem, this paper's results, analyses, and conclusions regarding "co-infections" of RSV-A and B need to be taken with caution.

Co-detection of RSV A and B were in 19 NPS collections yielding to 14 different infection episodes in 7 households. 4 out of the 7 households had more than one co-infection episode. While acknowledging the challenges of interpreting co-infection data, the outgoing observations of clustering of the co-detections by individual and co-infection episodes by households does not point to spurious findings.

7. Problems with co-infections with RSV and other respiratory viruses (Adeno, Rhinoviruses, and Coronaviruses). RSV infection causes massive shedding of respiratory epithelial cells into the airway lumen (which are then collected by the swab technique). Any persistent virus present in those shed cells will therefore be detected at higher frequency during an RSV infection. This is an alternative explanation of the correlation between viral load of RSV and detection of these other infections (Top of Pg. 8). This phenomenon carries much different implications than are assumed to be present by these authors.

The reverse argument is also possible. A higher proportion of shed cells would be infected with RSV. Unless concurrent infections were actually in the very same cells (rather than infecting other cells) then this sloughed material would actually have less co-infection than the remaining intact cells. In addition, the enhanced sloughing of the epithelial cells would have been prominent in symptomatic infection episodes where secretions will have 'washed' a larger surface area and the later is adjusted for in our analyses.

8. Associations with age and outcome (area under the curve viral load). The authors do an excellent job of assessing the associations between area under the curve and other factors. After observing the problems with the actual viral load area under the curve measurements themselves, (see points 1,2,3,5 above) it appears that their data shows that the primary host characteristic that affects area under the curve viral load is whether or not children have had a prior RSV infection. This fits with other data from the literature. However, it would be nice to know whether young age within the infant age category was predictive of increased area under the curve viral load. Unfortunately, the data set does not contain data on age in weeks, days, or months within the first year of life. I suspect that younger ages within the infant age range, i.e. the RSV-naïve population, will be related to the area under the curve viral load. Other databases will likely need to

be leveraged in order to determine this important question. Although the number of infant infections in this database will likely be small, the authors should attempt to break down the ages within the infant age range in order to try to answer this important question. The infant antiviral functional immune response is likely not fully developed at birth, and RSV can be used to interrogate the ontogeny of this immune response, which can have major implications on vaccine development.

A very good suggestion though our analyses showed no apparent relationship between age and AUC for infants under 1 year. This is why we did not break down age further for this category.

9. Associations with other factors and outcome (area under the curve viral load). The manuscript also indicates that presence of respiratory symptoms, intra-household acquisition of infection, being a first infection episode in the season for an individual, and having a co-infection were associated with increased amounts of RSV shed. We should think about the root causes of these associations so that we can understand principles of RSV infection. I do not know the intricacies of the subject acquisition and identification, but is it possible that intra-household transmission simply is a surrogate marker for earlier swabbing of the patient? We know that viral load peaks in the nose shortly after symptom onset, and that viral load declines thereafter. Therefore, having a first swab earlier would create a subject with a greater viral load AUC. An alternative explanation is that perhaps an inoculum effect is present and that children are inoculated within a household transmission with greater amounts of virus, thus causing greater AUC viral load. Past studies have tried to determine an inoculum effect for RSV and crowding does not seem to be a predictor of viral load. Likewise, in experimental RSV infection models of adults, there appears to be little effect of inoculum on AUC viral load. It is likely that RSV inoculum affects whether a person becomes infected, but once that person is infected that inoculum has little effect on viral load, AUC or disease severity. I have addressed the issue of co-infection (and alternative explanations for the observed data) in points 6 and 7 above.

Swabs were collected regardless of symptom status and days of collections were planned to coincide for all members in same household unless someone was away. Therefore swabbing in early phase of infection (when viral load was high) would have occurred by chance and would influence the observed associations towards null.

10. Waning control of RSV infection with advanced age. Perhaps the greatest contribution of this paper to the medical literature is the finding that area under the curve viral load begins to rise again in adulthood after the age of 40. This is a fascinating finding, and could result from either immune senescence, or reduced exposure to boosting infections of RSV, or a combination of both. Whether this adult age-related increase in RSV area under the curve viral load is similar in developed countries will need to be evaluated.

Indeed this is an intriguing observation. We have interpreted this finding with caution since we had rather small sample size for these age group i.e. 9 episodes from 8 infected participants.

DeVincenzo, J. P., T. Wilkinson, A. Vaishnav, J. Cehelsky, R. Meyers, S. Nochur, L. Harrison, P. Meeking, A. Mann, E. Moane, J. Oxford, R. Pareek, R. Moore, E. Walsh, R. Studholme, P. Dorsett, R. Alvarez and R. Lambkin-Williams (2010). "Viral load drives disease in humans experimentally infected with respiratory syncytial virus." *Am J Respir Crit Care Med* 182(10): 1305-1314.

Competing Interests: No competing interests were declared.

Referee Report 03 January 2017

doi:[10.21956/wellcomeopenres.11076.r18492](https://doi.org/10.21956/wellcomeopenres.11076.r18492)



Peter F. Wright

Department of Pediatrics, Geisel School of Medicine, Dartmouth College, Hanover, NH, USA

The report by the KEMRI team builds on a cohort of 47 households followed through an RSV season in coastal Kenya with frequent nasopharyngeal swabs for detection of RSV. The introduction and rationale for the study are very clearly provided.

More complicated is the analysis to determine minimum and maximum areas under the curve representing estimates of the total amount of virus shed based on a PCR cycle to detection of virus. Once this is accepted the analysis of factors influencing the amount of virus detected is straightforward and indicates age, symptoms, RSV type B, and whether infection is the initial infection during the sampling period. Each of these observations remained significant in a multiple linear regression analysis. None of these observations are surprising but all are of interest in this locale and within a community based study design. Might the large number of asymptomatic individuals include some who subsequently became symptomatic and were at the time of sampling in the incubation period of their illness?

The authors address some limitations of the study. Other things that might be considered include an analysis based gradation of symptoms in terms of virus recovered, whether frequency of sampling was influenced by symptoms, and patterns of spread within families. The latter is of particular interest in that I suspect that with many viral infections the peak titre will influence spread more than the total amount shed. The number of days to detection of secondary infections in the household might give a clue to the validity of this hypothesis.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: No competing interests were disclosed.

Author Response 25 Jan 2017

Miriam Wathuo, KEMRI Wellcome Trust Research Programme, Kenya

1. Might the large number of asymptomatic individuals include some who subsequently became symptomatic and were at the time of sampling in the incubation period of their illness?

Being symptomatic or not was defined both by individuals and by episodes.

An episode was described as asymptomatic if the individual never had any respiratory illness symptoms during the entire duration of the infection episode. An episode was described as symptomatic if the individual had respiratory illness symptoms at least once during the episode.

An individual was described as symptomatic if they had respiratory illness symptoms at least once during the study period, and asymptomatic if they never had any respiratory

illness symptoms throughout the study period.

Of the 71 individuals with asymptomatic episodes, 19 (26.8%) subsequently developed symptoms within one week. The subsequent symptoms were not linked to any virus detection in 15 (78.9%) of the 19 individuals.

2. Other things that might be considered include an analysis based gradation of symptoms in terms of virus recovered, whether frequency of sampling was influenced by symptoms, and patterns of spread within families.

Sampling was done irrespective of symptom status and the frequency of sampling was not influenced by symptoms. Mild symptoms (mainly cough and running nose) were observed among the symptomatic cases hence difficult to grade the disease severity – no severe or very severe pneumonia cases. Nevertheless, there was no association between symptom status and virus type. The patterns of spread within households and role of asymptomatic cases has been discussed in detail in previous papers:

1. ***Munywoki PK, Koech DC, Agoti CN, et al. The source of respiratory syncytial virus infection in infants: a household cohort study in rural Kenya. [electronic article]. 2014;209(11):1685–92.***
2. ***Munywoki, P. K., D. C. Koech, C. N. Agoti, A. Bett, P. A. Cane, G. F. Medley and D. J. Nokes (2015). "Frequent Asymptomatic Respiratory Syncytial Virus Infections During an Epidemic in a Rural Kenyan Household Cohort." *J Infect Dis* 212(11): 1711-1718.***
3. ***Munywoki, P. K., D. C. Koech, C. N. Agoti, N. Kibirige, J. Kipkoech, P. A. Cane, G. F. Medley and D. J. Nokes (2015). "Influence of age, severity of infection, and co-infection on the duration of respiratory syncytial virus (RSV) shedding." *Epidemiol Infect* 143(4): 804-812.***

Competing Interests: No competing interests were disclosed.

Referee Report 30 December 2016

doi:[10.21956/wellcomeopenres.11076.r18488](https://doi.org/10.21956/wellcomeopenres.11076.r18488)



H. Rogier van Doorn¹, Lien Anh Ha Do²

¹ Oxford University Clinical Research Unit- Hanoi, Wellcome Trust Major Overseas Programme, National Hospital of Tropical Diseases, Hanoi, Vietnam

² Murdoch Children's Research Institute, Melbourne, VIC, Australia

This is an interesting paper using area under curve (AUC), derived from non-quantitative PCR data using three different estimates, as a unit to measure viral shedding during an RSV season 8 years ago for a prospective cohort of 47 households in rural coastal Kenya. AUC can take variability of viral load due to different factors (technical, biological) into account and reflects the overall temporal changes of virus shed.

The authors use three models for calculating the AUC. While this is not my expertise, I cannot comment on these models, but I strongly believe that choosing an appropriate model is one of the most crucial parts of this study, any wrong assumption/ formula of these models will affect the results and its interpretations.

Hence, the introduction of AUC as a measurement unit for transmission studies is an interesting concept, however, it may need a validation step using data from animal models, human challenge models, or clinical data monitoring RSV load.

Below are some points requiring more clarification:

1. The interval between samples is 3-4 day, why did the author choose this big interval, while previous human challenge studies have suggested daily sampling or every 2 days could be the best follow-up (1). This is a major limitation as samples are taken only every 3-4 days and amounts of shed virus are therefore modelled based on little data and many assumptions, which may or may not equal out. The authors should definitely include more information on how their semi-quantitative data was calibrated.
2. As authors didn't use qPCR and use of a standard curve is not described, authors should clarify how they ensured PCR efficiency was stable between runs and what was used to calibrate runs and Ct values. This is a crucial thing to add, as all results in this paper are based on (semi)quantitative measurements and therefore, proper quantification is absolutely key.
3. At the beginning of the Methods section, the authors specified that the AUC was calculated using viral load in log scale, but later for the midpoint AUC, they used Ct-value, please clarify.
4. Reference 21 and 23 are not accessible
5. Without baseline demographics about the sampled cohort I find some of this information hard to interpret. Symptomatic infections would also be interesting to see by age group.
6. In Table 1, it is not clear how the total AUC was calculated. Surprisingly, the number of RSV infection episodes in the young age group (less than 1 year old and less than 5 years old) is much lower (denominator?) than those in the age group of 5-15 years old, while many epidemiological data have shown the re-infection rates are higher in the young age group. There is no comment/discussion on this result.

In addition, the number of RSV infection episodes and the AUC in "living with smokers" group were lower and similar, respectively, compared to those who do not live with smokers. This result is controversial with the common observations from many epidemiological studies on RSV susceptibility and there is no comment/discussion regarding this.

Specific minor comments:

1. P3c1p2: The comparison with human volunteer studies to suggest load / inoculum may be associated with infection is not very strong, this may only be true in the background of a well-developed immune history against RSV whereas the main burden of disease of RSV is in infants who will not have this history yet.
2. P3c1p3: I don't think excretion is proper terminology for the process of virus shedding (nor is viral density – which is used later on, p6).
3. P3c1p4: the number of sampled household members and the inclusion criteria should be added here.

4. P3c2p1: why were only adeno, corona and rhino targeted with the diagnostic PCR and not the influenzas, parainfluenzas and hMPV? This is a very limited definition of co-infection.
5. P6: Please clarify why data are presented as median (IQR) amounts of virus shed in one paragraph but then as mean amounts (95%CI) of virus shed in the other.
6. P8c1p2: In Table 2 sex is not significant, but its mentioned here as an associated factor.
7. P8c2p3: the result that timing within the season was (appeared?) associated with the amount of virus shed is presented in the discussion section, shouldn't this also be presented as a result, with data and a p-value?
8. P8c2p3: why is it of particular importance that 90.3% of virus shed is from individuals experiencing their first episode? Also, 179/208 episodes studied were first episodes.
9. Table 2: the number of primary and secondary cases per age group would be informative here

In summary, this paper reports important data on RSV transmission in a community setting; data on this subject is still scarce. This data is relevant for RSV vaccine and treatment development and can help suggest target groups for vaccination and implications of infection control measurement.

References

1. DeVincenzo JP, Wilkinson T, Vaishnav A, Cehelsky J, Meyers R, Nochur S, Harrison L, Meeking P, Mann A, Moane E, Oxford J, Pareek R, Moore R, Walsh E, Studholme R, Dorsett P, Alvarez R, Lambkin-Williams R: Viral load drives disease in humans experimentally infected with respiratory syncytial virus. *Am J Respir Crit Care Med*. 2010; **182** (10): 1305-14 [PubMed Abstract](#) | [Publisher Full Text](#)

We have read this submission. We believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however we have significant reservations, as outlined above.

Competing Interests: No competing interests were disclosed.

Author Response 25 Jan 2017

Miriam Wathuo, KEMRI Wellcome Trust Research Programme, Kenya

1. The interval between samples is 3-4 day, why did the author choose this big interval, while previous human challenge studies have suggested daily sampling or every 2 days could be the best follow-up (1). This is a major limitation as samples are taken only every 3-4 days and amounts of shed virus are therefore modelled based on little data and many assumptions, which may or may not equal out. The authors should definitely include more information on how their semi-quantitative data was calibrated.

We agree the sampling interval was not optimal for this type of analysis. However, it was not logistically possible to collect deep nasopharyngeal swabs from ~500 participants at a more intense frequency for six months. Furthermore, community acceptability of more frequent sampling was an important issue.

The following rationale was used to determine the twice-weekly (every 3.5 days) nasal sampling frequency. Assuming individuals shed virus with mean duration of between 3.5

and 9 days, with a constant rate of recovery from shedding, and an onset on average half way between any sampling interval, then the proportion of individuals predicted to remain shedding, and thus detectable, will range from 61%-82% (for 3.5-9 days duration) for a 3.5 day sampling interval. Given the need to detect infection in mild cases and in older children and adults with a likely lower range of shedding duration, sampling twice weekly was indicated.

2. As authors didn't use qPCR and use of a standard curve is not described, authors should clarify how they ensured PCR efficiency was stable between runs and what was used to calibrate runs and Ct values. This is a crucial thing to add, as all results in this paper are based on (semi) quantitative measurements and therefore, proper quantification is absolutely key.

This is a retrospective analysis of household RSV infection data and we share the referee's concerns about limitations on the use of Ct values. A qPCR would have been the best method to quantify viral load. However, a standardized multiplex RT-PCR protocol was used which included positive controls monitored each run to ensure their Ct values did not vary significantly.

3. At the beginning of the Methods section, the authors specified that the AUC was calculated using viral load in log scale, but later for the midpoint AUC, they used Ct-value, please clarify.

Viral load for calculation of minimum, midpoint and maximum AUC was presented in \log_{10} copies. Ct values are mentioned to explain how viral load for days in between collection days was estimated before being converted to \log_{10} copies. The original viral load values were in Ct values, which were then converted to \log_{10} copies and used to calculate AUC in all three scenarios (minimum, midpoint and maximum AUC). Please also see comments in response to Referee Cowling.

4. Reference 21 and 23 are not accessible

Thank you. We have rectified the errors in the two references. Reference 21 should be: 'Kreide, Donald; Lahr D. Trapezoid Rule and Simpson's Rule [Internet]. 2010. Available from: <https://math.dartmouth.edu/~m3cod/klbookLectures/406unit/trap.pdf>,

while reference 23 is the study dataset.

5. Without baseline demographics about the sampled cohort, I find some of this information hard to interpret. Symptomatic infections would also be interesting to see by age group.

We have provided a brief summary of the baseline characteristics in the first paragraph of the results section. However, more details could be accessed in our earlier publications arising from this study whose references are provided. Figure 4 of this manuscript provides distribution of symptomatic and asymptomatic infections by age.

6. In Table 1, it is not clear how the total AUC was calculated. Surprisingly, the number of RSV infection episodes in the young age group (less than 1 year old and less than 5 years old) is much lower (denominator?) than those in the age group of 5-15 years old, while many epidemiological data have shown the re-infection rates are higher in the young age group. There is no comment/discussion on this result.

Total AUC is the cumulative AUC for all infection episodes in a specified category eg age group. The higher number of infection episodes in the 5-15 year age group is most likely due to the higher number of individuals in this age category, and this translated to the higher total of the amount of virus shed by this age group. To compare the transmission potential per person by age group, an average amount of virus shed per infection episode might be a better statistic rather than the total AUC. However, from the perspective of contribution to community transmission then the cumulative AUC seems appropriate.

7. In addition, the number of RSV infection episodes and the AUC in “living with smokers” group were lower and similar, respectively, compared to those who do not live with smokers. This result is controversial with the common observations from many epidemiological studies on RSV susceptibility and there is no comment/discussion regarding this.

In this rural community, indoor smoke from solid fuels (mainly firewood and charcoal) was ubiquitous and likely to conceal any effects of smoke from cigarettes. In addition, we only had 18 smokers out of the 493 participants. So, relatively few people lived with smokers. Also, we cannot be sure how much contact members of a household had with a smoker and if the smoker did smoke in the house or outside. A comment has been added in the discussion to that effect.

Specific minor comments:

1. P3c1p2: The comparison with human volunteer studies to suggest load / inoculum may be associated with infection is not very strong, this may only be true in the background of a well-developed immune history against RSV whereas the main burden of disease of RSV is in infants who will not have this history yet.

This limitation in extrapolating results on association of infection acquisition and exposure and viral load from human challenge studies has been noted and discussed in the main text.

2. P3c1p3: I don't think excretion is proper terminology for the process of virus shedding (nor is viral density – which is used later on, p6)

Well noted. We have now used viral shedding consistently in the text.

3. P3c1p4: the number of sampled household members and the inclusion criteria should be added here.

This has been described in a previous paper as stated in the methods section

4. P3c2p1: why were only adeno, corona and rhino targeted with the diagnostic PCR and not the influenzas, parainfluenzas and hMPV? This is a very limited definition of co-infection.

Adeno, corona and rhino were the common respiratory viruses detected in a preliminary screening of the samples from 5 randomly selected households and taken forward for the screening of rest of the samples from the remaining households. We thus have limited

data to comment on the other respiratory pathogens.

5. P6: Please clarify why data are presented as median (IQR) amounts of virus shed in one paragraph but then as mean amounts (95%CI) of virus shed in the other

The AUC values were still slightly skewed even after log transformation. So, we presented medians in the descriptive analysis because it is the best measure for a skewed variable. However, the linear regression analysis coefficients are means not medians, so for the linear regression results, we had to present means and 95% CI.

6. P8c1p2: In Table 2 sex is not significant, but its mentioned here as an associated factor

We chose to adjust for sex a priori because sex has previously been considered an important risk factor to RSV infections. In the final model presented sex was not a significant covariate though.

7. P8c2p3: the result that timing within the season was (appeared?) associated with the amount of virus shed is presented in the discussion section, shouldn't this also be presented as a result, with data and a p-value?

This was a comparison of viral load values (not amount of virus) against time just to show the reader a visual of how viral load varied depending on the month of the year. This was to demonstrate that there is an increase in the number of positive viral load detections during the RSV season. This has been mentioned in the results and discussion sections.

8. P8c2p3: why is it of particular importance that 90.3% of virus shed is from individuals experiencing their first episode? Also, 179/208 episodes studied were first episodes

First episode in this case refers to the first infection detected over the study period, which were 179 (86%). As stated earlier, the total AUC gives an idea of which groups contribute most to community transmission of virus. It is, therefore, important to note that most virus shed comes from individuals experiencing their first episode of the RSV season.

9. Table 2: the number of primary and secondary cases per age group would be informative here

Thank you. This has been added.

Competing Interests: No competing interests were disclosed.

Referee Report 23 December 2016

doi:[10.21956/wellcomeopenres.11076.r18490](https://doi.org/10.21956/wellcomeopenres.11076.r18490)



Benjamin J. Cowling

WHO Collaborating Centre for Infectious Disease Epidemiology and Control, School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong (HKU), Hong Kong, China

This is an interesting manuscript about RSV shedding, in the context of household transmission. There are few data on this topic in the literature and this is a valuable contribution. I thought the manuscript was clear and well written and should be indexed. I only had one comment:

1. Please comment on the choice of log viral load in analyses

In Figure 1 please clarify the y-axis metric and you might like to explain this choice as well. Is it viral density (copies), or log₁₀ copies ?

It seems as if you are calculating the AUC based on log₁₀ viral copies but what is the biological interpretation of that? I was wondering whether the number of viruses could be a better correlate of transmission, not their log?

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Competing Interests: No competing interests were disclosed.

Author Response 25 Jan 2017

Miriam Wathuo, KEMRI Wellcome Trust Research Programme, Kenya

1. Please comment on the choice of log viral load in analyses

The absolute values (viral RNA copy numbers) for AUC were not normally distributed and were log transformed to stabilize the variances providing a suitable outcome variable for use in linear regression analysis. The mean (SD) for the untransformed and log-transformed values is 3.73×10^7 (1.34×10^8) and 6.07 (1.51), respectively. The choice of outcome has been clarified in the manuscript.

We have re-analysed the data by the alternative route of summing the absolute viral load values and then taking logs to estimate the AUC. The results are not substantially different from the current method which is a summation of the log-transformed values to estimate AUC. This is stated in the discussion.

2. In Figure 1 please clarify the y-axis metric and you might like to explain this choice as well. Is it viral density (copies), or log₁₀ copies?

Log₁₀ viral RNA copies as they are the units used in measuring the viral load.

3. It seems as if you are calculating the AUC based on log₁₀ viral copies but what is the biological interpretation of that? I was wondering whether the number of viruses could be a better correlate of transmission, not their log?

The relationship between viral shedding and infectiousness is unknown. The relationship between amount of virus shed and risk of infection might be linear or have a threshold effect (e.g. minimum dose), in which case a logarithmic scale would be more appropriate. The log-transformation allowed us to stabilize the variance of the outcome values, suggesting that the measurement error of amount of virus might vary non-linearly. The final conclusions (biological interpretation of determinants of amount of virus spread) is similar to when using the arithmetic values since an increase in log values is also linked to an increase in absolute values albeit at exponential scale.

The relationship between various factors such as age, symptoms, etc and untransformed amount of virus is the same as that between those factors and the log-transformed amount of virus. Using log-transformed values does not affect the final conclusions. Determining the relationship between measures of virus shed and risk of transmission requires a different analysis which we are undertaking. This is now reflected in the Discussion.

Competing Interests: No competing interests were disclosed.
