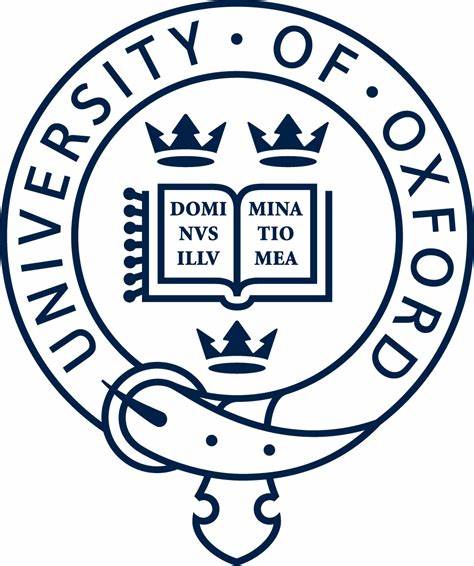
**Title**

**Estimating the time-dependent reproduction number for Ebola Virus Disease in an age-stratified host population**

**Candidate number 1079733**

**University of Oxford**



**Explanatory Narrative**

***Journal Name:*** Science Advances

As a graduate student involved in research on modelling for global health, the journal suitable for publishing my findings needed to have a focus on publishing multidisciplinary and impactful research papers. While choosing a reputable journal with a strong impact factor was important, I also needed it to have open access options so that my work would be accessible to a broader audience, including practitioners and policy-makers. The possibility of flexible article formats as well as the journal’s openness to rapid publication timelines was a significant advantage too. The manuscript content’s relevance to the area of study and its alignment with the journal’s scope is outlined further in the narrative.

The great importance of being able to accurately estimate the time-dependent reproduction number, that is the expected number of secondary cases caused by each infected individual at time *t*, during infectious disease outbreaks has become evident during recent years. Not all existing methods, however, account for the presence of age heterogeneity across a host population. Throughout this research project we aimed to investigate whether employing a two-host model, and thus differentiating between child and adult cases, is a more powerful framework for estimating the time-dependent reproduction number compared to a previously formulated one-host model. The focus was on the transmission of Ebola Virus Disease (EVD).

The two renewal equation models were tested with both computer-generated and the real-world data from the 2020 EVD outbreak in Équateur Province, Democratic Republic of the Congo, before and after introducing either age-specific serial interval distributions or pathogen transmissibility (the latter being tested with simulated datasets only).

Our findings suggest that, indeed, the two-host model is able to produce more accurate estimates of the time-dependent reproduction number early on in an outbreak when either age-specific serial interval distributions or pathogen transmissibility is introduced. For the computer-generated dataset, the one-host model was shown to improve its performance if the population serial interval distribution was parameterised correctly to account for the age-specific serial interval distributions.

While the conducted analyses resulted in valuable insights, there are limitations related to how readily available an estimate of the serial interval distributions is early on in an outbreak and the estimates used for the scaling of a standard contact matrix to obtain an effective contact matrix being context-sensitive. These are to be addressed in any further research.

Overall, this research project highlights that employing methods which account for age heterogeneity when estimating the time-dependent reproduction number early on in an infectious disease outbreak can be more reliable in a number of contexts. This is particularly beneficial for implementing and/or relaxing any relevant control measures both effectively and efficiently.

**Abstract**

During an infectious disease outbreak, it is crucial to obtain reliable estimates of quantities characterising pathogen transmission so that interventions are implemented in an effective and efficient manner. One such quantity is the time-dependent reproduction number, *Rt*, which is defined as the expected number of secondary cases caused by each infected individual at time *t*. Even though there exist several methods to estimate it, some of them do not account for heterogeneity across the host population. Here a two-host renewal equation model was formulated that relies on the numbers of observed new cases at successive times during an outbreak, an estimate of the serial interval distribution and a contact matrix. The model was then tested against a previously described one-host model with a computer-generated dataset and a dataset from an outbreak of an Ebola virus disease. For the computer-generated dataset we were able to demonstrate that: firstly, when the serial interval distribution and the pathogen transmissibility were not age-specific, the one-host model was able to estimate *Rt* as accurately as the two-host model and secondly, when age-stratified effective contact patterns were considered in combination with age-specific serial interval distributions the one-host model overestimated *Rt* early on in the outbreak and was outperformed by the two-host model. It was also found that the one-host model performance improved significantly when the population serial interval distribution was parameterised correctly. For the real-world dataset, it was shown that the two models produced very similar results when no difference was assumed between the mean serial intervals of the host types. When age-specific serial intervals were introduced, the models were not in agreement. This suggests that the two-host model inference of transmissibility could potentially be both more accurate and meaningful in terms of relevant interventions, given sufficient data had been collected.

**Introduction**

Infectious diseases have been a major and recurring threat for global health across the world. The time-dependent reproduction number, *Rt*, along with other parameters governing pathogen spread, is an important measure when assessing the effectiveness of control efforts and deciding whether additional interventions are required (1). A number of formal definitions for *Rt*  and methods to estimate it in real-time during epidemics have been proposed (2). One of the estimation approaches was developed by Wallinga and Teunis (2004). They applied their method to data from the 2003 SARS epidemic and showed that the time-dependent reproduction number decreased after control measures were implemented. The approach considers all possible transmission trees consistent with the observed epidemic data, and generates an estimated value of the time-dependent reproduction number at each timestep with observed cases. This method has been applied to estimate reproduction numbers during epidemics of diseases including, for example, Ebola virus disease (3,4). It has been extended to also permit inference in populations of multiple host types (5). Another method employing branching processes for estimating the time-dependent reproduction number was described by Cori et al. (2013) (6). Similarly to Wallinga and Teunis (7), two inputs are required: the serial interval distribution, i.e., the distribution of delays between symptom onset times of infectors and their infectees, and a disease incidence time series.  A potential drawback of the method arises from the assumption of homogeneity across all incident cases. Subsequent work has been done to address this, by considering a scenario in which imported cases are differentiated from cases that acquired infection locally (8). The effect of heterogeneity in the onwards transmission risk between local and imported cases was then explored further (9). Using COVID - 19 data from different regions worldwide, it was shown that different assumptions about the relative transmission risk between imported and local cases have a substantial effect on *Rt*estimates and consequently on implementation of interventions. The study hence emphasizes the need to collect data during outbreaks describing heterogeneities in transmission between different infected hosts, and to account for these heterogeneities in methods used to estimate *Rt*.

The model investigated in this report incorporates a different kind of heterogeneity between cases: specifically, age heterogeneity. The aim thus was to explore when, provided there is sufficient data, it is needed to employ the two-host model and when a previously formulated one-host model (6) might be sufficient to monitor the transmission of an infectious disease. The focus was on the transmission of Ebola Virus Disease (EVD), and we used data from the 2020 outbreak in Équateur Province, Democratic Republic of the Congo (DRC).

EVD is a haemorrhagic fever, usually fatal if left untreated, and is caused by the Ebola Virus (EV). Zoonotic spillover occurs when a human comes in close contact with blood or saliva of an infected animal. EVD is then spread by human-to-human transmission via direct unprotected contact with a body fluid such as saliva, urine, faeces or semen as well as with items contaminated by the body fluids of an infected person e.g., clothes. Burial practices that involve direct contact with an infectious dead body are another major driver of the spread of EVD (10). Such practices very often include washing and cleaning of the corpse. High viral loads and EV being able to sustain its viability for days after death make corpses of EVD cases extremely infectious (11). In DRC, similarly to other regions of West Africa, funeral rituals are performed by the relatives of the deceased and the local community, heightening the risk of transmission (12).

Since 1976, multiple EVD outbreaks have been recorded. The outbreaks were primarily located in remote areas of East and Central Africa. The DRC has been affected by more than ten EVD epidemics. The last outbreak in the province of Équateur was declared over on July 4, 2022 (13). Overall, the disease has caused hundreds of deaths across the country. Control measures for an Ebola outbreak include case surveillance, contact tracing, and promotion of safe burial practices. The DRC, similarly to most low- and medium-income countries, struggles with sufficient resources to adequately implement control measures against EVD outbreaks despite having acquired sufficient expertise in diagnostics, therapeutics and research activities (14).

**Methods**

*Data*

To test the methods both computer-generated data and daily EVD incidence counts, split by age between children and adults, from the 2020 outbreak in Équateur Province, DRC were used. Initially, the characteristics of the serial interval were taken from literature (15). Specifically, the continuous Ebola serial interval was assumed to be gamma distributed with mean 15.3 days and standard deviation 9.3 days. Then age-specific serial interval distributions were also considered. A discrete serial interval distribution was then obtained using the method outlined in Web Appendix 11 of Cori et al., 2013 and used for both computer-generated and real data.

The simulated datasets were generated from a two-host renewal equation model, described in the methods section. For every simulation, 10 initial cases were assumed in each of the age groups. Initially, the two age groups were assumed to have the same mean serial interval as well as the same infectiousness. Further in the report age-specific serial interval distributions were considered to explore several possible scenarios. 1000 simulations were performed whenever a model was tested.

*Effective Contact Matrix*

Interaction patterns between individuals are important to take into account when assessing the risk of contracting a directly transmissible infectious disease. Hence many mathematical models employ contact matrices to trace the spread of a certain pathogen. In this report, firstly, a synthetic contact matrix for DRC was taken from literature (16). The full contact matrix is a symmetric 16x16 matrix where each entry *Mij* corresponds to the average number of contacts an individual in age group *i* has with individuals in age group *j* at all locations each day. For the two-host model, the full matrix was converted into a 2x2 contact matrix C for children and adults by aggregating the relevant entries of the full contact matrix weighted by the population sizes. The population age composition is provided by The United Nations Population Division in five-year age intervals for the year 2020 (17).

Then for each entry of the new 2x2 matrix we have:

, where

and contain the indices of the smaller age groups being aggregated into larger age groups *i* and *j*;

*Mkr*– number of contacts an individual in age group *k* has with individuals in age group *r*;

*Pk*– population size of age group *k* (*k=1, 2,16*).

However, apart from community transmission, unsafe burial practices in West Africa cause a high after-death EVD transmission risk. As a result, standard contact matrices do not provide a complete picture of the risk of transmission. To account for this, the regular contact matrix C was adapted to include a higher proportion of transmission resulting in adult infectees (since the majority of transmissions at funerals involve adults becoming infected). Rather than using the contact matrix C in our analyses, we therefore use an effective contact matrix, Ceff.

From literature it was taken that ; *pd,a = 0.7*; *pd,c = 0.58*; *pf = 0.57* (18,19) , where *Rc* - the expected number of infections generated by a community case (excluding funeral transmission); *Rf* - the expected number of infections arising at an unsafe burial; *pd,a (pd,c)* – adult (child) case fatality ratio for community cases; *pf* - probability of unsafe burial given death in the community. It follows then that an adult or a child who undergoes an unsafe burial has 5.67 times as many effective contacts at the unsafe burial compared to when they are a community case. Around of adult cases and child cases undergo an unsafe burial. This gives

*Simulation two-host model*

A renewal equation model with two host types was formulated to estimate the time dependent reproduction number, *Rt*, on day given the incidence data and the discrete serial interval distribution. The host types considered were children (< 20 years) and adults (≥ 20 years). It was assumed that transmission within and between the age groups scales with the number of effective contacts. The number of cases each day was assumed to be drawn from a Poisson distribution:

where is the number of child cases on day , is the number of adult cases on day and is the total number of cases on day ; , are (discrete) serial interval distributions for children and adults, respectively; is proportional to the overall pathogen transmissibility on day and is assumed to be the same for both of the host types throughout the report (age-specific pathogen transmissibility is discussed in Supplementary File E).

In this model *Rt* is taken to be the dominant eigenvalue of the next generation matrix

, i.e., , where is the dominant eigenvalue of . Then equation (1) becomes

where

In principle, this model can be extended to any number of different age groups. If the elements of the effective contact matrix are all equal, i.e., no difference between contact patterns of host types is assumed, and the serial interval distribution as well as the pathogen transmissibility are not age-specific, the model is automatically reduced to a one-host model described in Cori et al., 2013:

where and is the population (discrete) serial interval distribution. Throughout the report we refer to it as the one-host model.

*Inference of Rt*

As in Cori et al., (2013), it was assumed that *Rt* is constant over a time period *[t-τ, t]*, where is the length of the time window over which *Rt* is estimated and is taken to be one week everywhere. Since it is assumed that is constant over a window of length days, the estimate on day is based on the number of infectees not only on day but also on the days prior to day . As such, the first day on which an estimate of can be generated is day is only defined from *t=2*). The probability of observing the incidence  , given the reproduction number *Rt* and conditional on the previous incidence data  , is

=

It is then possible to generate a distributional posterior estimate for *Rt* analytically by assuming that the prior for is a gamma distribution, conjugate to the Poisson likelihood, with shape parameter and scale parameter . In previous studies *a* and *b* have been set so that the prior for has mean and standard deviation both equal to 5 (6,8). The choice of large standard deviation and mean ensures a relatively uninformative prior and that meaningful posterior estimates of low or high values of *Rt* (e.g., *Rt <1* or *Rt >1*) are due to the data rather than the choice of a prior. The analysis results in the posterior distribution for *Rt* that is also a gamma distribution, with shape parameter and scale parameter :

The time-dependent reproduction number for children (adults) was defined as in Glass et al., 2011, i.e., as the average number of cases generated by a single infected child (adult) on day *t*. It is referred to as the age-specific time-dependent reproduction number throughout the report and is calculated as:

*Age-specific mean serial intervals*

Initially, the one-host and the two-host models were tested with children and adults having the same mean serial interval taken from literature. If such an estimate is taken from a trial that is based solely on transmission in adults, this would set the mean serial interval for children incorrectly. Hence for the one-host model the population mean serial interval distribution would need to be parameterised accordingly.

We note that:

,

where – number of cases in children during the nth generation of transmission (n = 1,2,…); – number of cases in adults during the nth generation of transmission (n = 1,2,…); – eigenvalues of the transposed effective contact matrix *Ceff (;* (i=1,2) – normalised eigenvectors of the transposed effective contact matrix Ceff ; – proportion of cases in children, – long-term proportion of cases in adults; *p* – pathogen transmissibility ;.

Then the (long-term) population serial interval distribution can be expressed as a weighted average , where , , are population, child and adult discrete serial interval distributions, respectively.

To explore the effect of having the population serial interval distribution corrected as above for the one-host model and to compare it with the two-host model, the mean child serial interval, was set to differ from the estimated adult mean serial interval, days, by a multiplicative factor of 0.25, 0.5 and 0.75.

Additionally, the effect of age-specific pathogen transmissibility on the one-host model performance with comparison to the two-host model was explored by having the models tested for several scenarios with incidence datasets generated by the two-host model. The details and the results can be found in the Supplementary File E.

To run the models RStudio version 4.3.1 was used. R packages used were: EpiEstim, sparsevar, zoo, ggplot2, dplyr, tydiverse, readxl. The code is provided in the supplementary materials.

**Results**

We begin by testing the one-host and the two-host models using simulated incidence datasets. The true underlying value of *Rt* was set to be 2.5 in the beginning of the outbreak and decreased gradually to 0.5. The incidence data was then generated by the two-host model with initial cases set to be 10 in both age groups and a sliding window of one week. The discrete serial interval distribution was first set to be the same for both age groups with a mean of 15.3 days and a standard deviation of 9.3 days. The effective contact matrix is:

The one-host and the two-host models were able to estimate the underlying values of *Rt*reasonably well and were in agreement with one another. As for the time-dependent age-specific reproduction numbers, *Rc,t* and *Ra,t*,the underlying values were 1.91 and 2.65 at the start and decreased to 0.38 and 0.53, respectively. Figure 1 shows these results. It can be noted that a wider credible interval is observed early on in the first half of the simulated outbreak for both of the tested models and that, because a sliding window is included in the inference of both models, the Rt estimates take some time to catch up with the underlying values after the decrease happens.

Next, age-specific mean serial intervals were considered and the two models were again tested with data generated by the two-host model. Figure 2 shows the results without the population serial interval distribution being corrected for the one-host model. Throughout the first half of the outbreak the one-host model overestimated the underlying value of *Rt* and was significantly outperformed by the two-host model. The error increased as the difference between the child and adult mean serial intervals increased. Figure 3 shows the results when the population serial interval distribution has been corrected as described in the methods section. Now the one-host model estimated the time-dependent reproduction number fairly accurately showing a noticeable improvement. The two-host model estimates were in close agreement with the underlying *Rt* values for all of the tested scenarios.

Finally, the methods were applied to real-world data from the 2020 Ebola outbreak in Équateur Province, DRC. Figure 4 presents the estimates yielded by the two models assuming no difference in mean serial intervals between children and adults. The one-host model estimates were very close to the two-host model estimates and ranged from 0.67 to 2.39. There are two periods during the outbreak when the *Rt* estimates are above 1, meaning the outbreak is estimated to not be under control during this time. The time-dependent age-specific reproduction numbers, *Rc,t* and *Ra,t* , were also estimated by the two-host model. Since an adult had a greater number of effective contacts compared to a child, the *Ra,t*estimates were higher than the *Rc,t*estimates ranging from 0.73 to 2.46 and from 0.53 to 1.77, respectively.

Similarly to the computer-generated data, age-specific mean serial intervals were tested with the real-world data. The same three scenarios were considered. Namely,  *= 0.75* , = 0.5 , and = 0.25 . The discrepancy between the two models was the greatest, although not particularly extreme, for the latter scenario whereas for the former one the estimates were in relatively close agreement. Correcting the population mean serial interval distribution did not result in reducing the discrepancy as effectively as for the simulated data. Figure 5 presents these results.

The results of introducing age-specific pathogen transmissibility are described in Supplementary File E and are presented in Figure S1.

**Discussion**

In this report we aimed to explore whether the two-host model might be more powerful at estimating *Rt* for EVD compared to the one-host model in the presence of age heterogeneity. For computer-generated data, we have discovered that the one-host model and the two-host model showed a similar accuracy in estimating *Rt* when the mean serial interval was not age-specific. However, the one-host model, unlike the two-host model, overestimated the underlying *Rt* value early on in the outbreak without the population serial interval distribution, , being corrected to account for age-specific serial intervals. The estimation accuracy of the one-host model improved if the correction was introduced. In Supplementary File E we also showed that the one-host model started to estimate *Rt* less accurately than the two-host model early on in the outbreak as the difference between child and adult pathogen transmissibility increased.

For the real-world data, when the same serial interval distribution was assumed for both host types, the two models showed no noticeable differences in their performance. However, differences did appear if children were taken to have a shorter mean serial interval than adults. Unlike for when the models were tested with the simulated data, even when the population mean serial interval distribution was corrected for the one-host inference, the discrepancy between the models was not eliminated. This can be explained by there not being sufficient consistency between the incidence data and the underlying renewal equation model.

An obvious advantage of opting for the two-host model is being able to extract the age-specific reproduction numbers, *Rc,t*and *Ra,t*. These can be useful when interventions acting on the infectiousness of a target age group are developed. Here *Rc,t*and *Ra,t* estimates did not differ to a large extent. It is, however, important to recognise that scaling up the standard contact matrix can potentially have a more dramatic effect on the time-dependent age-specific reproduction numbers given a different context. If additionally, it is assumed that adults are more likely to receive unsafe burials than children, *Ra,t*is even more likely to exceed *Rc,t*.. This in turn would indicate that such interventions as, for example, safe burial practices are to be focused on.

Among possible limitations to be addressed is the fact that the serial interval characteristics are not being estimated in parallel with *Rt*, and instead an estimate from literature is used. In reality, such an estimate may often not be readily available early on in an outbreak. This is mostly applicable to outbreaks of newly emerging infections. In this report we assume that natural history and characteristics of EVD are relatively well known. A statistical framework that enables to avoid this limitation, provided there is sufficient data, is described in Thompson et al., 2019.

Another limitation concerns the effective contact matrix used throughout the analysis. The matrix draws on assumptions such as that only adults take part in unsafe burial practices and so the standard contact matrix is scaled accordingly. Also, the values used for the scaling up are only rough estimates taken from previous studies and in principle can vary greatly across different outbreaks which in turn can affect any further analysis. Therefore, the effective contact matrix introduced in this report is to be treated as a purely illustrative tool that enables to highlight the importance of differentiating between it and the standard contact matrix as the latter is likely unsuitable when a disease with several transmission routes, additional to community transmission, like EVD, is considered.

Overall, from our analyses it is evident that accounting for age heterogeneity during infectious disease (EVD in this case) outbreaks is essential for obtaining accurate estimates of the time-dependent reproduction number in real-time, especially early on in an outbreak. This ensures that not only any relevant control measures are applied in a timely manner and target the right age groups but also that these measures are relaxed as soon as the risk of future transmission is estimated to be sufficiently low so that limited control resources are conserved. Finally, because the two-host model investigated in this report relies heavily on age-stratified data being readily available, it is vital to maintain rigorous surveillance to acquire up-to-date epidemiological data.

**Author contributions**

*Formal analysis:* DK.

*Software:* DK.

*Methodology:* DK, RNT, WSH.

*Supervision:* RNT, WSH.

*Writing – Original Draft Preparation:* DK.

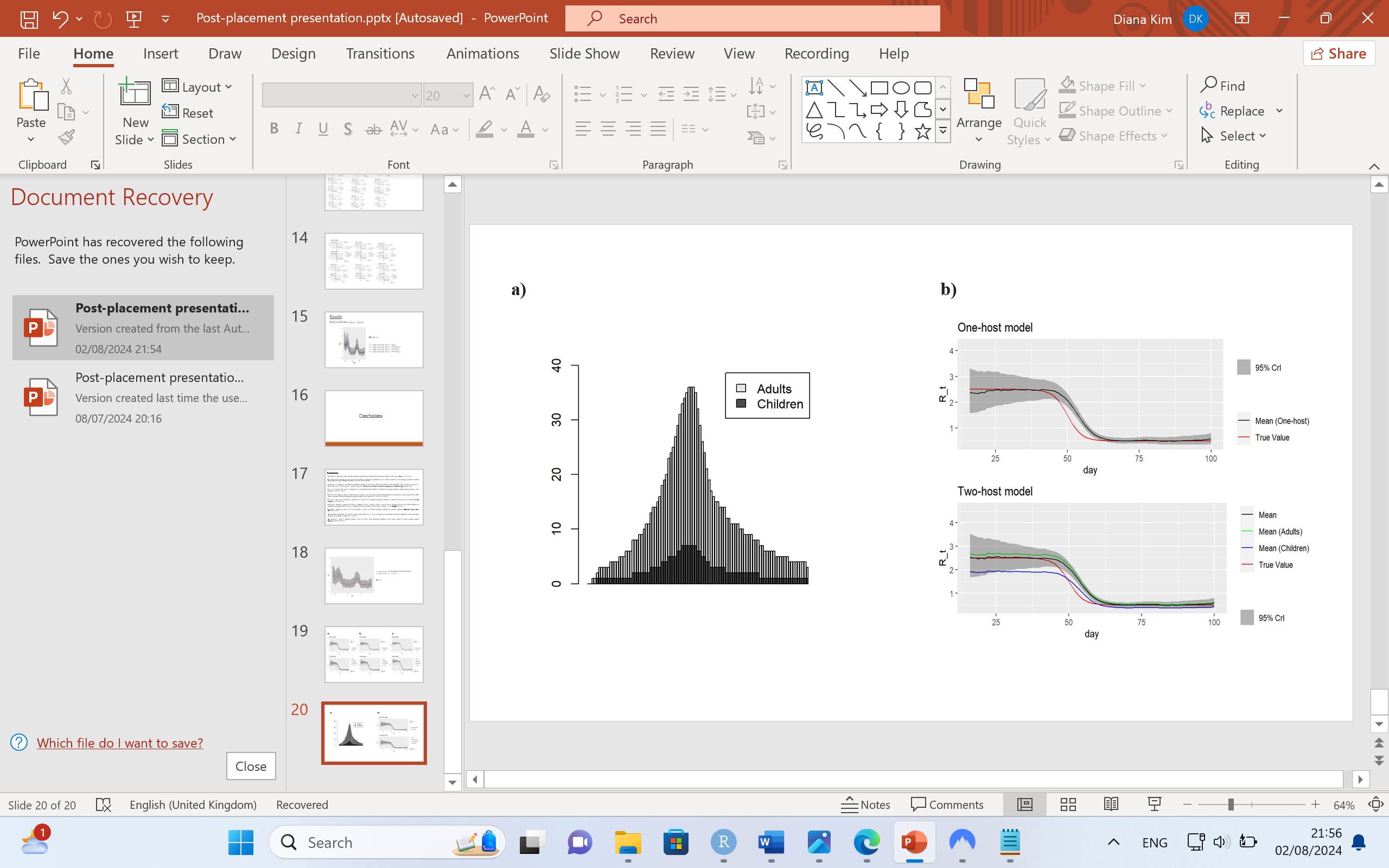
*Writing – Review & Editing:* RNT, WSH.

***Competing interests***

The author declares that no competing interests exist.

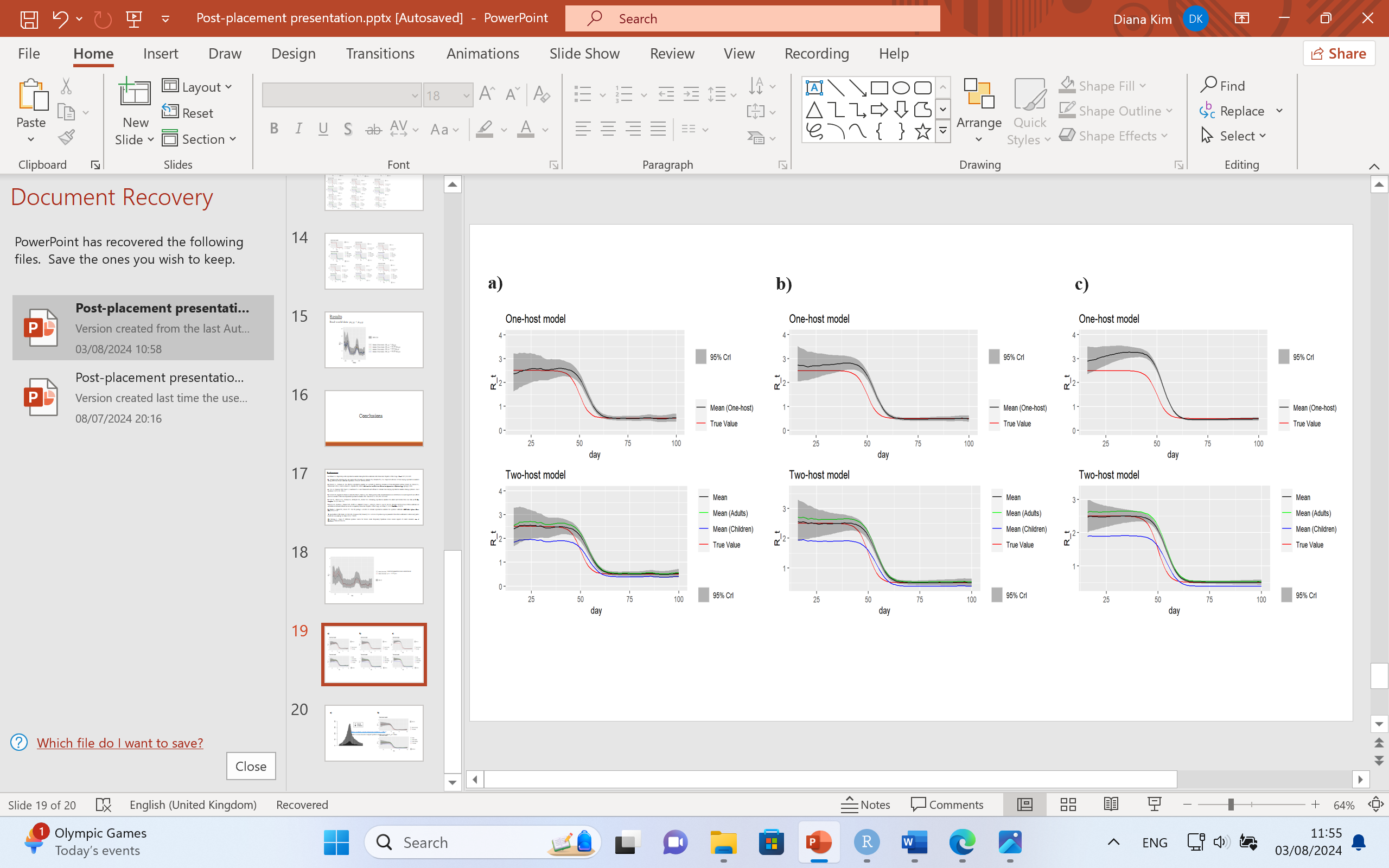
**Financial disclosure**

The author received no specific funding for this work.



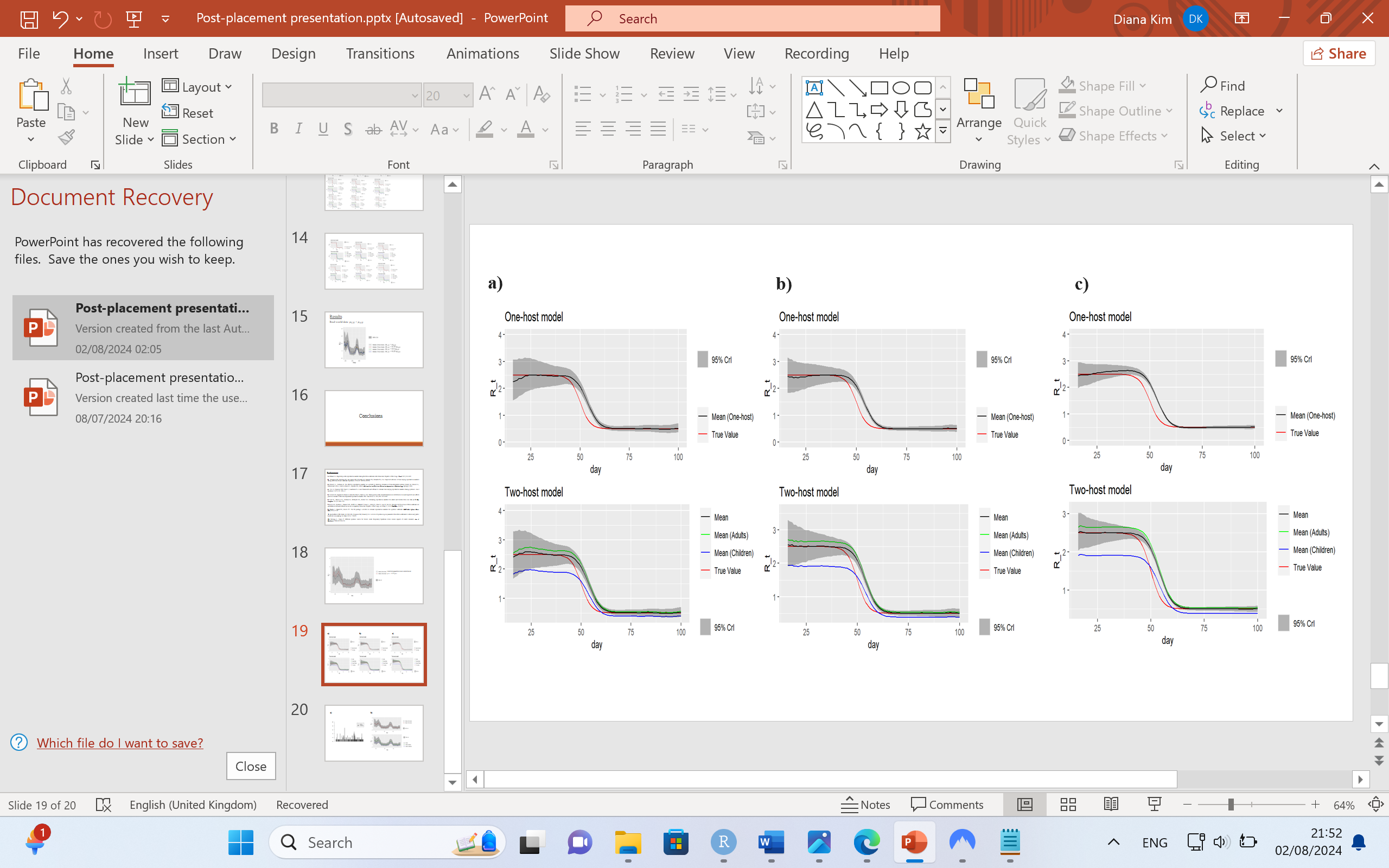
***Figure 1:******Testing the models with simulated data when the serial interval distribution is not age-specific***

***a)*** *the daily incidence counts by age generated by the two-host model;* ***b)*** *estimates of the time-dependent reproduction numbers ( Rt, Rc,t and Ra,t ) from computer-generated data with 95% credible intervals for Rt. The data are generated using an effective contact matrix for DRC for the year 2019* (16)*. The estimates are plotted as soon as the sliding window of one week and at least one average serial interval have passed to ensure the associated bias in the early estimates of Rt is reduced (6).*



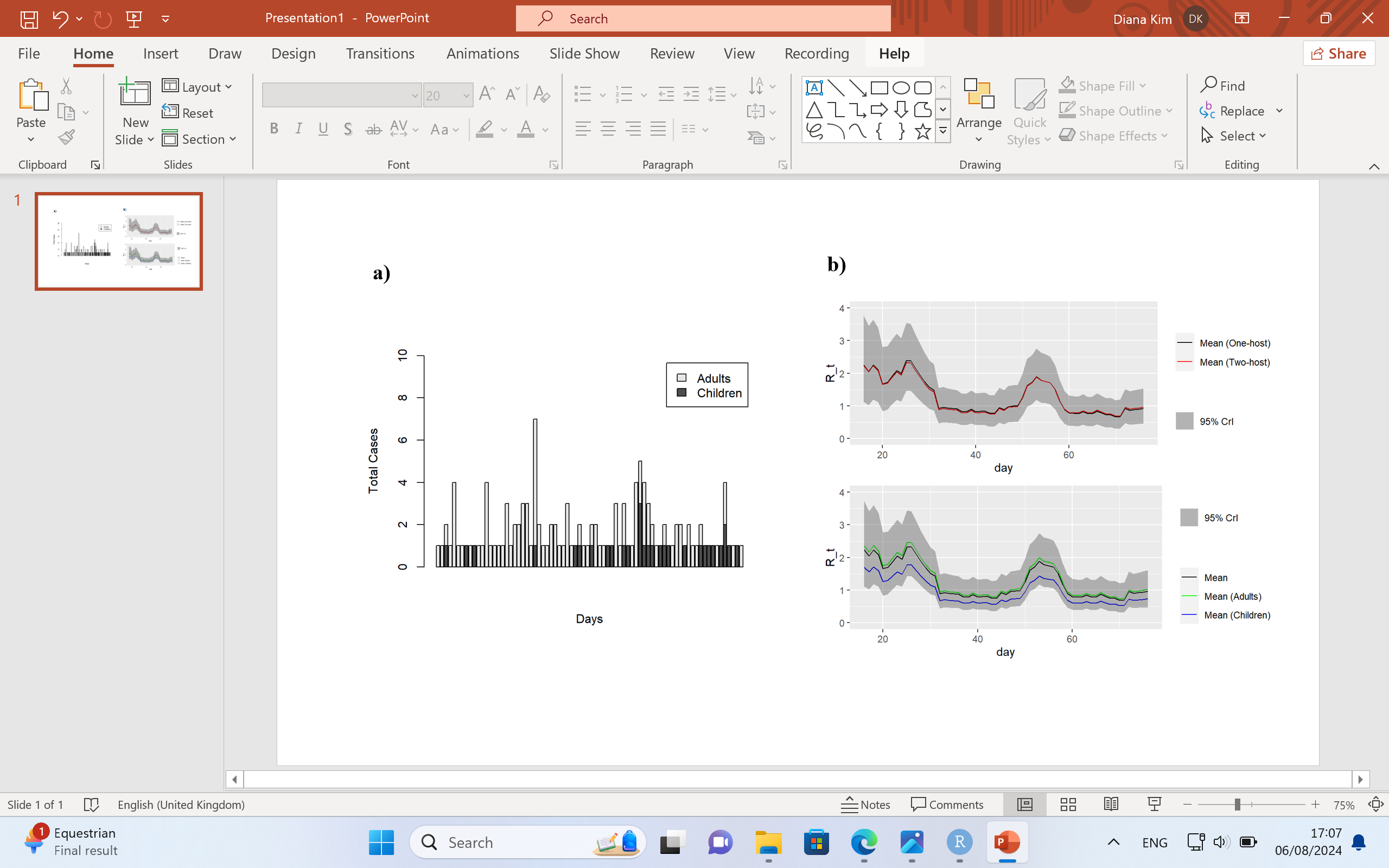
***Figure 2: Testing the models with simulated data when age-specific serial interval distributions are introduced with no correction for the one-host model***

*Estimates of the time-dependent reproduction numbers (Rt, Rc,t and Ra,t) by the one-host and the two-host models with 95% credible intervals for Rt when the population serial interval distribution, , is not corrected for the one-host model:* ***a)*** *= 0.75 ,* ***b)*** *= 0.5 ,* ***c)*** *= 0.25 . The estimates are plotted as soon as the sliding window of one week and at least one average serial interval have passed to ensure the associated bias in the early estimates of Rt is reduced (6).*



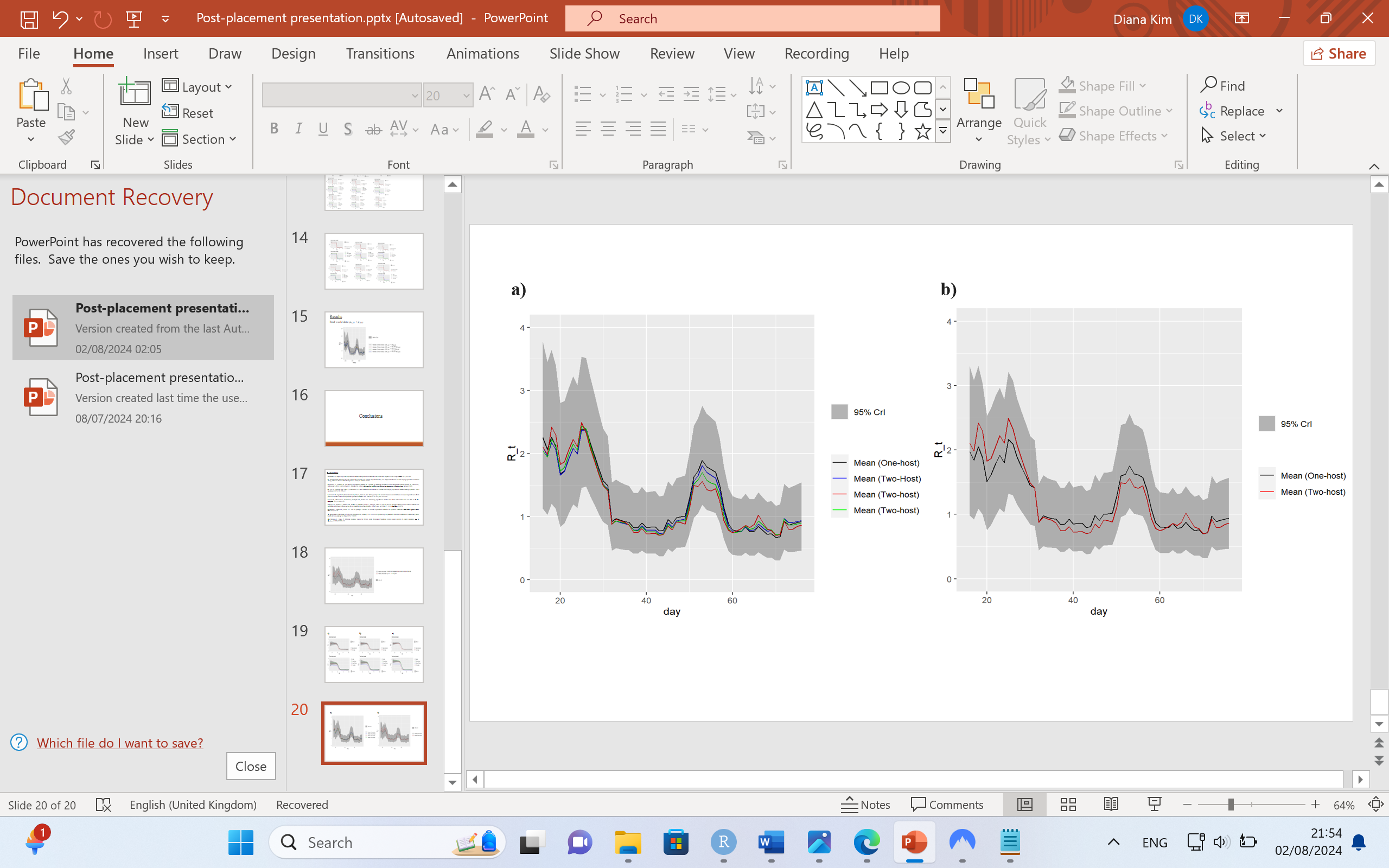
***Figure 3: Testing the models with simulated data when age-specific serial interval distributions are introduced with the correction for the one-host model***

*Estimates of the time-dependent reproduction numbers by the one-host and the two-host models with 95% credible intervals for the Rt estimates when the population serial interval distribution, , is corrected for the one-host model:* ***a)*** *= 0.75 ,* ***b)*** *= 0.5 , and* ***c)*** *= 0.25 . The estimates are plotted as soon as the sliding window of one week and at least one average serial interval have passed to ensure the associated bias in the early estimates of Rt is reduced (6).*



***Figure 4: Testing the models with the real-world data when the serial interval distribution is not age-specific***

***a)*** *the daily incidence counts by age during the 2020 EVD outbreak in Équateur Province, DRC;* ***b)*** *top: the one-host model Rt estimates and the two-host model Rt estimates for the real-world data with 95% credible intervals for the one-host model estimates; bottom: the two-host model estimates of the time-dependent reproduction numbers (Rt, Rc,t and Ra,t) with 95% credible intervals for the Rt estimates.* *The estimates are plotted as soon as the three criteria described in the Web Appendix 3 of Cori et al., 2013 for when to start estimating the reproduction number in an epidemic are satisfied, that is starting from day 16.*



***Figure 5:******Testing the models for the real-world data when age-specific serial interval distributions are introduced without the correction and with the correction ( = 0.25) for the one-host model***

***a)*** *the one-host model Rt estimates plotted together with the two-host model Rt estimates for the real-world data* from *the 2020 EVD outbreak in Équateur Province, DRC for the three scenarios: = 0.75 (in blue), = 0.5 (in green), = 0.25 (in red) without the population serial interval distribution, , being corrected for the one-host model;* ***b)*** *the one-host model Rt estimates plotted together with the two-host model Rt estimates for the real-world data when = 0.25 and the population serial interval distribution, , is corrected for the one-host model. The estimates are plotted as soon as the three criteria described in the Web Appendix 3 of Cori et al., 2013 for when to start estimating the reproduction number in an epidemic are satisfied, that is starting from day 16.*

**Supplementary Materials**

*Graphics*

S1 [Supplementary Figure S1 a).png, Supplementary Figure S1 b).png, Supplementary Figure S1 c).png] – age-specific pathogen transmissibility plots

Figures showing *Rt* estimates by the one-host and the two-host models plotted for several scenarios.

*Files*

SA [Supplementary File A.xlsx] – Daily incidence data by age

Excel file containing daily incidence by age from the 2020 EVD outbreak in Équateur Province, DRC.

SB [Supplementary File B] – Population structure data (DRC)

R file containing a data frame with population age composition in five-year age intervals for the year 2020 for 201 countries.

SC [Supplementary File C.rdata] - Contact data

R file containing 16x16 synthetic daily contact (all locations) matrices for 177 regions.

SD [Supplementary File D.rdata] – Formatted incidence data

Excel file containing daily incidence counts split into two age groups (< 20 and ≥ 20 years).

SE [Supplementary File E.docx] – Age-specific pathogen transmissibility analysis

Further detail and results of introducing age-specific pathogen transmissibility.

SF [Supplementary File F.rdata] – R code

R code for data formatting, model simulations and analysis relevant to the current submission.

**Data**

*Data Availability Statement*

The incidence, population structure and contact data used were publicly available and can be found using the following links:

Incidence data (also provided in Supplementary File A):

<https://zenodo.org/records/11091713/preview/will-s-hart/end-of-outbreak-v1.0.3.zip?include_deleted=0#tree_item11>

Population structure data (also provided in Supplementary File B):

<https://github.com/kieshaprem/synthetic-contact-matrices/raw/master/compare_contact_matrices/input/poptotal.rdata>

Contact data (also provided in Supplementary File C):

<https://github.com/kieshaprem/synthetic-contact-matrices/raw/master/output/syntheticmatrices/contact_all.rdata>

*Code Availability Statement*

All code for data formatting, model simulations and analysis relevant to the current submission is available publicly at <https://github.com/EstimatingRt/Age-stratified-model-for-Ebola-Virus-Disease> (also provided in Supplementary File F).

**References**

1. Nishiura H, Chowell G. The Effective Reproduction Number as a Prelude to Statistical Estimation of Time-Dependent Epidemic Trends. Math Stat Estim Approaches Epidemiol. 2009;103–21.

2. Obadia T, Haneef R, Boëlle PY. The R0 package: a toolbox to estimate reproduction numbers for epidemic outbreaks. BMC Med Inform Decis Mak. 2012 Dec 18;12:147.

3. Kelly JD, Worden L, Wannier SR, Hoff NA, Mukadi P, Sinai C, et al. Projections of Ebola outbreak size and duration with and without vaccine use in Équateur, Democratic Republic of Congo, as of May 27, 2018. PloS One. 2019;14(3):e0213190.

4. Althaus CL. Rapid drop in the reproduction number during the Ebola outbreak in the Democratic Republic of Congo. PeerJ. 2015 Nov 19;3:e1418.

5. Glass K, Mercer GN, Nishiura H, McBryde ES, Becker NG. Estimating reproduction numbers for adults and children from case data. J R Soc Interface. 2011 Sep 7;8(62):1248–59.

6. Cori A, Ferguson NM, Fraser C, Cauchemez S. A new framework and software to estimate time-varying reproduction numbers during epidemics. Am J Epidemiol. 2013 Nov 1;178(9):1505–12.

7. Wallinga J, Teunis P. Different Epidemic Curves for Severe Acute Respiratory Syndrome Reveal Similar Impacts of Control Measures. Am J Epidemiol. 2004 Sep 15;160(6):509–16.

8. Thompson RN, Stockwin JE, van Gaalen RD, Polonsky JA, Kamvar ZN, Demarsh PA, et al. Improved inference of time-varying reproduction numbers during infectious disease outbreaks. Epidemics. 2019 Dec;29:100356.

9. Creswell R, Augustin D, Bouros I, Farm HJ, Miao S, Ahern A, et al. Heterogeneity in the onwards transmission risk between local and imported cases affects practical estimates of the time-dependent reproduction number. Philos Trans R Soc Math Phys Eng Sci. 2022 Aug 15;380(2233):20210308.

10. Tiffany A, Dalziel BD, Kagume Njenge H, Johnson G, Nugba Ballah R, James D, et al. Estimating the number of secondary Ebola cases resulting from an unsafe burial and risk factors for transmission during the West Africa Ebola epidemic. PLoS Negl Trop Dis. 2017 Jun 22;11(6):e0005491.

11. Prescott J, Bushmaker T, Fischer R, Miazgowicz K, Judson S, Munster VJ. Postmortem Stability of Ebola Virus. Emerg Infect Dis. 2015 May;21(5):856–9.

12. Sikakulya FK, Ilumbulumbu MK, Djuma SF, Bunduki GK, Sivulyamwenge AK, Jones MK. Safe and dignified burial of a deceased from a highly contagious infectious disease ebolavirus: Socio-cultural and anthropological implications in the Eastern DR Congo. One Health. 2021 Aug 13;13:100309.

13. WHO | Regional Office for Africa [Internet]. 2022 [cited 2024 Jun 23]. The Democratic Republic of the Congo declares 14th Ebola outbreak over. Available from: https://www.afro.who.int/countries/democratic-republic-of-congo/news/democratic-republic-congo-declares-14th-ebola-outbreak-over

14. Mukadi-Bamuleka D, Nkuba-Ndaye A, Mbala-Kingebeni P, Ahuka-Mundeke S, Muyembe-Tamfum JJ. Impact of Ebola epidemics on the daily operation of existing systems in Eastern Democratic Republic of the Congo: a brief review. J Med Econ. 2024;27(1):184–92.

15. Van Kerkhove MD, Bento AI, Mills HL, Ferguson NM, Donnelly CA. A review of epidemiological parameters from Ebola outbreaks to inform early public health decision-making. Sci Data. 2015;2:150019.

16. Prem K, Zandvoort K van, Klepac P, Eggo RM, Davies NG, Group C for the MM of IDC 19 W, et al. Projecting contact matrices in 177 geographical regions: An update and comparison with empirical data for the COVID-19 era. PLOS Comput Biol. 2021 Jul 26;17(7):e1009098.

17. World Population Prospects - Population Division - United Nations [Internet]. [cited 2024 Jul 22]. Available from: https://population.un.org/wpp/

18. Drake JM, Kaul RB, Alexander LW, O’Regan SM, Kramer AM, Pulliam JT, et al. Ebola Cases and Health System Demand in Liberia. PLOS Biol. 2015 Jan 13;13(1):e1002056.

19. Li J, Duan HJ, Chen HY, Ji YJ, Zhang X, Rong YH, et al. Age and Ebola viral load correlate with mortality and survival time in 288 Ebola virus disease patients. Int J Infect Dis. 2016 Jan;42:34–9.